



# Understanding the role of microRNAs in wheat spike development

Sophie J. Carpenter

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

John Innes Centre

September 2024

*This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.*

# Abstract

The role of microRNAs, small RNA molecules which repress target genes, in the development of the wheat spike remains largely unknown. In this thesis, I have broadened and deepened our understanding of miRNA functions during early wheat spike development, a period when several yield-determining traits, such as spikelet number, are established.

Using a smallRNA-Seq timecourse, I have generated a comprehensive database of miRNAs that are expressed during early wheat spike development and their putative target transcripts. Using this dataset, I have identified natural and induced variation in miRNA binding sites; a large proportion of these SNPs likely confer dominant, gain-of-function mutations which are promising candidates for integration into breeding pipelines. I have found that natural miRNA binding site variation in the A. E. Watkins landrace collection is significantly associated with changes in spike phenotypes.

Additionally, I have shown how this dataset can be used alongside techniques such as spatial transcriptomics and luciferase assays to deepen our understanding of specific regulatory networks involving miRNAs. I have interrogated the miR172-*AP2* interaction in detail and found that almost identical miR172 family members have divergent expression profiles during wheat spike development, exemplifying the importance of characterising closely-related miRNAs independently. However, I have also shown that miR172 family members likely repress the closely related *AP2-2* and *AP2-5* genes with equivalent efficiency.

The functions of and interactions between *VRT-A2*, *AP2-5*, and miR172 appear to be mostly conserved between Arabidopsis and wheat. However, increases in *AP2-2* mRNA levels over time and decreases in response to increased and extended *VRT-A2* expression cannot be explained by current data. I have generated hypotheses to explain these unexpected observations and proposed experiments to test them. I have used our findings to create a model explaining how *VRT-A2*, miR172, and *AP2*-like genes work together to form a complete and fertile wheat spike.

## **Access Condition and Agreement**

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

# Table of Contents

Abstract .....	II
Table of Contents .....	III
List of tables.....	XII
List of figures.....	XIII
List of accompanying material.....	XVI
Acknowledgements.....	XVII
1 General Introduction .....	1
1.1 Wheat yields are not keeping up with projected demand .....	1
1.2 The nature of the wheat genome presents challenges for genetic studies .....	2
1.2.1 Recent developments mean that fundamental research is now more amenable in polyploid wheat.....	4
1.3 We are beginning to understand the processes that control wheat inflorescence development.....	6
1.3.1 The ABCDE model of wheat development is partially conserved between Arabidopsis and wheat .....	8
1.4 miRNAs play important roles in wheat spike development .....	10
1.4.1 miRNAs are crucial components of floral development in plants .....	10
1.4.2 miRNA biogenesis .....	11
1.4.3 miRNA-mediated repression.....	13
1.5 Improving our understanding of wheat inflorescence development will help address food security .....	14
1.6 Thesis aims .....	14
2 The role of microRNAs during key developmental transitions in the wheat inflorescence .....	15
2.1 Chapter summary.....	15
2.2 Introduction .....	15
2.2.1 There is no comprehensive database for miRNAs in wheat.....	15
2.2.2 miRNA binding sites are an untapped source of variation in wheat .....	16

2.2.3	Modifying miRNA binding sites could help us overcome the issue of redundancy in polyploid crops.....	17
2.2.4	Aims and hypotheses.....	18
2.3	Methods.....	18
2.3.1	Small RNA-Seq.....	18
2.3.2	Small RNA-Seq data analysis .....	20
2.3.2.1	Data trimming.....	20
2.3.2.2	Predicting miRNAs .....	20
2.3.2.3	Generating RPM values for candidate miRNAs.....	21
2.3.2.4	Predicting miRNA targets.....	21
2.3.3	Identifying variation in TILLING and Watkins lines .....	21
2.3.4	TILLING population phenotyping .....	22
2.3.5	Watkins population phenotyping.....	22
2.4	Results .....	23
2.4.1	Over 500 mature miRNAs are expressed in wheat spikes during early development .....	23
2.4.2	3,535 transcripts contain sequences which may be bound by the 527 predicted mature miRNAs.....	29
2.4.2.1	This pipeline identified known miRNA-mRNA interactions .....	29
2.4.2.1.1	miR156 targets SPL genes.....	29
2.4.2.1.2	miR165/166 targets HD-ZIP III transcription factors.....	30
2.4.3	Variation in miRNA target sites.....	31
2.4.3.1	There is induced variation in miRNA target sites in wheat TILLING populations.....	31
2.4.3.2	This dataset can be used to identify useful induced variation in miRNA target sites .....	35
2.4.3.3	Natural variation in miRNA target sites is correlated with spike phenotypes .....	38
2.5	Discussion.....	42
2.5.1	This dataset represents a comprehensive overview of miRNAs expressed during early wheat spike development .....	42
2.5.2	A novel locus translation tool is useful resource for the genomics community ... ..	44

2.5.3	A comprehensive wheat miRNA database constitutes a key resource for developmental biology .....	44
2.5.4	A comprehensive wheat miRNA database can be used for better wheat breeding .....	45
2.5.5	Future perspectives.....	47
3	How do <i>APETALA2-2</i> and <i>APETALA2-5</i> have different expression patterns in developing spikes when they are both regulated by microRNA172? .....	48
3.1	Chapter summary.....	48
3.2	Introduction .....	48
3.2.1	The interaction between miR172 and <i>AP2</i> -like genes is a cornerstone of wheat spike development.....	48
3.2.2	<i>AP2-2</i> and <i>AP2-5</i> have partially redundant roles in wheat spike development .....	50
3.2.3	Knowledge from other species shows we must properly disentangle the functions of miRNA family members .....	51
3.2.4	Aims and hypotheses.....	52
3.3	Methods .....	52
3.3.1	Phylogenetic analysis .....	52
3.3.2	Paragon RNA-Seq.....	52
3.3.3	Chinese Spring RNA-Seq.....	53
3.3.3.1	RNA Sequencing .....	53
3.3.3.2	Bioinformatic analysis.....	54
3.3.4	miR172 meta-analysis .....	54
3.3.5	Syntenic analysis.....	55
3.3.6	MERFISH .....	55
3.3.6.1	Plant tissue.....	55
3.3.6.2	Sample preparation.....	56
3.3.6.3	MERFISH imaging.....	56
3.3.6.4	Data analysis.....	56
3.3.7	sRNA-Seq .....	56
3.3.7.1	sRNA-Seq statistical analysis.....	57
3.3.8	Identifying miRNA target sites.....	57

3.3.9	Dual luciferase assay.....	57
3.3.9.1	Preparation of electro-competent <i>A. tumefaciens</i> .....	57
3.3.9.2	Plasmid construction .....	58
3.3.9.2.1	miRNA overexpression plasmids .....	58
3.3.9.2.2	Dual luciferase target plasmids.....	62
3.3.9.2.3	Transformation into <i>A. tumefaciens</i> .....	64
3.3.9.3	Infiltrations .....	67
3.3.9.4	Sample collection .....	69
3.3.9.5	Protein assay.....	70
3.3.9.6	Statistical analysis .....	71
3.3.10	Generating <i>AP2-5A</i> / <i>AP2-2A</i> miR172 target site swap transgenic lines .....	71
3.3.10.1	Plasmid construction.....	71
3.3.10.1.1	Constructing the L0 plasmids.....	71
3.3.10.1.1.1	pUC-GW-Kan-AP2-5A PROM+5'UTR L0.....	71
3.3.10.1.1.2	pUC-GW-Kan-AP2-5A CDS L0.....	72
3.3.10.1.1.3	AP2-5A CDS+AP2-2A L0 .....	72
3.3.10.1.1.4	pUC-GW-Kan-AP2-5A 3'UTR L0.....	73
3.3.10.1.1	Constructing the L1 plasmids.....	74
3.3.10.1.1.1	pICH47742-AP2-5A L1 .....	74
3.3.10.1.1.2	pICH47742-AP2-5A-AP2-2A miR172 binding site L1 .....	74
3.3.10.1.2	Constructing the L2 plasmids.....	75
3.3.10.2	Genotyping donor material .....	78
3.4	Results .....	78
3.4.1	<i>AP2-2</i> and <i>AP2-5</i> form separate phylogenetic clades .....	78
3.4.2	<i>AP2-2</i> and <i>AP2-5</i> have opposite patterns of expression during wheat spike development .....	79
3.4.3	There are three mature miR172 family members in <i>T. aestivum</i> .....	83
3.4.4	miR172a has better complementarity to <i>AP2-5</i> and miR172b has better complementarity to <i>AP2-2</i> .....	86
3.4.5	Hypotheses to explain the opposite expression patterns of <i>AP2-2</i> and <i>AP2-5</i> .....	87
3.4.6	Testing hypothesis 1: The expression domains of <i>AP2</i> -like genes and miR172 do not overlap spatially or temporally.....	89
3.4.6.1	The expression domains of <i>AP2-2</i> and <i>AP2-5</i> partially overlap.....	89

3.4.6.2	miR172a and miR172b abundance diverges after terminal spikelet stage during wheat spike development .....	96
3.4.6.3	sRNA-Seq reveals an additional miR172 family member which has predicted targets in <i>T. aestivum</i> .....	98
3.4.6.4	Hypothesis 1 conclusions .....	99
3.4.7	Testing hypothesis 2: Different mature miR172 family members target different <i>AP2</i> genes.....	103
3.4.7.1	TargetFinder predicts that miR172 targets <i>AP2-5</i> , but not <i>AP2-2</i> .....	103
3.4.7.2	miR172a represses <i>AP2-2</i> and <i>AP2-5</i> target sites with equal efficiency .....	103
3.4.7.3	miR172 target site swap transgenics .....	107
3.4.7.4	Hypothesis 2 conclusions .....	109
3.5	Discussion.....	109
3.5.1	The opposite expression patterns of <i>AP2-2</i> and <i>AP2-5</i> reflects the distinct functions of their orthologues in <i>A. thaliana</i> .....	109
3.5.2	Advances in technology and resources have allowed us to deepen our understanding of miR172 in wheat .....	110
3.5.3	Developing hypotheses to explain the opposite expression profiles of <i>AP2-2</i> and <i>AP2-5</i> .....	111
3.5.3.1	There are few regions where <i>AP2-2</i> is expressed in the absence of <i>AP2-5</i> .....	111
3.5.3.2	miR172a and miR172b expression profiles are unique, but both overlap with <i>AP2</i> -like gene expression.....	112
3.5.3.3	miR172a and miR172b repress <i>AP2-2</i> and <i>AP2-5</i> target sites equally..	112
3.6	Conclusions and future perspectives .....	113
4	Does <i>VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2)</i> act via microRNA172 to achieve its phenotypic effect? .....	115
4.1	Chapter summary.....	115
4.2	Introduction .....	115
4.2.1	A natural allele of <i>VRT-A2</i> leads to increased and extended expression.....	115
4.2.1.1	<i>Triticum polonicum</i> .....	115
4.2.1.2	<i>VRT-A2</i> was identified as the causal gene for the long glume and long grain phenotype in <i>Triticum polonicum</i> .....	116
4.2.1.3	A sequence substitution in intron 1 of <i>VRT-A2</i> is responsible for the <i>T. polonicum</i> phenotype .....	116

4.2.1.4	The sequence substitution in intron 1 of <i>VRT-A2b</i> causes increased and ectopic expression in vegetative and reproductive tissues .....	116
4.2.2	<i>VRT-A2</i> is an ortholog of <i>SVP</i> in <i>Arabidopsis</i> .....	117
4.2.2.1	The three <i>SVP</i> genes in wheat have different expression patterns .....	118
4.2.2.2	The <i>AtSVP</i> regulatory network has been well characterised .....	119
4.2.3	The regulatory network of <i>VRT-A2</i> needs to be investigated.....	120
4.2.3.1	<i>VRT2</i> has been shown to interact with <i>SQUAMOSA</i> proteins.....	120
4.2.3.2	An interaction orthologous to those between <i>SVP</i> , <i>AP2</i> and miR172 in <i>Arabidopsis</i> has not been tested in monocots .....	121
4.2.4	<i>AP2-5A</i> and miR172 mutants have similar phenotypes to <i>VRT-A2</i> mutants..	121
4.2.5	<i>AP2-2</i> orthologue loss-of-function mutants in barley have long glumes and grains .....	121
4.2.6	Aims and hypotheses.....	122
4.3	Methods .....	123
4.3.1	Paragon RNA-Seq.....	123
4.3.2	<i>ap2-5A VRT-A2b</i> phenotyping.....	123
4.3.2.1	Germplasm .....	123
4.3.2.2	Plant growth.....	123
4.3.2.3	DNA extraction.....	124
4.3.2.4	PACE genotyping .....	125
4.3.2.5	Phenotyping.....	125
4.3.2.6	Statistical analysis .....	126
4.3.3	<i>ap2-2 ap2-5A VRT-A2b</i> phenotyping .....	126
4.3.4	sRNA-Seq .....	127
4.3.5	miR172 stem-loop RT-qPCR.....	127
4.3.5.1	Plant growth.....	127
4.3.5.2	Sampling.....	128
4.3.5.3	RNA extraction.....	128
4.3.5.4	cDNA synthesis .....	129
4.3.5.5	Stem-loop RT-qPCR .....	129
4.3.5.6	Statistical analysis .....	129
4.3.6	Generating <i>VRT-A2b::FLAG</i> transgenic lines.....	130
4.3.6.1	L1 Plasmid construction.....	130

4.3.6.2	L2 plasmid construction .....	132
4.3.6.3	Transformation .....	133
4.3.6.4	Phenotyping.....	133
4.3.6.5	RT-qPCR.....	133
4.3.6.6	Western blot.....	134
4.4	Results .....	135
4.4.1	<i>AP2-5</i> expression is significantly higher in a NIL with ectopic <i>VRT-A2</i> expression.....	135
4.4.2	<i>ap2-5A VRT-A2b</i> plants are a near phenocopy of <i>ap2-5A ap2-2</i> mutants .....	137
4.4.2.1	Some <i>ap2-5A/VRT-A2b</i> phenotypes were simply combinations of single mutant phenotypes .....	138
4.4.2.2	There is a significant interaction between <i>AP2-5A</i> and <i>VRT-A2</i> for glume width and length and grain width.....	140
4.4.2.3	<i>VRT-A2b/ap2-5A</i> plants phenocopy <i>ap2-2/ap2-5A</i> mutants.....	142
4.4.3	<i>ap2-2 ap2-5A VRT-A2b</i> lines will provide further insight into the interactions between these key genes .....	147
4.4.4	<i>AP2-2</i> expression is significantly lower in a NIL with ectopic <i>VRT-A2</i> expression .....	149
4.4.5	<i>vrt2</i> spikes have higher miR172 abundance.....	150
4.4.6	Development of <i>VRT-A2b::FLAG</i> transgenics.....	151
4.4.6.1	<i>VRT-A2b::FLAG</i> transgenics have expected phenotypes .....	152
4.4.6.2	<i>VRT-A2::FLAG</i> protein can be detected by an anti-FLAG antibody.....	156
4.4.6.3	A CUT&RUN assay is being developed to identify DNA that <i>VRT-A2</i> protein binds to .....	156
4.5	Discussion.....	157
4.5.1	<i>VRT-A2</i> downregulates miR172 .....	157
4.5.2	Increased and extended <i>VRT-A2</i> expression leads to higher <i>AP2-5</i> expression .....	158
4.5.3	Increased and extended <i>VRT-A2</i> expression leads to lower <i>AP2-2</i> expression .....	159
4.5.4	The tagged transgenics will reveal direct targets of the <i>VRT-A2</i> protein .....	159
4.5.5	Conclusions and future perspectives .....	160
5	General Discussion.....	161

5.1	Thesis summary .....	161
5.2	I have developed novel resources for the wheat community .....	161
5.3	I have developed a more nuanced understanding of miR172-mediated repression of <i>AP2</i> -like genes .....	162
5.4	<i>VRT-A2</i> promotes <i>AP2-5</i> abundance and represses miR172 .....	163
5.4.1	<i>VRT-A2</i> represses miR172 abundance .....	163
5.4.2	<i>VRT-A2</i> promotes <i>AP2-5</i> abundance .....	163
5.5	Patterns of <i>AP2-2</i> mRNA abundance cannot be explained using current data ...	164
5.5.1	<i>AP2-2</i> mRNA abundance increases through developmental time .....	164
5.5.2	Increased and extended <i>VRT-A2</i> expression leads to lower <i>AP2-2</i> mRNA abundance .....	164
5.6	Generating hypotheses to explain the unexpected patterns of <i>AP2-2</i> mRNA abundance .....	165
5.6.1	Variation in the dominant mechanism of miR172-mediated repression (miR172 mechanism hypothesis) .....	165
5.6.1.1	The miRNA mechanism hypothesis relies on independent changes to the rate of <i>AP2</i> transcription over time .....	168
5.6.1.2	The miR172 mechanism hypothesis relies on changes to the rate of <i>AP2</i> transcription in the presence of increased and extended <i>VRT-A2</i> expression .....	168
5.6.1.3	Changes to the rate of <i>AP2</i> transcription in the presence of increased and extended <i>VRT-A2</i> expression may also be the result of AP2 protein repressing <i>AP2-2</i> transcription .....	169
5.6.2	miR172 is absent from lodicules (miR172 absent from lodicules hypothesis) .....	169
5.7	Generating models to summarise how <i>AP2-2</i> , <i>AP2-5</i> , <i>VRT-A2</i> and miR172 work together to control phytomer differentiation .....	170
5.7.1	Previous studies have hypothesised that <i>VRT-A2</i> expression affects the balance of leaf and axillary meristem repression .....	170
5.7.2	<i>AP2</i> -like genes promote a lemma-like identity, which correlates with differentiation of the axillary meristem into a floret .....	171
5.8	Concluding statement .....	172
	Appendix .....	173
A.1	Primer list .....	173

A.2 miR172 precursor sequences used in dual luciferase assay .....	175
A.3 Synthesised <i>AP2-5A</i> sequences .....	175
A.4 Synthesised 3x FLAG tag sequence .....	179
A.5 Candidate miRNAs .....	181
A.5 Predicted miRNA targets .....	188
A.6 transcript_to_genomic_loci code .....	198
01_Extracting_exon_information_for_loci_transcripts.sh .....	198
02_Translating_the_loci.r .....	204
Glossary .....	224
References .....	226

## List of tables

Table 2-1: Details of <i>Triticum aestivum</i> cv ‘Paragon’ spike samples used for the sRNA-Seq experiment.....	19
Table 2-2: Adapters used during sRNA-seq trimming.....	20
Table 2-3: Summary statistics for sRNA-Seq dataset generated from <i>T. aestivum</i> cv Paragon spikes .....	24
Table 2-4: SNPs in the A. E. Watkins collection located within putative miRNA binding sites which have statistically significant associations with phenotypic differences .....	40
Table 3-1: Reaction setup for double digestion of the dual luciferase miRNA overexpression plasmid inserts and vector.....	59
Table 3-2: Oligos used to generate target sites used in the dual luciferase assay.....	62
Table 3-3: Reaction setup for double digestion of the pGreen_dualluc_3'UTR_sensor plasmid. ....	64
Table 3-4: List of co-infiltration combinations used in dual luciferase experiment 1 .....	67
Table 3-5: List of co-infiltration combinations used in dual luciferase experiment 2 .....	68
Table 3-6: miR172 family members in <i>Triticum aestivum</i> .....	85
Table 3-7: Putative targets of miR172d in <i>Triticum aestivum</i> .....	101
Table 4-1: Details of germplasm used in <i>ap2-5A</i> x <i>VRT-A2b</i> phenotyping experiment .....	124
Table 4-2: Details of germplasm used in <i>ap2-2</i> x <i>ap2-5A</i> x <i>VRT-A2b</i> phenotyping experiment .....	127
Table 4-3: Details of germplasm used to quantify miR172 expression using step-loop RT-qPCR.....	128
Table A-1: List of primers used in this thesis.....	173
Table A-2: Candidate miRNAs generated by miRDeep-P2, miRador and ShortStack from an early wheat spike development sRNA-Seq timecourse referenced in this thesis .....	181
Table A-3: Predicted targets of candidate miRNAs referenced in this thesis .....	188

## List of figures

Figure 1-1: Global wheat yields from 1961-2022.....	2
Figure 1-2: The evolution of hexaploid bread wheat .....	4
Figure 1-3: Phytomers as repeating structures during wheat spike development. ....	7
Figure 1-4: Detailed structure of wheat spikelets .....	8
Figure 1-5: The ABCDE model of development in Arabidopsis and wheat.....	9
Figure 1-6: Summary of miRNA biogenesis in plants .....	12
Figure 1-7: The two mechanisms of miRNA-mediated repression, translational inhibition and mRNA cleavage .....	13
Figure 2-1: Representative images of wheat spikes sampled for sRNA-Seq.....	19
Figure 2-2: Mature miRNAs predicted by three prediction pipelines based on small RNA-Seq data.....	26
Figure 2-3: PCA of small RNA-Seq data generated from <i>T. aestivum</i> cv Paragon spikes.....	28
Figure 2-4: <i>TaSPL</i> genes predicted to be targeted by miRNA candidate-351 (miR156).. ....	30
Figure 2-5: Phenotypic data of Cadenza TILLING population 1 .....	33
Figure 2-6: Phenotypic data of Cadenza TILLING population 2 .....	34
Figure 2-7: TILLING lines with variation in miRNA binding sites .....	36
Figure 2-8: TGW for lines from the A. E. Watkins hexaploid landrace collection with specific SNPs in putative miRNA binding sites .....	41
Figure 3-1: The effect of <i>q</i> and <i>Q</i> alleles of <i>AP2-5A</i> on miR172 binding and spike phenotypes .....	49
Figure 3-2: Plasmid map of pGreen_GUS_competitor-miR172a .....	60
Figure 3-3: Plasmid map of pGreen_GUS_competitor-miR172b .....	61
Figure 3-4: Plasmid map of pGreen_dualuc_3' UTR_sensor .....	63
Figure 3-5: Plasmid map of pSoup .....	66
Figure 3-6: Sampling plan for the dual luciferase assay .....	70
Figure 3-7: Plasmid map of AP2-5A Level 2 .....	76
Figure 3-8: Plasmid map of AP2-5A/2A Level 2 .....	77
Figure 3-9: Phylogenetic tree of AP2-2 and AP2-5 proteins.....	79
Figure 3-10: Expression of <i>AP2-5</i> and <i>AP2-2</i> homoeologues during wheat spike development (Paragon dataset).....	81
Figure 3-11: Expression of <i>AP2-5</i> and <i>AP2-2</i> homoeologues during wheat spike development (Chinese Spring dataset) .....	82
Figure 3-12: Complementarity between <i>AP2-2</i> and <i>AP2-5</i> binding sites and miR172 family members.....	86
Figure 3-13: Images from MERFISH assay showing the expression of <i>AP2-2</i> and <i>AP2-5</i> ... ..	90

Figure 3-14: UMAP clustering of cells in a W5.0 <i>PI<sup>WT</sup></i> wheat spike.....	92
Figure 3-15: Overlap between <i>AP2-2</i> and <i>AP2-5</i> transcripts.....	94
Figure 3-16: Mean log normalised <i>AP2-2</i> and <i>AP2-5</i> transcript count per cell for cell clusters of a W5.0 <i>PI<sup>WT</sup></i> wheat spike.....	95
Figure 3-17: Abundance of mature miR172 family members in developing <i>PI<sup>WT</sup></i> spikes.....	97
Figure 3-18: Abundance of miR172d in developing <i>PI<sup>WT</sup></i> spikes.....	99
Figure 3-19: Constructs used in the dual luciferase assay .....	104
Figure 3-20: Dual luciferase assay testing binding between miR172 family members and <i>AP2</i> -like binding sites .....	105
Figure 3-21: Structure of the miR172 target site swap transgenic constructs and K3946 spike .....	108
Figure 4-1: Phylogenetic tree of the three <i>SVP</i> -like gene clades .....	118
Figure 4-2: Working model describing how increased and extended <i>VRT-A2</i> expression leads to long floral organs and a lax spike .....	122
Figure 4-3: Plasmid map of <i>VRT-A2b::FLAG</i> .....	132
Figure 4-4: Expression of <i>VRT2</i> and <i>AP2-5</i> homoeologues in developing spikes of <i>PI<sup>WT</sup></i> and <i>PI<sup>POL</sup></i> NILs .....	136
Figure 4-5: Primary F <sub>3</sub> spikes from the <i>VRT-A2b</i> x <i>ap2-5A</i> cross at physiological maturity .....	139
Figure 4-6: Glume size phenotypes of F <sub>3</sub> primary spikes from the <i>VRT-A2b</i> x <i>ap2-5A</i> cross .....	141
Figure 4-7: Grain size phenotypes of F <sub>3</sub> primary spikes from the <i>VRT-A2b</i> x <i>ap2-5A</i> cross .....	142
Figure 4-8: Dissected spikelets from F <sub>3</sub> primary spikes from the <i>VRT-A2b</i> x <i>ap2-5A</i> cross .....	143
Figure 4-9: Spikelet-level phenotypes of F <sub>3</sub> primary spikes from the <i>VRT-A2b</i> x <i>ap2-5A</i> cross .....	144
Figure 4-10: Spike phenotypes of <i>ap2-2 ap2-5A</i> and <i>VRT-A2b/ap2-5A</i> .....	146
Figure 4-11: Grain size phenotypes of F <sub>3</sub> primary spikes from the <i>VRT-A2b</i> x <i>ap2-5A</i> x <i>ap2-2</i> cross .....	148
Figure 4-12: Expression of <i>AP2-2</i> homoeologues in developing spikes of <i>PI<sup>WT</sup></i> and <i>PI<sup>POL</sup></i> NILs .....	149
Figure 4-13: miR172a and miR172b expression in <i>PI<sup>WT</sup></i> and <i>PI<sup>POL</sup></i> spikes .....	150
Figure 4-14: miR172 expression in germplasm with differential <i>VRT-A2</i> expression.....	152
Figure 4-15: Glume length from T <sub>0</sub> primary spikes, plotted against <i>VRT-A2b::FLAG</i> copy number .....	153
Figure 4-16: Primary spikes of <i>VRT-A2b::FLAG</i> T <sub>0</sub> plants .....	154

Figure 4-17: T <sub>1</sub> <i>VRT-A2b::FLAG</i> phenotypes at 14 dpa .....	155
Figure 4-18: Western blot of of <i>VRT-A2b::FLAG</i> T <sub>1</sub> transgenic lines .....	156
Figure 5-1: Models of the miR172 mechanism hypothesis to explain unexpected observations .....	166
Figure 5-2: A model summarising the putative roles of <i>AP2</i> -like genes and <i>VRT-A2</i> during phytomer differentiation .....	172

## List of accompanying material

Table S1: Candidate miRNAs generated by miRDeep-P2, miRador and ShortStack from an early wheat spike development sRNA-Seq timecourse.

Table S2: Predicted targets of candidate miRNAs.

Table S3: SNPs in the A. E. Watkins collection in predicted miRNA binding sites with statistically significant associations with phenotypic differences.

The above supplementary tables are included in a separate file.

## Acknowledgements

Firstly, I want to say a huge thank you to my supervisor Cristobal. Your enthusiasm for science is contagious, and you've built such a wonderful group to work with. Thank you for all the wisdom and for making me a better scientist.

I was incredibly lucky to join such a wonderful lab group who made me feel so welcome in the middle of a global pandemic! Thank you to everyone who's been part of the lab over the last four years, it's been a joy to work with you all every day. To Anna, Ricardo, Aura, Yunchuan, Jesus, Simmo, Andy, Vale, Marina, Isa, Max, Oscar and Katie - you've all helped to make the best lab group I could've asked for! A special shout out to Katie, Mac and Cheese for being there for me through thick and thin - you're superstars! Thank you to Nikolai - official biscuit supplier of this PhD. You've always known what to say whenever I hit a hurdle, and I'll miss chatting over a cuppa.

To all the cast members of the Biffen tea room, you've made me look forward to lunch even more than I already did (a hard feat). Mikki and Jake - you've brought joy to every day! Thank you to James, Richard, David, Calum and everyone who's made up the community I've relied upon at JIC. It's truly the most supportive, most fun place I think I'll ever work and it's because of you all. To the gals - Judit, Em, Cat and Deirdre. Thank you for all the coffees, brunches and pints! I was so lucky to have you all in my cohort and I can't believe we've done it.

Eternal thanks have to go to the women who have supported me from afar. To Louisa, Flora, Kitty, Katie and Imogen. I am forever grateful to call you all my friends, I never thought I'd be so lucky to have such rich friendships that have persisted through so much change and distance over the last four years. You are all my rocks and I don't know what I'd do without you!

And finally to my family. Thank you to Benny for being the best crazy ball of fur there is! Thank you to Nannie Sue and Grandad, and to Nanny Pam for being so supportive in everything I do. Thank you to Sam - you are the best brother I could ask for - I can't wait to take on the big smoke together! Finally, thank you to my parents. You have been there for me no matter what through this crazy journey. Thank you for celebrating with me, and for picking up the phone whenever I needed a chat. Thank you for everything. I could not have done this without you.

You have all made this PhD possible.

# I General Introduction

## I.1 Wheat yields are not keeping up with projected demand

In the 10,000 years since its domestication <sup>1</sup>, wheat (*Triticum spp.*) has become a vital crop of great importance to food security and the global economy. It was grown on more land (219 million ha) than any other crop in 2022 <sup>2,3</sup> and is a staple food in many countries, providing 19% of calories and 20% of protein consumed globally in 2018 <sup>2</sup>. The world's population is projected to continue increasing until at least 2100, when there will be an estimated 10.9 billion people to feed <sup>4</sup>. To ensure future food security, yields must increase by 2-3% per year <sup>5</sup>; wheat yields broadly satisfied this requirement until the turn of the millennium, but the rate of yield increase has slowed since then (Figure 1-1). Food security remains a major issue in the 21<sup>st</sup> century, with problems being compounded by the COVID-19 pandemic and climate change <sup>6</sup>. Levels of hunger (measured by the prevalence of undernourishment) have risen since 2019, from 7.5% to 9.1% of the global population in 2024 <sup>7</sup>.

Additionally, Russia's invasion of Ukraine in February 2022 exposed the fragility of global wheat production <sup>8,9</sup>. Ukraine produced 12% of wheat exports in 2020 and in the three months after Russia's invasion exports dropped by 90%, particularly affecting low- and middle-income importers such as Egypt which imported 30% of its wheat from Ukraine in 2021 <sup>9</sup>. Additionally, wheat prices rose by 19.7% in March 2022 <sup>10</sup>. The impacts of this conflict on global food security have increased political focus on developing sustainable and robust food systems <sup>11</sup>, including by maximising production <sup>9</sup>. Stable food supplies and prices have also been highlighted as being politically significant and fluctuations in wheat supplies and prices have been cited as key causes of political instability, such as the Arab Spring <sup>12</sup>.

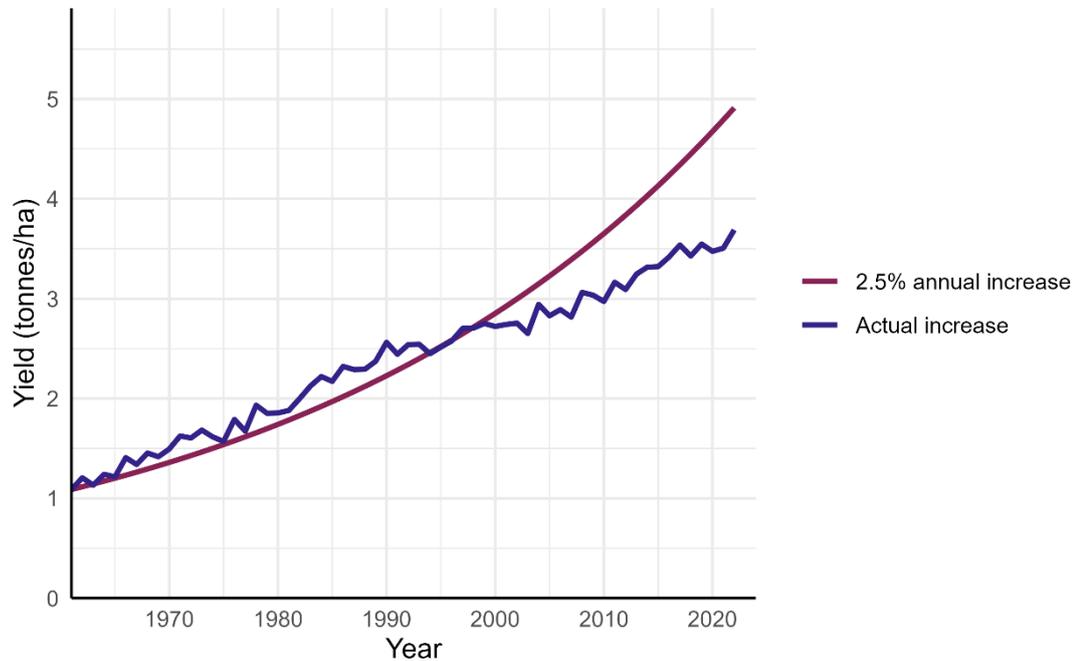


Figure 1-1: Actual global wheat yields from 1961-2022 compared against a 'necessary' 2.5% increase <sup>4</sup> since 1961. Data from UN FAOSTAT <sup>2</sup>.

Wheat is also very important economically. Its global export value was US\$66 billion in 2022 <sup>2</sup> and it is a valuable commodity in both developed and developing nations, with production being approximately equally split between the two <sup>12</sup>. Wheat is worth £2.9 billion to the UK economy alone, 24% of all crop value <sup>13</sup>. Therefore, it is vital that we continue to strive to increase wheat yields, to protect both the economy and food security.

## 1.2 The nature of the wheat genome presents challenges for genetic studies

Compared to model plant species such as *Arabidopsis* (*Arabidopsis thaliana*), wheat is inherently difficult to study, with bread wheat (*Triticum aestivum*) having a hexaploid 16 Gbp <sup>14</sup> genome consisting of 85% repetitive elements <sup>15</sup>. These challenges are compounded by a relatively long generation time of three months for spring wheat (which does not require vernalisation) and six months for winter wheat (which requires vernalisation) <sup>16</sup>. The large and repetitive nature of the wheat genome has meant that knowledge has lagged behind other major crops. A large genome is costly to sequence at sufficient depth for *de novo* assembly, and repetitive genomes present challenges for assembly pipelines. After wheat, maize (*Zea mays*) and rice (*Oryza sativa*) were grown on the most land globally in 2022 <sup>2</sup>. The rice and maize genomes were completed in 2005 <sup>17</sup> and 2009 <sup>18</sup>, respectively, while a fully

assembled and annotated wheat genome (for the cultivar Chinese Spring) was not published until 2018 <sup>15</sup>.

Hexaploid bread wheat is the product of multiple hybridisation events between closely related species (Figure 1-2). These hybridisations brought together the A, B and D wheat genomes and occurred relatively recently in evolutionary time <sup>19</sup>. Less than 820,000 years ago, a hybridisation event between the A genome and B genome progenitors (*Triticum urartu* and a close relative of *Aegilops speltoides*, respectively <sup>19</sup>) formed *Triticum turgidum*, an allotetraploid <sup>19,20</sup>. A domesticated subspecies of *T. turgidum*, *T. turgidum* ssp. *durum*, is grown for pasta and other semolina products; approximately 5% of wheat grown today is *T. turgidum* ssp. *durum* <sup>21</sup>. A second hybridisation event between *T. turgidum* and the D genome progenitor, *Aegilops tauschii*, occurred less than 430,000 years ago <sup>20</sup> to form bread wheat (*T. aestivum*).

67% of genes in bread wheat exist as part of a triad meaning there are homoeologues in the A, B and D genomes <sup>22</sup>. This presents a challenge for studies attempting to elucidate the function of wheat genes as homoeologues share on average 95-98% sequence identity within coding regions <sup>15</sup>.

Borrill, *et al.* <sup>23</sup> summarised one of the main challenges that recent polyploidy creates: functional redundancy. Often, the effect of mutating a gene is masked by its functionally redundant homoeologues <sup>24</sup>. Therefore, all three homoeologues of a gene may need to be disrupted to reveal the full phenotypic effect <sup>25,26</sup>. Due to wheat's long generation time, it takes up to 18 months to generate a triple loss-of-function mutant; this significantly slows the generation of knowledge and the identification of candidate alleles for breeding. Because of functional redundancy, humans have predominantly selected for dominant, gain-of-function alleles, leaving a large pool of recessive variation relatively unexplored <sup>24</sup>.

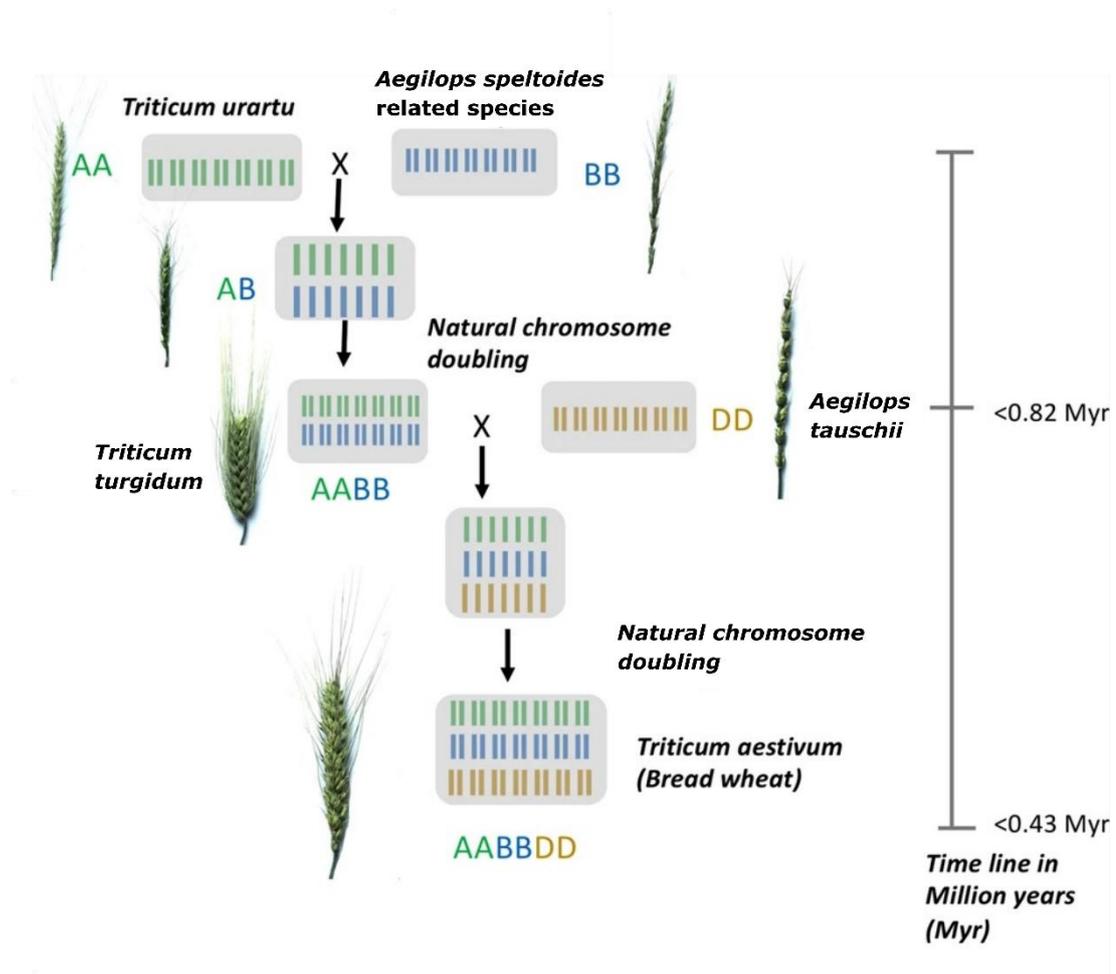


Figure 1-2: The evolution of hexaploid bread wheat. 'X's denote natural hybridisation events. Letters such as "AA" and "AABBDD" represent the genomes present and the ploidy level. *Triticum urartu* and a close relative of *Aegilops speltoides* hybridised to form tetraploid wheat (*Triticum turgidum*) less than 0.82 million years ago (Myr). *Triticum turgidum* subsequently hybridised with *Aegilops tauschii* less than 0.43 Myr to form hexaploid bread wheat (*Triticum aestivum*). Adapted from Rosyara, et al. <sup>27</sup>.

### 1.2.1 Recent developments mean that fundamental research is now more amenable in polyploid wheat

Since the fully annotated Chinese Spring reference genome assembly was published in 2018 <sup>15</sup>, knowledge and resources in wheat have developed at pace. In 2020, additional genome assemblies from 15 cultivars were released <sup>14</sup>, including *T. aestivum* cv Cadenza, which is often used in research. Long generation times have become less of an obstacle to wheat research with the release of 'speed breeding' protocols in 2018 which can be used to halve generation times <sup>28</sup>. Other key resources include a platform for visualising and analysing

RNA-Seq datasets <sup>29</sup> and the publication of a comprehensive developmental gene expression atlas in a *T. aestivum* cv Azhurnaya background <sup>22</sup>. These and other resources <sup>30</sup> have enabled research that would not have been possible ten years ago.

The decreasing cost of sequencing has also opened the door to using a wider range of germplasm, unlocking previously unknown variation. In 2017 a collection of ethyl methanesulfonate (EMS)-mutagenised tetraploid and hexaploid Targeting induced local lesions in genomes (TILLING) lines were released <sup>31</sup>. Exome capture and sequencing was used to catalogue mutations in these lines and it was found that this collection contained at least one truncation or deleterious missense mutation for over 90% of captured genes <sup>31</sup>. This is a critical resource within the wheat community and has been used countless times in the last seven years to understand gene function (*e.g.*, Debernardi, *et al.* <sup>32</sup>) and generate novel germplasm for breeding pipelines (*e.g.*, Simmonds, *et al.* <sup>33</sup>).

A new resource for the wheat community is the recent whole-genome sequencing of the A. E. Watkins hexaploid landrace collection <sup>34</sup>. A detailed description of the collection can be found in Cheng, *et al.* <sup>34</sup> but in brief, the collection comprises 827 pure lines gathered by A. E. Watkins in the 1920s and 1930s from 32 countries. The collection represents germplasm available in local markets at the time, before the advent of modern intensive wheat breeding <sup>34</sup>. This dataset contains a huge amount of untapped variation; 62% of single nucleotide polymorphisms (SNPs), 57% of insertion-deletions (indels), and 53% of copy number variants (CNVs) identified in the collection are absent from modern wheat. Published in 2024, the potential of this resource has only just begun to be explored.

The sequenced TILLING and A. E. Watkins collections are two of the most important sources of novel induced (TILLING) and natural (Watkins) variation for the wheat research community today. The well-characterised variation within these collections accelerates wheat breeding and advances our knowledge of wheat development. The resources available in wheat as well as improvements in experimental protocols, such as the improvement in wheat transformation rates in recalcitrant varieties through the use of a *GROWTH-REGULATING FACTOR 4 (GRF4)* - *GRF-INTERACTING FACTOR 1 (GIF1)* chimera <sup>35</sup>, mean that fundamental research can now be carried out in wheat directly. As Borrill <sup>36</sup> suggested in 2019, the boundaries between model and crop species have become blurred. By combining developmental genetics carried out directly in wheat with knowledge from model species such as *Arabidopsis*, we can maximise the efficiency with which we can translate work in the lab into gains in the field.

### **I.3 We are beginning to understand the processes that control wheat inflorescence development**

As the structure which bears the grains we eat, it is important that we understand how the wheat inflorescence, or spike, develops. By understanding how the inflorescence forms, we can better focus our efforts towards crop improvement.

Shoots in all higher plants are comprised of repeating structural units called phytomers which consists of a leaf, internode, and axillary meristem (Figure 1-3)<sup>37,38</sup>. Programmed repression or release of the leaf and internode of each phytomer determines the architecture of the wheat spike.

During vegetative growth, most axillary meristems are suppressed while leaves develop. Some axillary meristems are initiated to form tillers, but this occurs in a minority of phytomers<sup>38</sup>. The transition from vegetative to reproductive development begins as the shoot apical meristem (SAM) expands along the vertical axis, giving rise to alternating phytomers. The double ridge stage of spike development (Figure 1-3), is so named for the visual appearance of these phytomers. The repeating double ridge structure consists of a lower leaf meristem and upper axillary meristem; the internodes between the meristems (which will later form the rachis) are very short at this stage or are perhaps not yet determined. This phase of spike development is characterised by suppression of the leaves and initiation of the axillary meristems, which differentiate into spikelet meristems. On average, 15-20 phytomers, and therefore spikelets, form per spike<sup>39</sup>. The wheat spike, in contrast to closely related species such as barley<sup>40</sup>, is a determinate structure; the most apical phytomer is rotated 90 ° and forms the terminal spikelet.

This structural plan is repeated for the indeterminate spikelets. Phytomers are formed along the spikelet axis (Figure 1-3). The internodes form a central rachilla. Again, a different programme of leaf and axillary meristem repression and release is observed in the spikelet. Each spikelet is subtended by two vegetative structures called glumes which form from the leaf meristems of the first two phytomers. The axillary meristems of these phytomers are repressed. Florets arise from the subsequent phytomers; the leaf meristem differentiates into a lemma, and the axillary meristems initiate into florets, including the palea, lodicules, stamens and carpel (Figure 1-4). Two glumes and four to six florets develop per spikelet<sup>39</sup>. A higher number of phytomers are formed along the spikelet axis, as spikelets in wheat are indeterminate, however they are aborted and do not fully develop.

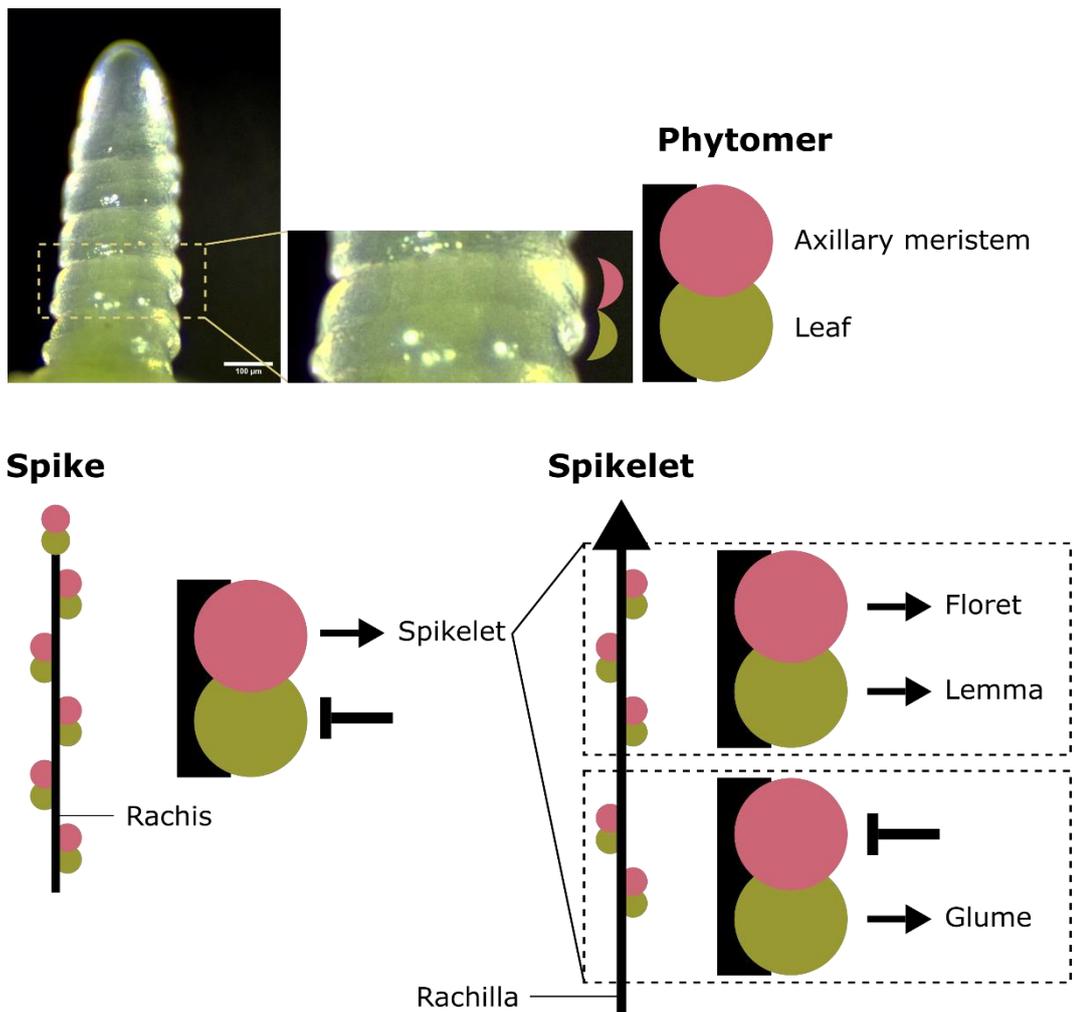


Figure 1-3: Phytomers as repeating structures during wheat spike development. Image of a wheat spike at the W2.5 (double ridge) stage according to Waddington, et al.<sup>41</sup>. Leaf meristems are shown in green, and axillary meristems are shown in pink. Wheat spikes are determinate structures with a terminal spikelet, while spikelets are indeterminate as shown by the arrow at the apex of the rachilla. Along the spike axis, leaf meristems are repressed while axillary meristems differentiate into spikelets. Along the spikelet axis, the first two phytomers develop differently to subsequent phytomers. In the first two phytomers, the leaf meristems differentiate into glumes, while the axillary meristems are repressed. In subsequent phytomers, the leaf meristems differentiate into lemma while the axillary meristems differentiate into florets.

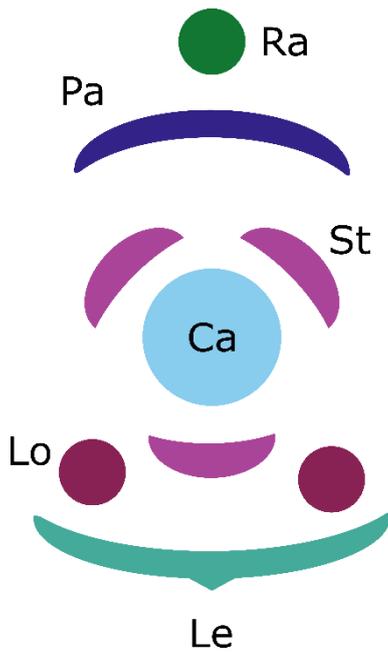


Figure 1-4: Detailed structure of a wheat floret and surrounding structures, shown as a cross-section. Based on Debernardi, et al. <sup>42</sup>. Florets alternate along the central rachilla (Ra), subtended by a lemma (Le). Each floret consists of a palea (Pa), two lodicules (Lo), three stamens (St) and a carpel (Ca) which contains an ovule.

We have begun to unravel the mechanisms which determine whether leaves and axillary meristems are repressed or initiated. Current evidence suggests that the downregulation of *VEGETATIVE TO REPRODUCTIVE TRANSITION 2* (*VRT2*) plays a role in the suppression of leaf ridges in the wheat spike, and that a delay to this downregulation leads to the formation of developmentally-delayed rudimentary basal spikelets <sup>43</sup>. A second paper has proposed a unified model to explain the role of *SHORT VEGETATIVE PHASE* (*SVP*), *SQUAMOSA*, and *SEPALLATA* (*SEP*) genes in the regulation of leaf and axillary meristem initiation and suppression <sup>44</sup>. However, there is significant scope to develop a fuller understanding of how these patterns of meristem repression and initiation are determined. I will investigate the downstream targets of *VRT2* in Chapter 4.

### 1.3.1 The ABCDE model of wheat development is partially conserved between Arabidopsis and wheat

Although visually very distinct from the flowers of eudicots, wheat flowers (florets) are also arranged in whorls according to the ABC model of floral development <sup>45</sup>. The ABC model was first proposed by Coen & Meyerowitz in 1991 <sup>45</sup> and has since been expanded to include D- and E-class genes <sup>46,47</sup>. The genes involved in the ABC model are well characterised in Arabidopsis (Figure 1-5). All of the genes in this model are transcription factors, and all are

MADS-box genes with the exception of *APETALA2* (*AP2*) (which is a member of the AP2/ERF transcription factor family) <sup>48</sup>.

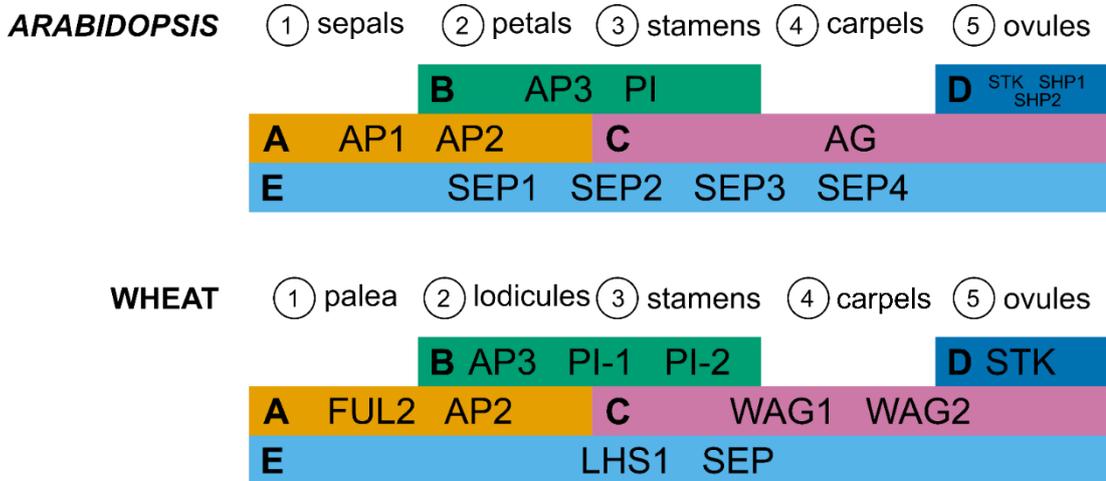


Figure 1-5: The ABCDE model of development in *Arabidopsis* (*Arabidopsis thaliana*) and wheat (*Triticum aestivum*). Flowers in *Arabidopsis* and florets in wheat consist of orthologous whorls, which are listed here in order from outermost to innermost, for example “(1) sepals”. There are five key gene classes which define the identity of each whorl, named A-E. The gene classes expressed in each whorl and example protein names from each class are shown here. Adapted from Murai <sup>49</sup>.

In *Arabidopsis*, flowers consist of five whorls. Each whorl is defined by the expression of different combinations of the ABCDE model transcription factors which are hypothesized to form tetrameric complexes <sup>46</sup>. The whorls in *Arabidopsis* differentiate into 1) sepals, 2) petals, 3) stamens, 4) carpels and 5) ovules <sup>49</sup>. Much of the work done to elucidate this model relied upon homeotic mutant characterisation <sup>45</sup>. These mutants do not contain functional copies of one or more model classes, which results in a homeotic transition where the organ in a particular whorl is replaced with another. For instance, according to the model in Figure 1-5, a knockout of the C-class *AGAMOUS* (*AG*) gene results in ectopic petals in whorl three <sup>45</sup>, a prediction which was confirmed by Bowman, *et al.* <sup>50</sup>.

Floral development has been shown to be largely conserved among land plants. The organs that make up the whorls in monocots have been debated in the literature. It has been argued that the first whorl consists of both lemma and palea <sup>51,52</sup>, lemma comprise the first whorl and palea comprise the second whorl <sup>53</sup>, or that lemma are modified leaves <sup>49,54</sup>. In this thesis, I have built my models according to the hypothesis that lemma are modified leaves. There is strong evidence that lodicules comprise the third whorl as highly modified petals (reviewed in Yoshida <sup>55</sup>) and the fourth and fifth whorls are conserved between *Arabidopsis* and grasses <sup>49</sup>. In wheat, florets contain a single palea (putative whorl 1), two lodicules (whorl 2),

three stamens (whorl 3), one carpel (whorl 4) and one ovule (whorl 5). The genes involved in the ABCDE model in grasses have not been completely understood, and there is some evidence that monocot floral development has diverged from that of eudicots including examples of gene duplication and sub-functionalization <sup>49</sup>.

Genes outside the ABCDE model that affect floral development have also started to be characterised in wheat. For example, ectopic *VRT-A2* expression has been shown to increase grain length in wheat <sup>56</sup>. Also, genes within the ABCDE model have been shown to have functions outside of what the model would predict, for example in barley *AP2* expression affects grain width and length <sup>57</sup>; the expression domain of *AP2* does not include carpels or ovaries according to the ABCDE model (Figure 1-5). These findings show that there is significant scope for deepening our understanding of floral development in grasses including wheat.

As both *VRT-A2* and *AP2* expression have been shown to affect grain length in wheat and barley respectively, it is possible they act within a shared genetic pathway. Altering the expression of *VRT-A2* or *APETALA2-5A* (*AP2-5A*) in wheat affects some of the same phenotypes, such as plant height, heading date, and spikelet density <sup>32,56</sup>. However, *AP2* expression has been shown in barley to affect grain width <sup>57</sup>, while *VRT-A2* expression has no effect on grain width in wheat <sup>56</sup>. This raises the possibility that these two genes affect grain morphology through distinct downstream regulatory pathways. I will investigate the interactions between *VRT-A2* and *AP2*-like genes during floral development and begin to disentangle their effects on wheat grain morphology in Chapter 4.

## 1.4 miRNAs play important roles in wheat spike development

### 1.4.1 miRNAs are crucial components of floral development in plants

Although the genes controlling floral development in wheat have begun to be unravelled, the role that non-coding RNAs play during this critical period of development have only just begun to be studied. miRNAs are a class of small non-coding RNAs that repress target mRNA transcripts.

We know that microRNAs (miRNAs) play critical roles during development in *Arabidopsis* <sup>58</sup>. For example, miR156 and miR172 help to determine the vegetative to reproductive transition. Studies have suggested that these two miRNAs have opposing functions; miR156 specifies juvenile reproductive growth, while miR172 specifies adult reproductive growth <sup>59</sup>. miR156

targets the SQUAMOSA-promoter binding protein-like (SPL) family<sup>59</sup>, which promote flowering<sup>60</sup>, while miR172 regulates *AtAP2*<sup>61</sup>, a floral repressor<sup>62</sup>.

miRNAs are also central to floral development. *AtAP2* also acts as an A-class floral development gene which specifies sepal and petal identity<sup>45</sup> (Figure 1-5). miR164 regulates *CUP SHAPED COTYLEDON1 (CUC1)* and *CUP SHAPED COTYLEDON2 (CUC2)*, which play essential roles in the formation of floral organ boundaries<sup>58,63</sup>, and miR159 regulates *MYB33* and *MYB65* which contribute to proper anther development<sup>64</sup>. Some of these interactions have been shown to be conserved in wheat<sup>32,65,66</sup>.

The evidence above shows that miRNAs play key roles during the vegetative to reproductive transition as well as floral development. This supports the hypothesis that we cannot form a complete understanding of floral development without also understanding the role of miRNAs.

#### 1.4.2 miRNA biogenesis

miRNAs are very short (20-24 bp<sup>67</sup>) single-stranded RNA molecules present in both animals<sup>68</sup> and plants<sup>69</sup>. There have been several excellent reviews of miRNA biogenesis<sup>70,71</sup>. In brief, miRNAs are encoded by *MIR* genes which are transcribed by RNA polymerase II (Pol II)<sup>72</sup> into primary-miRNAs (pri-miRNAs)<sup>70</sup> (Figure 1-6). Pri-miRNAs are characterised by a stem-loop structure<sup>71</sup>, which is cleaved by DICER-LIKE1 (DCL1) into a precursor-miRNA (pre-miRNA) stem-loop<sup>70,71,73</sup>. DCL1 and HUA ENHANCER1 (HEN1) then cleave and methylate the pre-miRNA, respectively, into a miRNA duplex which contains two distinct miRNA sequences on opposite strands<sup>74-76</sup>. It is generally thought that one strand (referred to as the miRNA/guide miRNA) is active while the other (the passenger miRNA/miRNA\*) is degraded<sup>71,74,77</sup>. However, there have been studies showing that the miRNA\* can also be abundant and functional in Arabidopsis<sup>78</sup> and rice<sup>79</sup>. Therefore, it may be more appropriate to refer to miRNAs as 5p or 3p according to their position on the *MIR* gene (as used by Hu, *et al.*<sup>79</sup> and recommended by Budak, *et al.*<sup>80</sup>) instead of naming them based on relative abundance (which is often based on a small number of timepoints and tissues).

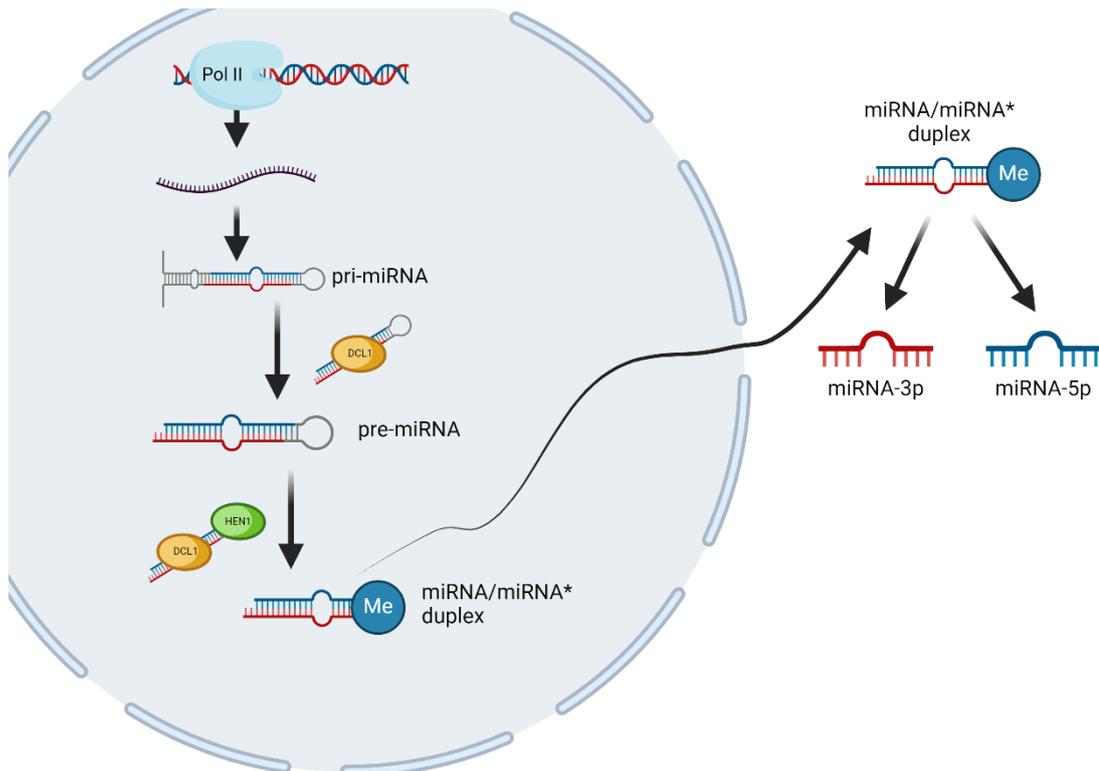
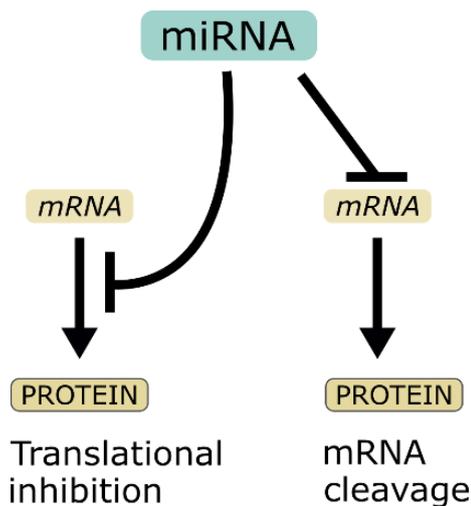


Figure 1-6: Summary of miRNA biogenesis in plants. In the nucleus, RNA Polymerase II (Pol II) transcribes *MIR* genes into stem-loop primary-miRNAs (*pri-miRNAs*). *Pri-miRNAs* are cleaved by *DICER-LIKE1* (*DCL1*) into stem-loop precursor-miRNAs (*pre-miRNAs*). The *pre-miRNA* is further cleaved and methylated by *DCL1* and *HUA ENHANCER* (*HEN1*) into a *miRNA/miRNA\** duplex which is then exported into the cytoplasm. The *miRNA/miRNA\** duplex gives rise to two complementary miRNA molecules, *miRNA-3p* and *miRNA-5p* (named according to their position on the original *MIR* gene). Created in BioRender. Carpenter, S. (2022) [BioRender.com/i50g648](https://BioRender.com/i50g648).

Multiple *MIR* genes can produce identical mature miRNAs. miRNAs often exist in families of almost identical mature sequences with up to two mismatches (although in certain cases miRNAs with up to four mismatches have been assigned to the same family)<sup>81</sup>. For example, the miR172 family in *Arabidopsis* consists of five *MIR* genes which produce three unique mature miR172 sequences<sup>82</sup>. There have been inconsistencies in the nomenclature used for miRNAs<sup>80</sup>, however a common system<sup>80</sup> (and the one I will use in this thesis) uses a number to denote the miRNA family (e.g., miR172). Unique mature sequences within each family are differentiated using a character suffix (e.g., miR172a and miR172b). *MIR* genes which encode the same mature sequence are differentiated using a numerical suffix (e.g., *MIR172A-1* and *MIR172A-2*).

### 1.4.3 miRNA-mediated repression

Mature miRNA sequences repress complementary target sites on messenger RNA (mRNA) transcripts via mRNA degradation and/or translational repression (Figure 1-7) <sup>74,83</sup>. It was thought that, as in animals, the level of complementarity between a plant miRNA and its mRNA target determined the mode of miRNA-mediated repression <sup>84-86</sup>. As miRNAs in plants show a higher level of complementarity to their targets compared to animal miRNAs, it was thought that mRNA cleavage was the dominant mode of miRNA-mediated repression in plants <sup>84</sup>. However, this has since been shown to not be the case as there are several examples in *Arabidopsis* which break this rule; for instance *AP2* has been shown to be repressed by miR172 predominantly via translational inhibition despite high levels of complementarity between miR172 and *AP2* mRNA <sup>61</sup>. A recent paper proposed that the secondary structure of the mRNA target affects the efficacy of miRNA-mediated mRNA cleavage <sup>87</sup>. To date the processes that determine the way in which a particular miRNA represses a particular target are still unclear.



*Figure 1-7: The two mechanisms of miRNA-mediated repression, translational inhibition and mRNA cleavage. miRNAs inhibit translation by binding to target mRNA molecules and physically blocking translation initiation of these mRNAs into protein. miRNAs cleave mRNAs by binding to them as part of miRNA-ARGONAUTE complex; the ARGONAUTE protein cleaves the mRNA, so reducing the abundance of the target mRNA.*

Target mRNAs are degraded (a process also known as slicing) by miRNA-ARGONAUTE (AGO) complexes <sup>84</sup>. Cleavage of the mRNA is carried out by the P-element induced wimpy testis (PIWI) domain of the AGO protein <sup>84,88</sup> and the resulting fragments are degraded by exonucleases <sup>84,89,90</sup>. Transcriptional inhibition is also mediated by the AGO protein amongst others and functions by inhibiting translation initiation, as shown by Iwakawa and Tomari <sup>91</sup>. The effects of mRNA cleavage can be observed via changes to mRNA levels using

transcriptomic techniques and by sequencing the cleavage products using Parallel Analysis of RNA Ends Sequencing (PARE-Seq), while the effects of translational inhibition can currently only be observed by quantifying protein abundance using techniques such as Enzyme Linked Immunosorbent Assay (ELISA) which rely on the availability of high quality, specific antibodies, which are not readily available for plant proteins.

## 1.5 Improving our understanding of wheat inflorescence development will help address food security

A more complete knowledge of the mechanisms behind inflorescence development in wheat, such as the components of the ABCDE model, would allow for a more integrated yield-improvement strategy to be implemented<sup>39</sup>. Yield is a complex trait with multiple components that can broadly be divided into grain number and grain weight<sup>92</sup>; there is often a trade-off between these traits<sup>93</sup>. This trade-off can be overcome<sup>93</sup>, however this requires a more detailed understanding of developmental processes<sup>39</sup>. Genes controlling inflorescence development, such as *AP2*-like genes and *VRT-A2*, and miRNAs such as miR172, provide promising avenues of research for ensuring future food security.

## 1.6 Thesis aims

The overall aim of this thesis is to develop our understanding of how miRNAs affect wheat spike development. I will approach this aim in two ways:

- Expanding our knowledge of the miRNAs that are expressed in the developing wheat spikes
- Developing a nuanced understanding of a gene network involving miR172, including examining any differences in the expression or function of miR172 family members

Specifically, I will use a broad range of experimental approaches, including spatial and non-spatial transcriptomics, transgenics, phenotypic characterisation, and luciferase-based assays, to answer the following questions:

1. What is the role of miRNAs during key developmental transitions in the wheat inflorescence? (Chapter 2)
2. How do the expression patterns of *APETALA2-2* (*AP2-2*) and *APETALA2-5* (*AP2-5*) differ in developing spikes when they are both regulated by miR172? (Chapter 3)
3. Does *VRT-A2* act via miR172 to achieve its phenotypic effect? (Chapter 4)

## 2 The role of microRNAs during key developmental transitions in the wheat inflorescence

### 2.1 Chapter summary

In this chapter, I generated a detailed summary of miRNAs present in the wheat inflorescence during early development. Using smallRNA-Sequencing (sRNA-Seq) and bioinformatic prediction pipelines, I identified 527 candidate miRNAs and predicted which transcripts they target. The pipelines predict known miRNA-mRNA interactions, increasing my confidence in the sensitivity of this analysis. I have identified natural and induced variation in miRNA binding sites and have linked both to phenotypic data. This variation offers promising new avenues for both fundamental developmental biology and wheat breeding targets.

### 2.2 Introduction

#### 2.2.1 There is no comprehensive database for miRNAs in wheat

There are conflicting lists of wheat miRNAs in the literature and databases. miRbase has been referred to as “the centralized registry and database for miRNA annotations across all studied species”<sup>74</sup>. There are 122 *T. aestivum* entries in the most recent miRbase release (22.1) which correspond to 100 miRNA families. This is very unlikely to be a complete list, as there are 222 Arabidopsis miRNA families in miRbase<sup>94</sup>. Also, several miRNAs which have been well-characterised in wheat are absent, such as miR165/166<sup>95</sup> and miR172<sup>32,66</sup>. sRNAanno is a newer miRNA database which includes entries for *T. aestivum*<sup>96</sup>. It contains entries for 616 *MIRNA* loci which are processed into mature sequences from 110 miRNA families. The International Wheat Genome Sequencing Consortium (IWGSC) v1.0 wheat reference genome annotation released in 2018 includes 71,747 unique *MIRNA* loci<sup>15</sup>, however, these correspond to just 111 miRNA families. The mature miRNA sequences released with the IWGSC v1.0 annotation are from just 49 miRNA families<sup>15</sup>. No updated lists of miRNAs in wheat have been released in subsequent IWGSC v1.1 and v2.1 annotations<sup>97,98</sup>. Current annotations of miRNAs in wheat are likely to be incomplete based on evidence from Arabidopsis<sup>94</sup>, and there are inconsistencies between the annotations in different key databases<sup>15,94</sup>.

The entries in these databases, however, provide an excellent starting point for research into the role of miRNAs during development, albeit they clearly tell an incomplete story. In

addition, data on the expression patterns of miRNAs during wheat spike development is limited. There are two studies which have used next generation sequencing (NGS) to quantify miRNA abundance in young wheat spikes<sup>99,100</sup>. Feng, *et al.*<sup>99</sup> sequenced the small RNA of spikes at four stages from W2.0 (double ridge) to W7.5 (which they defined as the “tetrad stage”<sup>99</sup>)<sup>41</sup>. The sequencing was relatively shallow (8-18 million reads per sample) and for one sample, only 26.9% of reads mapped to the wheat genome (compared to > 90% for the other samples). This is suggestive of potential contamination. The second paper, Li, *et al.*<sup>100</sup>, sequenced miRNAs to a higher depth (>34 million reads per sample), however, young spikes from different stages were pooled in the same samples. Therefore, we cannot see how miRNA abundance changes through time from this dataset. Also, although the sequencing depth for these samples was higher, miRNAs from each timepoint would be diluted within the pooled sample. Therefore, only the most abundant miRNAs from each stage or those which are consistently expressed through time would be detected using this method. The sensitivity of miRNA predictions from these pooled samples would be lower than the >34 million reads per sample would suggest.

To fully understand the effect that miRNAs have on wheat spike development, it is important to have a complete list of miRNAs that are present during this period of development and understand their abundance through time.

## 2.2.2 miRNA binding sites are an untapped source of variation in wheat

For continued wheat improvement, we must find or create novel variation within wheat germplasm. miRNAs present an almost completely untapped source of variation which already exists in well-characterised germplasm. A handful of studies have revealed the effects of specific mutations in miRNA binding sites on spike phenotypes.

Debernardi, *et al.*<sup>32</sup> and Greenwood, *et al.*<sup>66</sup> showed how many of the traits of domesticated wheat, such as free-threshing grain and a compact spike, are the result of a SNP in a miR172 binding site in the gene *Q*. Dixon, *et al.*<sup>95</sup> found that an induced SNP in the miR165/166 binding site of *HOMEODOMAIN-2 (HB-2)* conferred a striking paired spikelet phenotype (two spikelets being formed at a single rachis node).

Induced variation in miRNAs and their binding sites has been used to adapt and improve crop species. For example, in rice, overexpression of a miR319-resistant version of *TEOSINTE BRANCHED/ CYCLOIDEA/ PROLIFERATING CELL FACTOR1 (OsTCP21)* conferred resistance to Rice ragged stunt virus (RRSV)<sup>101</sup>. In wheat, clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9 (CRISPR-Cas9) was used to mutate

the miR156 binding site in *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 13* (*TaSPL13*), which led to a decrease in flowering time and an increase in both grain size and number<sup>102</sup>. The potential of modifying miRNAs and their targets has been identified as a promising route to crop improvement in several reviews<sup>103,104</sup>.

### 2.2.3 Modifying miRNA binding sites could help us overcome the issue of redundancy in polyploid crops

As described in Section 1.2, wheat is a young polyploid crop. This presents challenges for crop improvement, such as a large genome size (16 Gbp) and functional redundancy. As a hexaploid species, *T. aestivum* genes are most frequently present in three copies on the A, B and D genomes<sup>23</sup>. It has been shown that 72.5% of syntenic triads have balanced expression, *i.e.* the abundance of the A, B and D homoeologues is approximately equal<sup>22</sup>, providing no clear target for manipulation. Often, there is functional redundancy between homoeologues; when one copy is knocked out, the other homoeologues will compensate<sup>24</sup>. In cases such as these all three copies must be mutated (preferably via knockout mutation) for a phenotypic effect to be observed. Because of these challenges, breeders have historically selected, unconsciously, for dominant gain-of-function mutations as these are more easily identified in phenotypic evaluations<sup>23</sup>. An added advantage is that they can be tracked through breeding pipelines using a single molecular marker. Mutations in miRNA binding sites often (but not always) lead to reduced binding between a miRNA and its target, conferring a gain-of-function phenotype. miRNA binding sites have the potential to reveal a new range of dominant variation, which could be easily integrated into existing breeding pipelines. Also, this variation is likely to cause more subtle effects than constitutive over-expression, as expression of the target gene would still be limited by its endogenous expression domain and transcription rate. This could be beneficial, as there are examples in the literature where subtle increases in expression can be preferable, for example the ectopic expression allele of *VRT-A2* (*VRT-A2b*) confers an increase in TGW with no major developmental defects<sup>56</sup>. In contrast, both loss-of-function *vrt2* and *UBI<sub>pro</sub>:VRT2* (*VRT2* driven by the maize *UBIQUITIN* constitutive overexpression promoter) plants have severe morphological defects such as axillary spikes/spikelets at nodes below the spike in *vrt2* and branching at basal spike nodes in *UBI<sub>pro</sub>:VRT2*. In a study in maize, 13 promoter/intron combinations were tested with the *MADS67* transcription factor to optimise grain yield<sup>105</sup>. Increased and extended *MADS67* expression using a moderate constitutive promoter combined with the *Ubi1* intron (the first intron of the maize *UBIQUITIN* gene) was found to confer the most benefit, while avoiding negative pleiotropic effects, outperforming commonly used strong constitutive promoters<sup>105</sup>.

## 2.2.4 Aims and hypotheses

The aim of this chapter was to generate a comprehensive database of miRNAs which are present during the key stages of early spike development. I used this dataset to identify natural and induced variation in miRNA binding sites and link this variation to phenotypic data.

## 2.3 Methods

### 2.3.1 Small RNA-Seq

$PI^{WT}$  and  $PI^{POL}$  near isogenic line (NIL) seed was germinated as described in Simmonds, *et al.*<sup>106</sup>.  $PI^{WT}$  and  $PI^{POL}$  NILs are described in Adamski, *et al.*<sup>56</sup>. In brief, these lines are *Triticum polonicum* x *T. aestivum* cv ‘Paragon’ BC<sub>6</sub> NILs (Paragon as the recurrent parent) carrying either the wildtype *VRT-A2a* allele ( $PI^{WT}$ ) or the *VRT-A2b* allele ( $PI^{POL}$ ) which leads to increased and extended *VRT-A2* expression across the spike. *T. polonicum*, also known as Polish wheat<sup>56</sup>, is a tetraploid wheat with a distinctive long glume and long spike phenotype<sup>107</sup> which was recently mapped to the *VRT-A2* locus by Adamski, *et al.*<sup>56</sup> and Liu, *et al.*<sup>108</sup>. *T. polonicum* carries the derived *VRT-A2b* allele<sup>56</sup>. I conducted smallRNA-Seq using germplasm containing the *VRT-A2a* and *VRT-A2b* alleles to test hypotheses relating to the effect of increased and extended *VRT-A2* expression on miRNAs. These hypotheses will be discussed in depth in Chapter 4.

I grew plants in 24- or 40-cell trays of “John Innes Cereal Mix” (65% peat, 25% loam, 10% grit, 3 kg/m<sup>3</sup> dolomitic limestone, 1.3 kg/m<sup>3</sup> PG mix, 3 kg/m<sup>3</sup> osmocote extract) in growth chambers at 70% humidity with 16 h photoperiods and 20 °C/16 °C day/night temperatures. Trays contained single genotypes and I arranged the trays randomly in the growth chamber. I dissected primary spikes at Waddington stages W2.5 (double ridge), W4.0 (terminal spikelet), and approximately W5.0 (spike length >0.5 cm, <1.5 cm); approximately 17, 30, and 36 days after sowing, respectively (Figure 2-1)<sup>41</sup>. I dissected all spikes except for sample 0071 which was dissected by Kirsten Lim. I grew plants in batches which were sampled on different dates (Table 2-1).

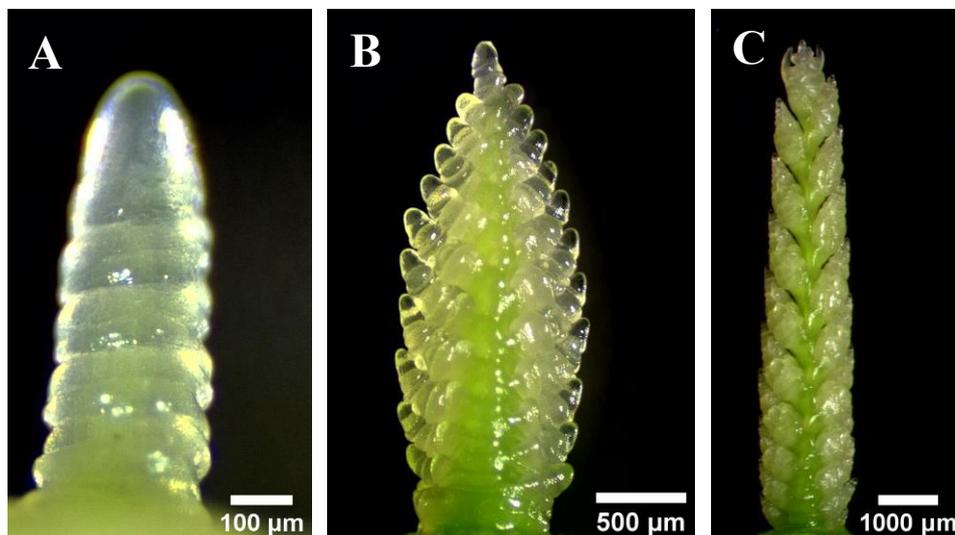
I pooled spike tissue for sequencing: 13 spikes per sample for W2.5, four per sample for W4.0, and two per sample for W5.0. Before and during the dissections I cleaned all surfaces and tools with 70% ethanol (VWR International GmbH) and Blitz RNase Removal Spray (Severn Biotech Ltd). I stored spikes in 2 mL microcentrifuge tubes on dry ice during dissection, followed by flash-freezing in liquid nitrogen and storage at -70 °C. I ground samples using steel ball bearings sterilised with 1 M HCl (Fisher Scientific). I placed two 3 mm tungsten carbide beads (QIAGEN) in each microcentrifuge tube and ground the tissue using a

**Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence**

Geno/Grinder (SPEX SamplePrep) at 1500 RPM for three 30 s intervals. I extracted total RNA from ground samples using the Zymo Directzol RNA Microprep kit (Zymo Research) according to manufacturer's instructions <sup>109</sup>, including the on-column DNase I treatment. Single-end 50 bp sRNA-seq was carried out by Novogene (Cambridge, UK) using the Illumina NovaSeq 6000. Three biological replicates were sequenced per timepoint and genotype combination, except for *PI<sup>WT</sup>* at W5.0 (four biological replicates).

*Table 2-1: Details of Triticum aestivum cv 'Paragon' spike samples used for the sRNA-Seq experiment. Timepoints are described according to the Waddington scale <sup>41</sup>: W2.5 = double ridge, W4.0 = terminal spikelet, W5.0 = carpel extending round three sides of ovule (0.5 cm < spike length < 1.5 cm). The sampling date is the date the spikes were dissected and flash-frozen for each sample ID.*

Timepoint	Sampling date	Sample ID
2.5	02/11/2022	0031, 0036
	04/11/2022	0032
	05/11/2022	0033, 0038
	28/11/2022	0040
4.0	16/11/2022	0041, 0042, 0045, 0046, 0047, 0048
5.0	31/05/2023	0062, 0063, 0068
	05/08/2023	0071, 0072, 0073, 0074



*Figure 2-1: Representative images of Triticum aestivum cv 'Paragon' spikes sampled for the sRNA-Seq experiment. (A) W2.5 (double ridge), (B) W4.0 (terminal spikelet) and (C) approximately W5.0 (carpel extending round three sides of ovule, 0.5 cm < spike length < 1.5 cm).*

## 2.3.2 Small RNA-Seq data analysis

### 2.3.2.1 Data trimming

I used FastQC v0.11.9 to carry out quality control on the raw data <sup>110</sup>. I removed low-quality reads and trimmed adapters using Trimmomatic v0.39 (Table 2-2) with a seed mismatch value of 2, palindrome clip threshold of 30, and simple clip threshold of 10 (Bolger, *et al.* <sup>111</sup>). I used sliding window trimming with a window size of 10 and an average required quality of 20. I set these parameters according to recommendations by Garg and Varshney <sup>112</sup>. I removed poly-A tails and reads outside of the required length (18-34 bp) using Cutadapt v3.4 (Martin <sup>113</sup>). I used FastQC to assess the quality of cleaned reads and test for the expected size distribution for small RNAs with peaks at 21 and 24 nucleotides (Andrews <sup>110</sup>).

*Table 2-2: Adapters used during sRNA-seq trimming with Trimmomatic <sup>111</sup>. The sequences are written 5' to 3'.*

Adapter Name	Sequence
RNA5AdapterRA5	GTTCAGAGTTCTACAGTCCGACGATC
RNA3AdapterRA3	AGATCGGAAGAGCACACGTCT

I also aligned the cleaned reads to the IWGSC v1.0 assembly <sup>15</sup> to quality control the reads, checking for sequencing accuracy and any contamination. I aligned the reads using Bowtie v1.2.2 (Alaux, *et al.* <sup>114</sup>) to an internal Uauy group Bowtie index of the IWGSC v1.0 assembly using the `--best` and `-v 0` options.

### 2.3.2.2 Predicting miRNAs

I used three different pipelines to predict miRNAs from the trimmed sRNA-Seq data.

I used miRador <sup>115</sup> with the trimmed FASTQ files using default parameters (including `RPMThreshold = 3.0`, which has been removed in subsequent releases). Using `RPMThreshold` filters out candidate miRNAs with a low level of expression (< 3 reads per million (RPM)).

I used ShortStack v4.0.3 (Axtell <sup>116</sup>) with BAM files generated from the trimmed data. I aligned the trimmed data to the IWGSC v1.0 assembly <sup>15</sup> using Bowtie v1.2.2 (Langmead, *et al.* <sup>117</sup>) with `-v 0 -best` parameters (maximum zero mismatches, reporting the best singleton alignment). I used samtools v1.15.1 (Li, *et al.* <sup>118</sup>) to convert the SAM files to BAM format, then sort and index the BAM files. I ran ShortStack using these sorted and indexed BAM files. I used the `--known_miRNAs` option with a FASTA file of miRbase

release 22.1 (Kozomara, *et al.* <sup>94</sup>) mature sequences to identify known miRNAs in my dataset. I used the `--dn_mirna` option to also search *de novo* for miRNA loci.

I used miRDeep-P2 v1.1.4 (Kuang, *et al.* <sup>119</sup>) on a FASTA file containing a non-redundant list of reads from all the small RNA-Seq libraries. I generated this file using `cat` to concatenate the trimmed FASTQ files. I converted the concatenated FASTQ file into a FASTA file using `cat $data_dir/combined_reads.fq | paste - - - - | sed 's/^@/>/g' | cut -f 1,2 | tr '\t' '\n' > $data_dir/combined_reads.fa`. I used the FASTX toolkit v0.0.13.2 (Hannon Laboratory <sup>120</sup>) `fastx_collapser` tool to collapse the concatenated reads into unique entries, which were then formatted using `sed -i 's/_/_x/' $data_dir/unique_tags.fa`. I used miRDeep-P2 (with dependencies Perl v5.16.3, ViennaRNA v2.5.1 (Lorenz, *et al.* <sup>121</sup>), and Bowtie v1.2.2 (Langmead, *et al.* <sup>117</sup>)) to predict miRNAs from this collapsed, formatted FASTA file using options `--large-index -L 100`. The Rfam reference FASTA file I used comprised of all *T. aestivum* transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) entries in the Rfam 14.10 release <sup>122</sup>.

### 2.3.2.3 Generating RPM values for candidate miRNAs

I generated an index of candidate miRNAs predicted by all three pipelines described in Section 2.3.2.2 using Bowtie v1.2.2 (Langmead, *et al.* <sup>117</sup>). I then used Bowtie v1.2.2 (Langmead, *et al.* <sup>117</sup>) to align the clean sRNA-Seq reads to the list of candidate miRNAs using the `--best` and `-v 0` options. I used Samtools v1.12 (Li, *et al.* <sup>118</sup>) to convert the SAM alignment files into BAM format, then sort and index the BAM files. I then used the Samtools `idxstats` function <sup>118</sup> to generate counts for each candidate miRNA.

I performed all analysis of count data using R v4.3.2 (R Core Team <sup>123</sup>). I calculated RPM values by multiplying the count value by 1,000,000 and dividing by the total number of reads mapped to the list of candidate miRNAs. I performed principal component analysis (PCA) on the RPM values. I removed all miRNAs which had an RPM variance value of 0 across all samples. I used the `prcomp()` function to perform the PCA.

### 2.3.2.4 Predicting miRNA targets

I used TargetFinder <sup>124-126</sup> to identify transcripts the candidate miRNAs bind to. I used the default options to predict miRNA targets in the IWGSC v1.0 annotation <sup>19</sup>.

## 2.3.3 Identifying variation in TILLING and Watkins lines

I used the bcftools v1.8 (Danecek, *et al.* <sup>127</sup>) `view` command to subset VCF files containing details of SNPs in the EMS-mutagenised and sequenced tetraploid and hexaploid TILLING

populations<sup>31</sup> and the A. E. Watkins hexaploid landrace collection<sup>34</sup>. The VCF files for the TILLING populations were obtained from Ensembl Plants release 59<sup>128</sup>. The VCF file for the Watkins collection was kindly provided by the authors of Cheng, *et al.*<sup>34</sup>.

### 2.3.4 TILLING population phenotyping

Phenotype data associated with the EMS-mutagenised and sequenced hexaploid TILLING population<sup>31</sup> is an internal Uauy group resource generated by James Simmonds.

TGW, grain width and grain length were evaluated in two field experiments in 2013/14 (winter sown) and 2015 (spring sown). The 2013/14 winter trial was sown the John Innes Centre (JIC) Experimental trials site in Bawburgh, UK (approximately 52°37'42.4"N 1°10'13.8"E) and the 2015 spring trial was sown at the JIC Experimental trials site in Bawburgh, UK (approximately 52°37'38.6"N 1°10'44.9"E). For both trials mutant lines were sown as unreplicated 1 m rows. Measurements were made using a MARViN seed analyser (MARViTECH) from all seed from each plot (except for ten primary spikes per line for the 2013/14 trial). Mean thousand grain weight (TGW), grain width and grain length values from the two field trials were used for further analysis.

All other phenotypes (number of rudimentary apical spikelets, number of rudimentary basal spikelets, spikelet density, number of seeds per spike, seeds per spikelet, spike length, seed density, yield per spike, and spikelets per spike) were evaluated during the 2013/14 winter trial described above. Ten primary spikes were collected and phenotyped per line.

I carried out statistical analyses using R v4.3.2<sup>123</sup>. I ranked the mutant lines independently for each phenotype and identified those in the top and bottom 10% of ranks ('outliers'). I used a chi-squared test to evaluate whether lines with SNPs in miRNA binding sites were overrepresented in phenotypic outliers.

### 2.3.5 Watkins population phenotyping

All data were collected as part of a collaboration between the Uauy and Griffiths groups at the JIC, UK before the start of this project. 823 lines from the A. E. Watkins hexaploid landrace collection were grown during the 2019/2020 field season at the JIC Experimental trials site in Bawburgh, UK (approximately 52°37'42.4"N 1°10'13.8"E). Lines were sown as unreplicated 1 m rows using an alpha-lattice design (sparse, partially replicated). Ten primary spikes were harvested per line and five were evaluated for the number of spikelets per spike, the number of infertile apical spikelets per spike, glume length, the number of infertile basal spikelets per spike, spike length, the maximum number of florets per spikelet, grain area, grain circularity, grain length to width ratio, grain length, grain width, and TGW. Best linear unbiased estimates (BLUEs) were calculated by Dr Anna Backhaus for each phenotype using the statgenSTA R

package <sup>129</sup>. Spatial row-column analysis was performed according to the statgenSTA manual <sup>130</sup>.

To test for differences in phenotypes between Watkins collection lines with the reference and alternative nucleotide in SNPs within candidate miRNA binding sites, I performed a one-way analysis of variance (ANOVA) in R v4.3.2 <sup>123</sup> for each phenotype. I generated adjusted *p*-values to correct for multiple testing using `p.adjust()` with the Benjamini-Hochberg procedure <sup>131</sup> (the correction was performed independently for each phenotype).

## 2.4 Results

### 2.4.1 Over 500 mature miRNAs are expressed in wheat spikes during early development

To better understand the role that miRNAs play during wheat spike development, I generated an sRNA-seq dataset from spikes at Waddington stages W2.5 (double ridge), W4.0 (terminal spikelet) and approximately W5.0 (0.5 cm < spike length < 1.5 cm) <sup>41</sup>. The spikes sampled were from *PI<sup>POL</sup>* and *PI<sup>WT</sup>* NILs which are described in detail by Adamski, *et al.* <sup>56</sup>. In brief, both of the NILs are in a *T. aestivum* cv ‘Paragon’ background, however *PI<sup>WT</sup>* contains the ancestral *VRT-A2a* allele, while *PI<sup>POL</sup>* contains the *VRT-A2b* increased and extended expression allele found in *T. polonicum*.

I extracted total RNA from pooled spikes (three or four biological replicates per genotype and timepoint combination) and small RNA (18-34 bp) was sequenced at a depth of >50 million 50 bp single end reads per sample (Table 2-3). This is higher than other sRNA-Seq experiments in wheat which generally sequence to a depth of 10-20 million reads per sample <sup>99,132-134</sup>. Across samples, over 83% of reads aligned perfectly to the IWGSC v1.0 assembly, suggesting the sequencing was accurate and the RNA was relatively free from contamination; the literature suggests an alignment rate of over 70% is indicative of accurate sequencing with low levels of contaminating DNA <sup>135</sup>. This alignment rate is also comparable to other high-quality datasets <sup>43</sup>, particularly considering the high level of stringency used during this alignment (zero mismatches allowed). Reads may not align to the reference assembly due to sequencing errors, or due to variation between the NIL parental lines (*Triticum aestivum* cv ‘Paragon’ and *T. polonicum*) and *T. aestivum* cv ‘Chinese Spring’, which the IWGSC v1.0 reference assembly is based on <sup>15</sup>.

Table 2-3: Summary statistics for the sRNA-Seq dataset generated from *Triticum aestivum* cv ‘Paragon’ spikes. Timepoints are given according to the Waddington scale <sup>41</sup>. W2.5 = double ridge, W4.0 = terminal spikelet, W5.0 = carpel extending round three sides of ovule (0.5 cm < spike length < 1.5 cm). Genotypes are near-isogenic lines described in Adamski, et al. <sup>56</sup>; in brief  $PI^{WT}$  contains the ancestral VEGETATIVE TO REPRODUCTIVE TRANSITION-A2a (VRT-A2a) allele, while  $PI^{POL}$  contains the VRT-A2b increased and extended expression allele. Trimming involved removing poor quality reads, adapters, poly-A tails and reads >34 bp or <18 bp using Cutadapt <sup>113</sup> and Trimmomatic <sup>111</sup>. Alignment against miRNAs was carried out using Bowtie <sup>117</sup> against a list of 527 putative miRNAs generated by ShortStack <sup>116</sup>, miRDeep-P2 <sup>119</sup> and miRador <sup>115</sup> from this data. Alignment against the IWGSC v1.0 assembly <sup>15</sup> was carried out using Bowtie <sup>117</sup>.

Sample ID	Genotype	Timepoint	Total raw reads	Reads passed trimming	Cleaned reads aligned to IWGSC v1.0 (%)	Cleaned reads aligned to miRNAs (%)
0031	$PI^{WT}$	2.5	93,151,277	80,555,074	93.42	1.26
0032	$PI^{WT}$	2.5	51,831,909	47,027,220	94.72	0.50
0033	$PI^{WT}$	2.5	56,859,068	51,702,145	94.38	0.55
0036	$PI^{POL}$	2.5	92,921,266	64,149,762	93.18	1.58
0038	$PI^{POL}$	2.5	54,543,913	51,142,702	86.78	0.61
0040	$PI^{POL}$	2.5	51,888,351	47,352,976	93.41	0.43
0041	$PI^{WT}$	4.0	145,713,034	101,315,174	92.58	0.92
0042	$PI^{WT}$	4.0	54,477,110	45,915,250	91.79	1.06
0045	$PI^{WT}$	4.0	59,903,628	52,061,735	92.49	0.64
0046	$PI^{POL}$	4.0	102,251,969	63,200,583	91.84	1.42
0047	$PI^{POL}$	4.0	50,169,386	42,352,786	91.88	0.88
0048	$PI^{POL}$	4.0	65,007,806	57,180,398	93.00	0.78
0062	$PI^{WT}$	5.0	61,612,777	51,096,110	86.48	1.70
0063	$PI^{WT}$	5.0	68,156,221	47,620,339	83.62	1.18
0068	$PI^{POL}$	5.0	85,865,526	52,380,380	89.05	1.64

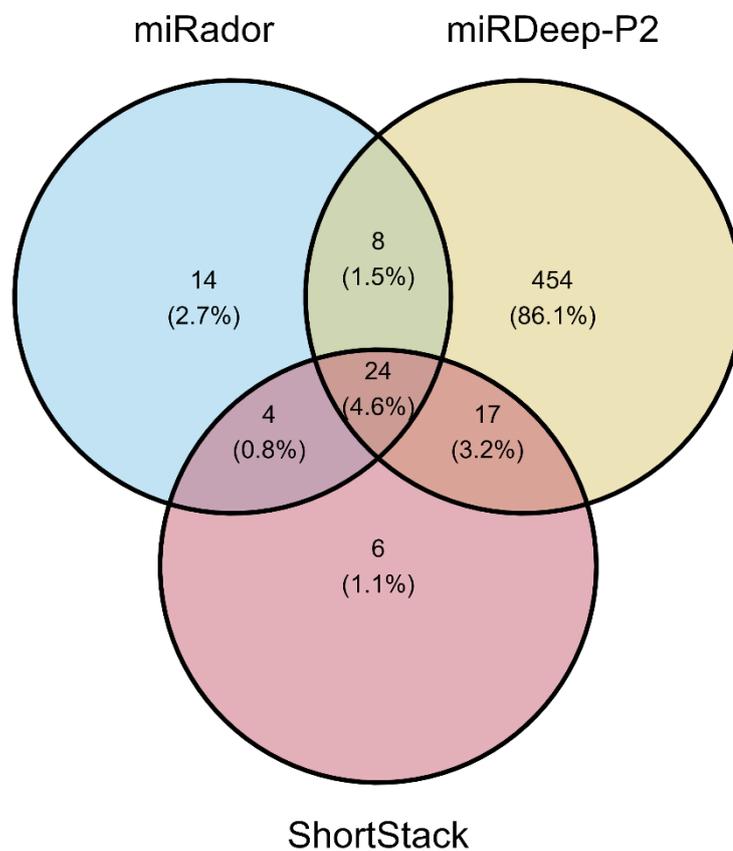
**Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence**

0071	<i>PI<sup>WT</sup></i>	5.0	65,207,217	53,431,174	86.40	2.06
0072	<i>PI<sup>WT</sup></i>	5.0	53,569,030	44,623,233	87.25	1.79
0073	<i>PI<sup>POL</sup></i>	5.0	73,821,421	51,159,083	88.28	2.10
0074	<i>PI<sup>POL</sup></i>	5.0	58,433,845	39,786,024	87.96	1.95

---

I used this small RNA-Seq dataset as input for three different miRNA prediction pipelines: ShortStack <sup>116</sup>, miRDeep-P2 <sup>119</sup> and miRador <sup>115</sup>. I combined the results from all three pipelines to generate a list of 2,182 putative *MIRNA* loci which encode 527 unique mature miRNAs.

The three pipelines predicted different mature miRNAs (Figure 2-2). miRDeep-P2 predicted 503 mature miRNAs, while miRador and ShortStack predicted 50 and 51, respectively. It should be noted that some of these mature miRNAs are one nucleotide longer or shorter than each other so are likely to be the result of different pipeline methodologies predicting the same miRNA in different ways. miRador and ShortStack predictions were relatively similar with 38.3% of their predictions being shared. miRDeep-P2 generated a high number of unique predictions: 86.1% of its predictions were not predicted by either miRador or ShortStack.



*Figure 2-2: Mature miRNAs predicted by three prediction pipelines based on small RNA-Seq data generated from *T. aestivum* cv Paragon spikes from Waddington scale 2.5-5.0 (double ridge to carpel extending round three sides of ovule) <sup>41</sup>. The pipelines used were ShortStack <sup>116</sup>, miRador <sup>115</sup> and miRDeep-P2 <sup>119</sup>.*

miRador and ShortStack assign predicted miRNAs to known miRNAs from miRBase. Unfortunately, miRDeep-P2 does not offer this function, therefore the 454 candidate miRNAs uniquely predicted by miRDeep-P2 were not assigned to a miRNA family. I therefore used the

*Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence*

nucleotide basic local alignment search tool (BLASTn) algorithm to search for closely related miRNAs in the miRBase database and used the top hit for each miRNA candidate with an expectation value (E-value) cutoff of 1.0 to assign the candidate sequences to miRNA families. According to this analysis, the combined results of all three pipelines identified members of at least 143 miRNA families present in wheat spikes, which is a similar number in the miRbase<sup>94</sup> and sRNAanno<sup>96</sup> databases and the IWGSC v1.0 annotation<sup>15</sup>. This analysis only includes miRNAs expressed in wheat spikes during the early stages of development (W2.5 – W5.0<sup>41</sup>). Also, 145 of the candidate miRNAs were not assigned to a family as they were not assigned to one by ShortStack or miRador and they did not produce a hit within the miRBase database with an E-value <1.0. Some of these unassigned candidates may be false positives, however, some were predicted by multiple pipelines, for example miRNA candidate-120 was predicted by ShortStack, miRDeep-P2 and miRador. The number of miRNA families identified suggests that this dataset represents a more comprehensive overview of miRNAs present in the wheat spike during early development.

The family designations made using BLAST, although a useful indicator, should be treated with some scepticism. The family-level miRNA annotation by ShortStack is based on identifying known miRNAs in the dataset in the first instance, so are very high-confidence. The miRador pipeline uses BLAST to search for candidate miRNA precursor sequences in the miRBase database. As it uses the precursor sequence (as opposed to the mature miRNA sequence I have used for my BLAST search), it is more likely to produce true positive hits against known miRNAs. In summary, miRNA family annotations from ShortStack and miRador can be treated with a higher level of confidence than the mature sequence BLAST search I carried out.

I aligned the sRNA-Seq reads to this new list of putative miRNAs and found that between 0.5 and 2.1% of clean reads aligned to miRNAs (Table 2-3). I carried out PCA on this expression data to visualise the variance between samples (Figure 2-3).

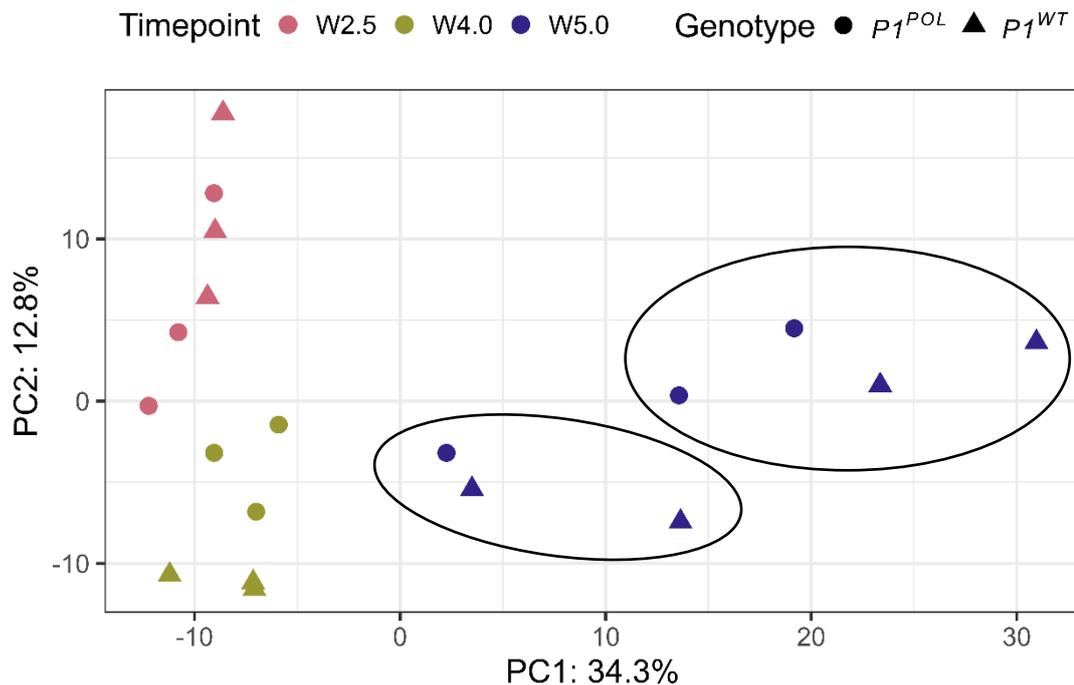


Figure 2-3: Principal Component Analysis (PCA) of small RNA-Seq data generated from *Triticum aestivum* cv ‘Paragon’ spikes. Each point represents a different biological replicate. The timepoint of each biological replicate is defined in the legend according to the Waddington scale<sup>41</sup> and differentiated in the PCA plot by colour; W2.5 (double ridge) = pink, W4.0 (terminal spikelet) = green, W5.0 (carpel extending round three sides of ovule, 0.5 cm < spike length < 1.5 cm) = blue. The genotype of each biological replicate is also defined in the legend and differentiated in the PCA plot by point shape; circle =  $P1^{POL}$ , triangle =  $P1^{WT}$ . Genotypes are near-isogenic lines described in Adamski, et al.<sup>56</sup>; in brief  $P1^{WT}$  contains the ancestral VEGETATIVE TO REPRODUCTIVE TRANSITION-A2a (VRT-A2a) allele, while  $P1^{POL}$  contains the VRT-A2b increased and extended expression allele. Percentages on x- and y-axes refer to the percentage of variance each Principal Component (PC) describes. Black circles denote W5.0 samples with different sampling dates.

The first two PCs described 47.1% of variance observed between the samples. Clustering was observed according to timepoint, indicating that samples of the same timepoint were more similar to each other than samples at different timepoints, suggesting that spike staging was accurate. There was some clustering observed according to the sampling date. The W5.0 samples clustered according to sampling date (black circles in Figure 2-3, Table 2-1).

## 2.4.2 3,535 transcripts contain sequences which may be bound by the 527 predicted mature miRNAs

I used TargetFinder<sup>124-126</sup> to predict transcript targets for this novel list of miRNAs within the IWGSC v1.0 genome annotation<sup>15</sup>. TargetFinder output contains transcript-relative target loci (the target location is given as a position within the mRNA sequence). To facilitate further analysis (particularly to identify variation in these target sites), I created a script to translate transcript-relative locations to genome-relative locations (Section A.6, available at [https://github.com/Uauy-Lab/transcript\\_to\\_genomic\\_loci](https://github.com/Uauy-Lab/transcript_to_genomic_loci)). As far as I am aware, a tool similar to this which can be applied to any genome does not currently exist.

The number of predicted targets per miRNA varied from zero to 136, with each miRNA targeting 7 transcripts on average (median).

### 2.4.2.1 This pipeline identified known miRNA-mRNA interactions

To test the sensitivity of the database I had generated I looked for miRNAs and miRNA-mRNA interactions which have been well characterised in wheat.

#### 2.4.2.1.1 miR156 targets SPL genes

Nineteen *MIR156* loci were identified in this analysis, which produce two mature miR156 sequences. The first, miRNA candidate-351, was found in a BLASTn<sup>136</sup> search against the miRbase mature miRNA database<sup>94</sup> to be most similar to the *Asparagus officinalis* miRNA aof-miR156a (accession: MIMAT0049765). The second, miRNA candidate-403, was most similar to the *Arabidopsis lyrata* miRNA aly-miR156h-3p (accession: MIMAT0017412). The sequences of miRNA candidate-403 and miRNA candidate-351 are substantially different, and given the results of the BLASTn search, it is likely that they are the 3p and 5p sequences of miR156, respectively. This is supported by the sequence of miRNA candidate-351 being identical to previously reported miR156 sequences in wheat<sup>137</sup>.

miRNA candidate-351 is predicted by TargetFinder to target 14 transcripts from 11 genes in the IWGSC v1.0 annotation<sup>15</sup>. These include nine *SPL* genes: *TaSPL14-A1* (TraesCS7A01G246500), *TaSPL14-D1* (TraesCS7D01G245200), *TaSPL17-D1* (TraesCS5D01G273900), *TaSPL16-A1* (TraesCS7A01G260500), *TaSPL16-D1* (TraesCS7D01G261500), *TaSPL18-B1* (TraesCS5B01G286000), *TaSPL18-D1* (TraesCS5D01G294400), *TaSPL4-A1* (TraesCS6A01G155300), and *TaSPL3-A1* (TraesCS6A01G110100) (Figure 2-4)<sup>102</sup>. The miRNA candidate-351 binding sites are all within the last exon of these *TaSPL* genes. This result is consistent with studies which have shown experimentally that miR156 targets *SPL* genes in wheat<sup>137,138</sup>. miRNA candidate-351 is also predicted to target TraesCS2B01G029900 and TraesCS4A01G341400, although these



The identification of known interactions in our data provides us with confidence that this analysis is producing true positive predictions from the dataset and that it has sufficient sensitivity to identify key known interactions and potentially novel interactions.

### 2.4.3 Variation in miRNA target sites

Variation in miRNA target sites presents a novel source of material for testing questions in developmental biology and for integration into breeding pipelines. I will focus on two types of variation in this section:

- Induced variation in EMS-mutagenised and sequenced tetraploid and hexaploid TILLING populations <sup>31</sup>
- Natural variation in the A. E. Watkins hexaploid landrace collection <sup>34</sup>

By utilising different populations, I hoped to identify the broadest possible range of variation in miRNA binding sites.

#### 2.4.3.1 *There is induced variation in miRNA target sites in wheat TILLING populations*

I analysed the wheat TILLING populations <sup>31</sup> in both *T. aestivum* cv ‘Cadenza’ and *T. turgidum* cv ‘Kronos’ backgrounds to identify lines with variation in the 3,535 putative miRNA binding sites predicted in Section 2.4.2. There are 2,590 and 1,306 variants in miRNA binding sites in the Cadenza and Kronos TILLING populations, respectively. Due to the polyploid nature of wheat, loss-of-function mutations in one gene are often phenotypically masked by its homoeologues <sup>31</sup>. As a result, mutant lines must often be crossed to generate a triple null mutant line and reveal the full phenotypic effect. Variation in miRNA binding sites may lead to better miRNA binding (decreasing target gene expression) or less effective miRNA binding (increasing target gene expression). SNPs which decrease the ability of a miRNA to bind to a gene are likely to have dominant phenotypic effects <sup>32,66,95</sup>, reducing the need for lengthy crossing schemes and enabling their immediate use in breeding.

47.8% of all Cadenza TILLING lines contain variation in miRNA binding sites. I linked this data with phenotypic data which was collected by James Simmonds. I first wanted to identify broad patterns and try to detect whether extreme phenotypic values (*i.e.*, those in the top or bottom 10% of values by rank) were enriched for mutant lines with variation in miRNA binding sites. I analysed the following phenotypes: the number of basal and apical rudimentary basal spikelets, spikelet density, the number of seeds per spike, the number of seeds per spikelet, spike length, seed density, the number of spikelets per spike, thousand grain weight (TGW), mean grain width and length, and yield per spike. In brief, I ranked the TILLING lines for each phenotype and assigned the top and bottom 10% of lines as ‘outliers’ for that phenotype (there were no lower outliers for the rudimentary basal and apical spikelet

**Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence**

phenotypes as a large number of lines had no rudimentary basal or apical spikelets) (Figure 2-5 and Figure 2-6). Lines with variation in miRNA binding sites are not significantly enriched in ‘outliers’ for any phenotype (chi-square test: spike length  $X^2(1) = 0.13$ ,  $p = 0.72$ ; spikelets per spike  $X^2(1) = 1.57$ ,  $p = 0.21$ ; number of rudimentary basal spikelets  $X^2(1) = 0.27$ ,  $p = 0.60$ ; number of rudimentary apical spikelets  $X^2(1) = 0.01$ ,  $p = 0.91$ ; seeds per spikelet  $X^2(1) = 0.42$ ,  $p = 0.52$ ; yield per spike  $X^2(1) = 0.66$ ,  $p = 0.42$ ; seeds per spikelet  $X^2(1) = 0.25$ ;  $p = 0.62$ ; seed density  $X^2(1) = 0.66$ ,  $p = 0.42$ ; spikelet density  $X^2(1) = 1.19$ ;  $p = 0.28$ ; TGW  $X^2(1) = 1.27$ ,  $p = 0.26$ ; grain width  $X^2(1) = 0.41$ ,  $p = 0.52$ ; grain length  $X^2(1) = 0.03$ ;  $p = 0.87$ ).

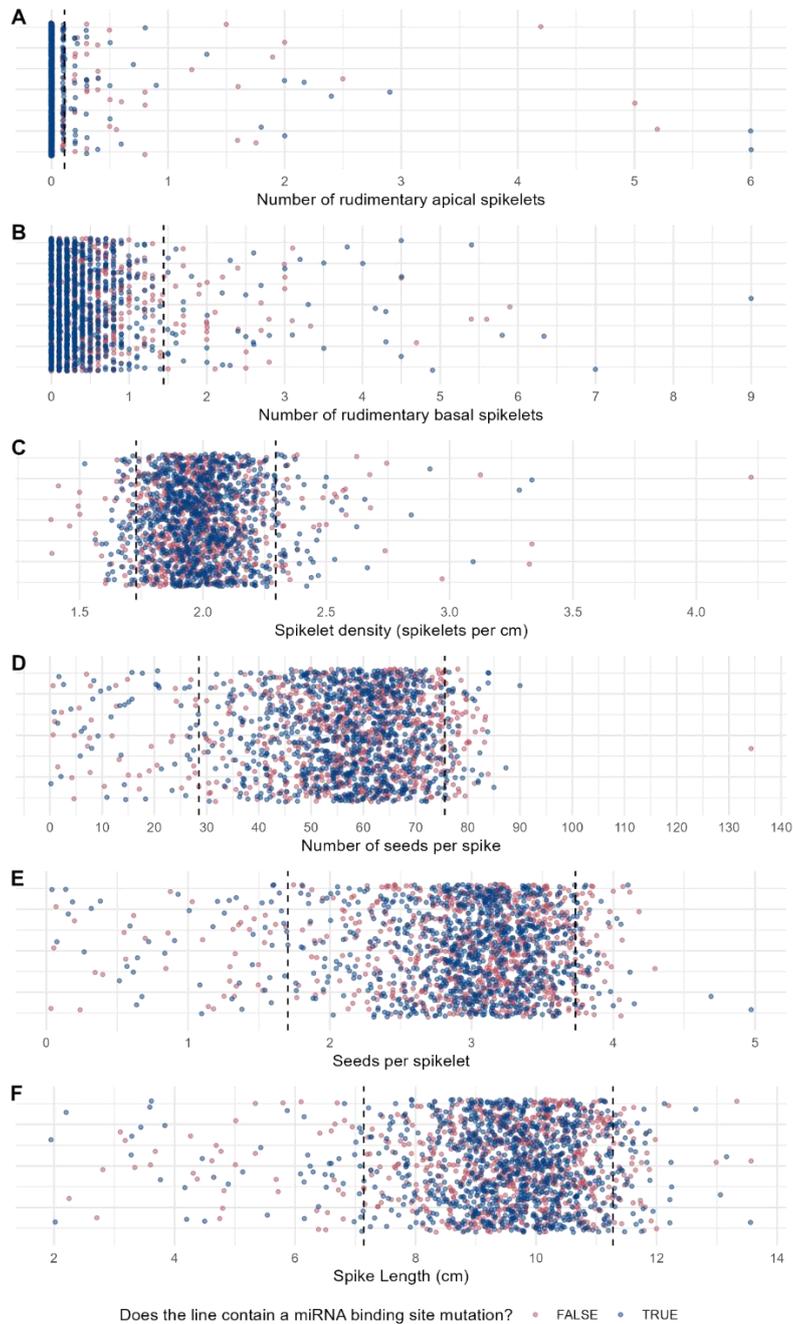


Figure 2-5: Phenotypic data of *Triticum aestivum* cv 'Cadenza' TILLING population<sup>31</sup> for (A) number of rudimentary apical spikelets, (B) number of rudimentary basal spikelets, (C) number of spikelets per cm, (D) number of seeds per spike, (E) number of seeds per spikelet and (F) spike length. All data points are average values across 10 primary spike samples. Blue data points correspond to TILLING lines which contain at least one SNP within a predicted miRNA binding site. Pink data points correspond to TILLING lines which do not contain any SNPs in predicted miRNA binding sites. Vertical dotted lines represent the threshold for the top (A-F) and bottom (C-F) 10% of data points by rank.

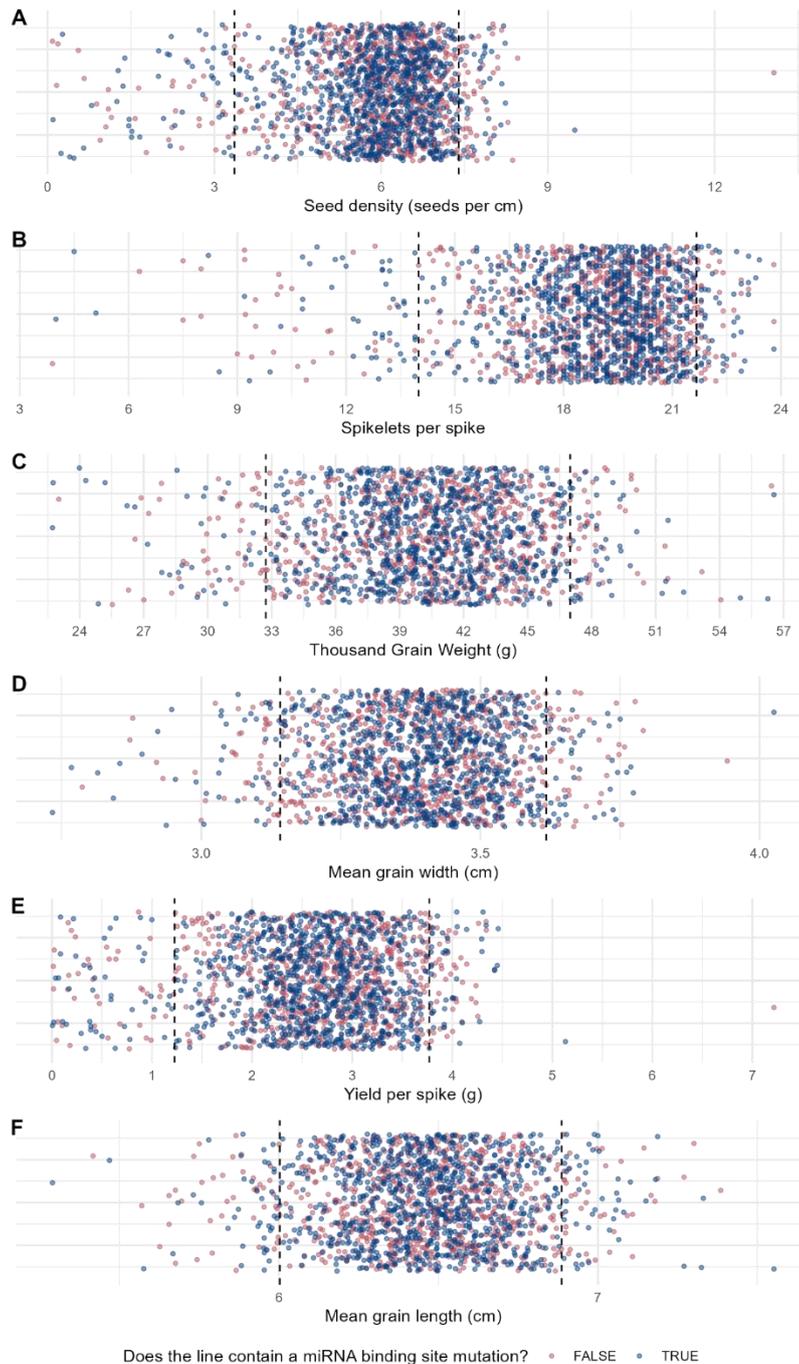


Figure 2-6: Phenotypic data of *Triticum aestivum* cv 'Cadenza' TILLING population<sup>31</sup> for (A) number of seeds per cm, (B) number of spikelets per spike, (C) thousand grain weight, (D) grain width, (E) yield per spike and (F) grain length. Data points for (A, B and E) are average values across 10 primary spike samples. Data points for (C, D and F) are average values from unreplicated 1 m rows. Blue data points correspond to TILLING lines which contain at least one SNP within a predicted miRNA binding site. Pink data points correspond to TILLING lines which do not contain any SNPs in predicted miRNA binding sites. Vertical dotted lines represent the threshold for the top and bottom 10% of data points by rank.

**2.4.3.2** *This dataset can be used to identify useful induced variation in miRNA target sites*

To take advantage of this novel dataset, I identified TILLING lines<sup>31</sup> with variation in miRNA binding sites in genes we know to be involved in wheat spike development (Figure 2-7). *SPL* genes are known to be targeted by miR156 and as shown in Section 2.4.2.1.1. There are two TILLING lines with SNPs in the *TaSPL4-A1* (TraesCS6A01G155300) miR156 binding site (Figure 2-7A). Kronos2942 has a synonymous SNP at position 4 (from the miRNA 5' end) of the binding site which results in an additional mismatch between *TaSPL4-A1* and miR156. Cadenza0670 has a missense SNP at position 9 of the binding site which also generates an additional mismatch. Both SNPs are located in the miRNA seed region (defined as positions 2-12 from the miRNA 5' end<sup>141-143</sup>), increasing the likelihood of them having a significant effect on miR156 regulation of *TaSPL4-A1*. These lines are currently being grown for phenotyping. The rice orthologue of *TaSPL4-A1*, *OsSPL11*, has been shown to positively regulate grain length and TGW<sup>144</sup>. However, in the phenotyping data described in Section 2.4.3.1, the length of Cadenza0670 grains was in fact lower than the wildtype (6.3 cm and 6.7 cm, respectively). The TGW was also lower in Cadenza0670 (39.7 g, compared to 45.7 g for the wildtype). TILLING lines contain, on average, 5,351 mutations per hexaploid line<sup>31</sup>. Therefore, it would be important to backcross both candidate TILLING lines to accurately quantify any effect these variants have on spike phenotypes.

There is also variation in the seed region of the *TaSPL14-A1* (TraesCS7A01G246500) miR156 binding site within the TILLING populations (Figure 2-7B). Cadenza1807 and Kronos3819 have an identical missense mutation at position 11 which confers an additional mismatch. *TaSPL14-A1* has been shown to positively regulate grain number and size<sup>145</sup>. Cadenza1807 had a lower number of seeds per spike than the WT in the field trials described in this Chapter (60.6, compared to 81.2 for the wildtype), but a higher TGW of 50.1 g compared to 45.7 g for the wildtype. These lines are also being grown for phenotyping.

**Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence**

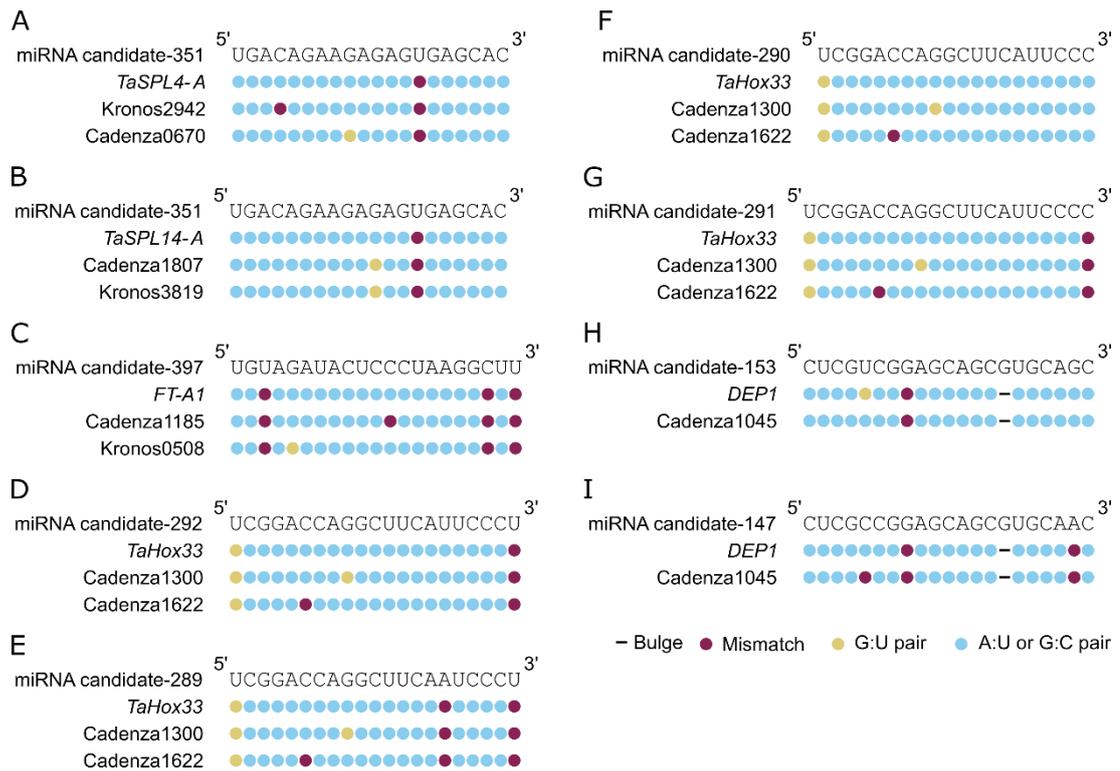


Figure 2-7: TILLING lines<sup>31</sup> with variation in miRNA binding sites in both tetraploid *Triticum turgidum* cv ‘Kronos’ and hexaploid *T. aestivum* cv ‘Cadenza’ backgrounds. The complementarity between miRNA candidates, their predicted wildtype targets and the EMS-induced mutant alleles of these targets are shown using a dot diagram. The 5’ to 3’ miRNA candidate sequence is shown at the top of each figure. Each coloured dot represents a binding site nucleotide in putative target sites, from 3’ to 5’. Each coloured dot represents a binding site nucleotide. Blue dots represent a perfect Watson-Crick base pair (A:U or G:C), gold dots represent a G:U pair, and red dots represent a mismatch. Dashes represent a gap in the alignment which would form a bulge during miRNA-mRNA binding. (A) miRNA candidate-351 binding to wildtype *TaSPL4-A* (*TaSQUAMOSA PROMOTER BINDING-LIKE 4-A*) and TILLING lines with variation in the binding site, (B) miRNA candidate-351 binding to wildtype *TaSPL14-A* (*TaSQUAMOSA PROMOTER BINDING-LIKE 14-A*) and TILLING lines with variation in the binding site, (C) miRNA candidate-397 binding to wildtype *FT-A1* (*FLOWERING LOCUS T-A1*) and TILLING lines with variation in the binding site, (D) miRNA candidate-292 binding to wildtype *TaHox33* and TILLING lines with variation in the binding site, (E) miRNA candidate-289 binding to wildtype *TaHox33* and TILLING lines with variation in the binding site, (F) miRNA candidate-290 binding to wildtype *TaHox33* and TILLING lines with variation in the binding site, (G) miRNA candidate-291 binding to wildtype *TaHox33* and TILLING lines with variation in the binding site, miRNA candidate-153 binding to wildtype *DEP1* (*DENSE AND ERECT PANICLE 1*) and a TILLING line with

**Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence**

variation in the binding site, and (I) miRNA candidate-147 binding to wildtype *DEP1* and a *TILLING* line with variation in the binding site.

*FLOWERING LOCUS T (FT)* genes are central regulators of flowering time<sup>146</sup>. It has been shown that *FT* genes are targeted by miR5200 in *Brachypodium distachyon*<sup>147</sup> and this interaction was predicted in my analysis in wheat. There are two *TILLING* lines with variation in the seed region of the miR5200 (candidate-397) binding site in *FT-A1* (TraesCS7A01G115400) (Figure 2-7C). Cadenza1185 contains a synonymous SNP at position 12 which confers an additional mismatch to miR5200, while Kronos0508 has a missense SNP at position 5 which also reduces the complementarity between miR5200 and *FT-A1*. These lines are currently being grown under long day (16 h light 20 °C, 8 h dark 16 °C) and short day (10 h light 20 °C, 14 h dark 16 °C) conditions and flowering time will be quantified for mutant and control lines. *FTI* has been shown to accelerate flowering in wheat<sup>146</sup>, so I would expect the flowering time of these lines may be altered, and that any differences may depend on the *Photoperiod-1* allele of the parent line<sup>148</sup> and on whether the plants are grown under short day or long day conditions<sup>149</sup>.

In parallel to my work, PhD student Isabel Faci identified candidate genes which may contribute to the control of branching during wheat spike development. Two of these genes, *TaHOX33* (TraesCS5D01G052300) and *DENSE AND ERECT PANICLE 1 (DEP1)* (TraesCS5D01G216900), could also be targeted by miRNAs based on my analysis. *TaHOX33* is predicted to be targeted by candidate-292, 289, 290 and 291 which are all members of the miR165/166 family. The best BLASTn hit for candidate-289 against the miRBase database was ata-miR5168-3p, however, despite it being designated as a member of the miR165 family by miRador. As outlined in Section 2.4.1, the miRador and ShortStack annotations should be treated with a higher confidence level than the top BLASTn hit. The interaction between miR165/166 and *TaHOX33* has been shown previously<sup>150</sup>. There are two *TILLING* lines with variation in the seed region of the miR165/166 binding site in *TaHOX33* (Figure 2-7D-G). Cadenza1300 has a missense/splice region variant SNP at position 9 and Cadenza1622 contains a missense SNP at position 6 which both reduce the complementarity between miR165/166 and *TaHOX33*. A miR165/166-resistant copy of *OsHOX33* has been overexpressed in rice, resulting in phenotypes such as rolled and radialised leaves (where there is a loss of either abaxial or adaxial identity<sup>151</sup>). I expect the Cadenza1300 and Cadenza1622 mutant lines would show a similar, although perhaps more subtle phenotype as the miR165/166-resistant *TaHOX33* allele is not overexpressed.

*DEP1* is predicted to be targeted by candidate-153 and 147 which are all members of the miR531 family. Cadenza1045 contains a missense SNP at position 5 which increases

complementarity to candidate-153 and reduces complementarity to candidate-147 (Figure 2-7H-I). Candidate-147 is more highly expressed at all timepoints than candidate-153 (>43 RPM in all samples, compared to <21 RPM in all samples). Therefore, I would predict that the SNPs effect of reducing complementarity to candidate-147 would have a greater impact than its effect on increasing complementarity to candidate-153. *OsDEP1* is a major rice grain yield QTL<sup>152</sup>. An allele which creates a premature stop codon, *depl*, enhances meristematic activity, suggesting that the *DEP1* gene represses the development of inflorescence meristems<sup>152</sup>. The same study which characterised *OsDEP1* generated transgenic wheat containing a *pUbi:RNAi-TaDEP1* construct which knocked-down the expression of *TaDEP1*<sup>152</sup>. The presence of this construct resulted in a lax ear with a lower number of spikelets<sup>152</sup>. Therefore, I hypothesise that Cadenza1045 would have a more compact ear with a higher number of spikelets per spike. Cadenza1045 is not an outlier for spikelets per spike as described in Section 2.4.3.1. The lowest value for upper outliers is 21.67, so it is just below the cutoff. Cadenza1045 has an above average number of spikelets per spike with 21.40; the average across all Cadenza TILLING lines is 18.79 and Cadenza wildtype has 21.29 spikelets per spike. This initial phenotyping data suggests that the Cadenza1045 SNP in the miR531 binding site may have a positive effect on the number of spikelets per spike which is consistent with studies in rice<sup>152</sup>.

Cadenza1045, Cadenza1300 and Cadenza1622 are currently being grown to be phenotyped for spike architecture traits. These lines with variation in miRNA binding sites promise exciting new avenues for modifying key traits such as spike architecture which can be deployed to accelerate crop improvement.

#### *2.4.3.3 Natural variation in miRNA target sites is correlated with spike phenotypes*

To test whether variation in miRNA target sites correlates with phenotypic changes, I combined recently published whole-genome sequencing data of the A. E. Watkins hexaploid landrace collection<sup>34</sup> with multi-trait phenotypic data from a field trial conducted in 2020 as a collaboration between the Uauy and Griffiths groups at the JIC, UK. I tested for significant phenotypic differences between lines with either the reference or alternate SNP in miRNA target sites. I have identified 187 promising SNPs in miRNA target sites which significantly correlate with at least one phenotype in the Watkins collection (Table 2-4, Table S3).

I have begun to investigate the potential of two SNPs which correlate with TGW (Table 2-4). A SNP at position chr2B:89,555,683 is within the gene *VACUOLAR SORTING RECEPTOR HOMOLOG 1-B1 (TaVSRI-B1)* (TraesCS2B01G122400) and lines with the alternative nucleotide at this position have a significantly higher TGW compared to lines with the reference nucleotide (one-way ANOVA:  $F(1, 669) = 70.40$ , adjusted  $p < 0.001$ ) (Figure 2-8A).

TraesCS2B01G122400 is predicted by TargetFinder to be targeted by miRNA candidate-125, which has most sequence similarity to the stem-loop sequence of *tae-MIR5384* in the miRBase database<sup>94</sup> and the two miR5384 sequences from *T. urartu* in the sRNAanno database (Tur-miR5384b-Known and Tur-miR5384a-Known)<sup>96</sup>. A study investigating miRNAs which are expressed in the wheat grain found that miR5384 had relatively high expression during the middle stages of grain development and, interestingly, that it was differentially expressed in high and low nitrogen conditions<sup>153</sup>. One paper has been published on the gene *TaVSR1-B1*<sup>154</sup>, which showed that increasing *TaVSR1-B1* confers deeper roots with a higher dry weight, although there was no significant difference in TGW between the wildtype and lines with higher *TaVSR1-B1* expression<sup>154</sup>. The increase in *TaVSR1-B1* expression in this paper was conferred by the absence of a miniature inverted-repeat transposable element (MITE) which is present in the promoter regions for some haplotypes, therefore the changes to *TaVSR1-B1* expression would likely be different between the lines used in this paper and those with the SNP I identified in my analysis. No published studies have investigated the role of *TaVSR1-B1* in developing grains, however data from the Azhurnaya developmental gene expression atlas shows that *TaVSR1-B1* is expressed >5 TPM in anthers at anthesis, grain at milk grain and soft dough stages, and the endosperm at dough stage<sup>22,155</sup>. The *AtVSR1* gene in *Arabidopsis* has been shown to sort storage proteins in seeds, and the seeds of *stvsr1* loss-of-function plants had smaller protein storage vacuoles than wildtype seeds<sup>156</sup>.

The synonymous SNP in *TaVSR1-B1* is located in position 19 of the binding site (from the 5' end of the miRNA, outside the putative seed region<sup>141-143</sup>). The 288 lines with a G in this position, which confers a Watson-Crick base pair at this position, had a mean TGW of 43.8 g, compared to the 383 lines with an A at this position which confers a mismatch and had a mean TGW of 48.2 g, a 10.0% increase. The modern varieties with genome assemblies hosted by Ensembl Plants have a mixture of alleles at this position, suggesting that the alternative allele which is correlated with an increase in TGW may not be present in some elite varieties. ArinaLrFor, Jagger, Julius, and Mace all contain an adenine at this position, while Lancer, Landmark, Norin61, and Stanley contain a guanine.

**Chapter 2: The role of microRNAs during key developmental transitions in the wheat inflorescence**

Table 2-4: SNPs in the A. E. Watkins collection located within putative miRNA binding sites which have statistically significant associations with phenotypic differences according to a one-way ANOVA. *p*-values have been adjusted using the Benjamini-Hochberg procedure. SNP locations are given from the IWGSC v1.0 annotation. The full dataset can be found in Supplementary Table S3.

SNP location	<i>F</i> value	Adjusted <i>p</i> value	Phenotype
chr2B:89555683	11.79	0.01	Glume length
	19.22	< 0.001	Spike length
	33.60	< 0.001	Maximum number of florets per spikelet
	53.27	< 0.001	Grain area
	30.43	< 0.001	Grain length
	43.54	< 0.001	Grain width
	70.40	< 0.001	TGW
	chr7B:526865360	16.37	0.003
7.59		0.04	Infertile basal spikelets
17.90		< 0.001	Maximum number of florets per spikelet
74.55		< 0.001	Grain area
55.46		< 0.001	Grain length
46.08		< 0.001	Grain width
64.03		< 0.001	TGW

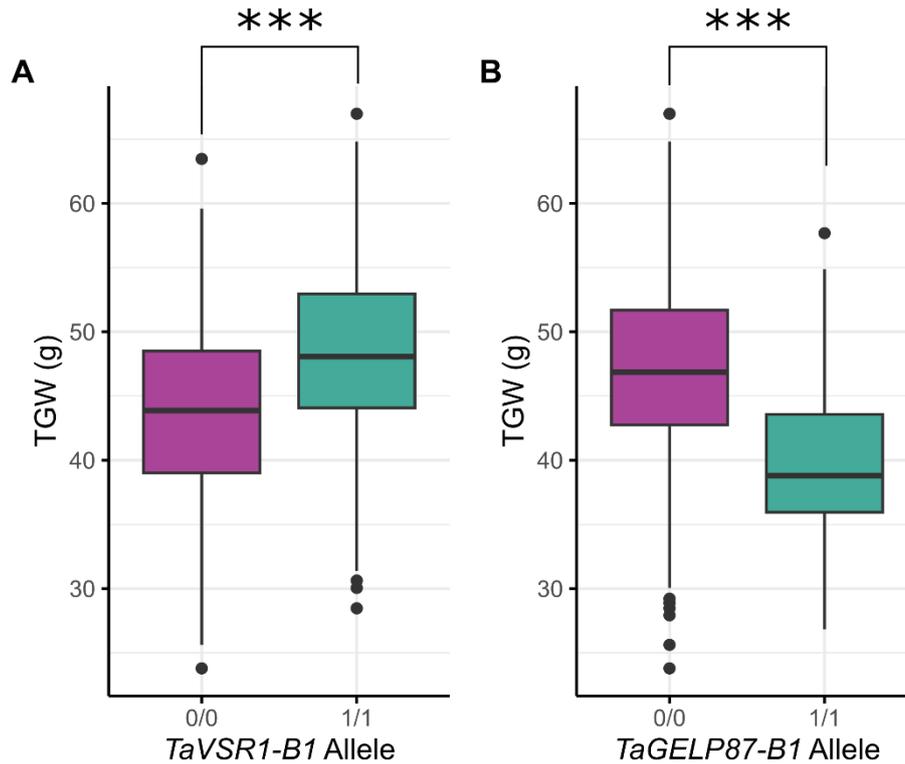


Figure 2-8: Thousand Grain Weight (TGW) in grams for lines from the A. E. Watkins hexaploid landrace collection<sup>34</sup> with specific SNPs in putative miRNA binding sites in (A) *TaVSR1-B1* (*VACUOLAR SORTING RECEPTOR HOMOLOG 1-B1*) and (B) *TaGELP87-B1* (*GDSL ESTERASE/LIPASE PROTEIN 87-B1*). 0/0 corresponds to the reference nucleotide at this position (chr2B:89,555,683 for *TaVSR1-B1* and chr7B:526,865,360 for *TaGELP87-B1* in the IWGSC v1.0 assembly<sup>15</sup>), 1/1 corresponds to the alternative nucleotide at this position. Significance levels according to a one-way ANOVA are indicated with asterisks: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

A SNP at position chr7B:526,865,360 is within the gene *GDSL ESTERASE/LIPASE PROTEIN 87-B1* (*TaGELP87-B1*) (TraesCS7B01G290600) (named according the orthologue in rice) and lines with the alternative nucleotide at this position have a significantly lower TGW compared to lines with the reference nucleotide (one-way ANOVA:  $F(1, 675) = 64.04$ , adjusted  $p < 0.001$ ) (Figure 2-8B). Little is known about *TaGELP87-B1* or its orthologue in rice. There are three orthologues in Arabidopsis reported by Ensembl Plants (*GUARD-CELL-ENRICHED GDSL LIPASE 14* (*AtGGL14*), AT2G42990, and *GUARD-CELL-ENRICHED GDSL LIPASE 23* (*AtGGL23*))<sup>128</sup>, but little is known about any potential role in seed development. In the Azhurnaya developmental gene expression atlas, *TaGELP87-B1* is expressed >5 TPM in the coleoptile at seedling stage, the shoot axis and first leaf sheath at the first leaf stage, spikelets at 50% spike emergence, and in grains at the hard dough stage<sup>22,155</sup>. Highest *TaGELP87-B1*

expression is in grains at 13.3 TPM<sup>22,155</sup>. Expression in tissues which could logically affect TGW means that this is also a promising avenue for future research.

The synonymous SNP in *TaGELP87-B1* is in the putative binding site of miRNA candidate-367, which shares the most sequence similarity to tae-miR9664-3p in miRBase<sup>94</sup>. tae-miR9664-3p is differentially expressed in high and low nitrogen conditions<sup>153</sup>. The alternative nucleotide confers an additional match with miRNA candidate-367 compared to the wildtype which has a mismatch at this position. The SNP is in position 20 of the binding site (from the 5' end of the miRNA), outside the miRNA seed region<sup>141-143</sup>. The 612 lines with the reference adenine at this position have a mean TGW of 47.0 g, while the 65 lines with the alternative cytosine at this position have a mean TGW of 39.9 g, a 15.1% reduction. The modern varieties ArinaLrFor, Jagger, Julius, Lancer, Landmark, Mace, Norin61, and Stanley all contain the alternative allele which confers a higher TGW, which presents the possibility that it has been unconsciously selected for by breeders.

## 2.5 Discussion

### 2.5.1 This dataset represents a comprehensive overview of miRNAs expressed during early wheat spike development

In this chapter I aimed to generate a comprehensive database of miRNAs which act during the key stages of early spike development. To do so, I generated a high-quality sRNA-Seq dataset. I maximised the sensitivity of this study by sequencing to a high depth.

The alignment rate for my sRNA-Seq dataset was relatively low against my list of candidate miRNAs at 0.43-2.10%. In the literature, mapping rates for sRNA-Seq datasets vary. However, studies where miRNAs were either enriched using a glass-fibre filter (*e.g.*, *mirVana*<sup>TM</sup> miRNA Isolation Kit (ambion)) or size-selected by polyacrylamide gel (as the data here was) also produced low alignment rates against lists of known miRNAs ranging from 0.11-8.46%<sup>157-159</sup>. My data had a high alignment rate against the full wheat genome, and the read size distribution showed the expected peaks at 21 and 24 nucleotides<sup>160,161</sup>. Combined, this evidence means that I remain confident in the sequencing data quality. The reads that did not align with the candidate miRNAs are likely to originate from other small non-coding RNAs (such as tRNA and snRNA) or degraded mRNA. In future, I would use an extraction technique such as the *mirVana*<sup>TM</sup> miRNA Isolation Kit (ambion) which includes a glass-fibre filter which generates samples enriched for RNA <200 nt. I could then evaluate whether this increases the proportion of NGS reads originating from miRNAs.

I used three independent pipelines to identify known and novel miRNAs in this dataset, again to increase the level of sensitivity. Previous studies have shown that the three pipelines I used (miRDeep-P2, miRador and ShortStack) have variable performance according to the dataset, and that despite some overlap, each produce unique predictions <sup>115</sup>. Hammond, *et al.* <sup>115</sup> recommended that multiple pipelines could be used to provide maximum benefit.

miRDeep-P2 predicted more unique miRNAs than miRador and ShortStack. 90.3% of miRDeep-P2-predicted miRNAs were unique (not predicted by either one of the other pipelines), compared to 28.0% and 11.8% for miRador and ShortStack respectively. miRDeep-P2 does not collapse closely related alignments in the same way that miRador and ShortStack do; single nucleotide positional variants (where the miRNA sequence is identical but includes a single additional nucleotide at the 5' or 3' end) are annotated by miRDeep-P2 as unique candidates. This methodology may partially explain why the number of miRNAs that miRDeep-P2 predicts is so much higher than for ShortStack or miRador. It would be beneficial to consolidate single nucleotide positional variants in my dataset to avoid redundancy.

All three pipelines use the criteria in Axtell and Meyers <sup>74</sup> to filter miRNA candidates, therefore the level of precision ( $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$ ) for each pipeline should be comparable. Hammond, *et al.* <sup>115</sup> calculated the precision of miRador, ShortStack and miRDeep-P2 for datasets in Arabidopsis, rice, and maize against the miRBase database and sPARTA (small RNA-PARE target analyser) <sup>162</sup> (generating a list of miRNAs by predicting miRNA targets and validating them using PARE). The level of precision in relation to sPARTA and miRBase varied according to species <sup>115</sup>. Despite producing a higher number of candidates, miRDeep-P2 was only the least precise tool when analysing a maize sRNA-Seq dataset in relation to the miRBase database <sup>115</sup>. In my dataset, some well-characterised miRNAs such as miR156 (candidate-351) were only predicted by miRDeep-P2. This suggests that despite the substantially higher number of predictions made by miRDeep-P2, it does not have a lower level of precision than miRador or ShortStack. By combining the results of three different pipelines I have made best use of the sRNA-Seq dataset I generated.

The analysis I carried out on the dataset I generated predicted miRNAs that have been characterised in wheat such as miR156 <sup>137,138</sup> and miR165/166 <sup>95</sup>. My analysis also correctly predicted known targets of these miRNAs as described in Section 2.4.2.1. This gives me confidence in the sensitivity of my analysis and in any novel predicted miRNAs and targets.

## 2.5.2 A novel locus translation tool is useful resource for the genomics community

As part of my analysis, I needed to translate the transcript-relative loci reported by TargetFinder to genome-relative loci so that I could identify variation in miRNA binding sites. Despite an extensive search, no appropriate tools are currently available that can be used on non-model species such as wheat. Therefore, I generated a bash/R-based script which can accurately translate transcript-relative locations to genome-relative locations. This script (Section A.6) is publicly available at [https://github.com/Uauy-Lab/transcript\\_to\\_genomic\\_loci](https://github.com/Uauy-Lab/transcript_to_genomic_loci) and will provide a simple-to-use tool for this common task for the genomics community.

## 2.5.3 A comprehensive wheat miRNA database constitutes a key resource for developmental biology

The 527 candidate miRNAs from at least 141 families and 3,535 candidate targets which I have predicted *in silico* provide a valuable novel dataset summarising miRNAs expressed during early wheat spike development. 59 of these belong to miRNA families which are not present in the IWGSC v1.0 annotation<sup>15</sup>, miRBase<sup>94</sup> or sRNAanno<sup>96</sup> and I have not assigned 145 miRNA candidates to any known miRNA family. These data will open interesting new avenues for research, some of which are already being implemented (Section 2.4.3.2). By developing a more comprehensive understanding of the miRNAs that are present in the wheat spike during this crucial period of development, we can help improve our understanding of wheat inflorescence biology more generally.

The sRNA-Seq dataset I have generated will be useful to understand the abundance of miRNAs during wheat spike development. A high-quality sRNA-Seq dataset means that one can distinguish between the expression of individual miRNA family members, something which I took advantage of in the detailed characterisation of the miR172-*AP2* interaction I describe in Chapter 3. As I will further describe in Chapter 3, there is evidence in Arabidopsis that closely related miRNA family members can have different functions<sup>163,164</sup>. Stem-loop RT-qPCR (real-time quantitative reverse transcription PCR) is unlikely to reliably distinguish between miRNA family members, particularly those with differences at their 5' end<sup>165</sup>. Therefore, sRNA-Seq is the best avenue for characterising the expression profile of individual mature miRNAs.

miRNAs play roles in key genetic pathways which affect wheat spike development and mature morphology such as the regulation of *AP2*-like genes by miR172, and *SPL* genes by miR156. Modifications to these interactions result in significant changes to spike architecture traits such

as spikelet density<sup>32,66</sup> and grain number<sup>102</sup>. Disrupting specific miRNA interactions by using miRNA overexpression constructs<sup>32</sup>, knocking down miRNA abundance using target mimicry<sup>32,166</sup> (the expression of MIM (mimic) constructs which sequester miRNAs), and mutagenizing miRNA binding sites<sup>66</sup>, can help to characterise miRNA-target interactions.

I have identified candidate TILLING lines with variation in miRNA binding sites (Section 2.4.3.2). I have predicted the phenotypic consequences of the SNPs in these TILLING lines and am currently growing the lines to quantify the relevant phenotypes. By interfering with these miRNA-mRNA interactions I expect these lines to show differences in agronomically important traits such as flowering time and spikelet density. By improving our knowledge of wheat spike development, we can effectively target breeding efforts to modify specific genetic interactions.

#### **2.5.4 A comprehensive wheat miRNA database can be used for better wheat breeding**

Potential miRNA-related targets for wheat breeding can be identified from this dataset. There is existing natural and induced variation in miRNA binding sites (Section 2.4.3) which would often confer a dominant increase in target protein abundance. miRNA and their target sites are highly conserved across land plants<sup>167</sup>, which may limit the natural variation that can be found in wheat populations such as the A. E. Watkins landrace collection<sup>34</sup>. In my analysis, I found 385 unique miRNA target site variants in the hexaploid Watkins collection (0.37 variants per line). This is compared to the hexaploid Cadenza TILLING lines which contain 1.49 miRNA target site variants per line. It is tempting to speculate that this four-fold increase in induced mutants compared to natural variation reflects the fact that nature and humans have actively selected for conserved miRNA binding sites due to their key functions which are sequence-dependent. Induced variation, on the other hand, is not under artificial selection pressure and SNPs must only be non-lethal and not confer sterility to be carried forward. To more accurately compare the levels of miRNA binding site variation in these two contrasting populations, it would be important to establish the level of variation in surrounding regions, as SNPs may be more abundant in the exons of TILLING lines compared to lines in the Watkins collection.

I have identified both natural and induced variation with potential for use in future wheat breeding. TILLING lines were identified which have induced miRNA binding site variation in genes which we know affect wheat spike development. I also identified natural variation in miRNA binding sites in the Watkins collection which correlates with differences in key phenotypes such as TGW and spike length. The associations made between SNPs in miRNA binding sites and phenotypes in the Watkins collections are likely to include false positives, as we know that variation is often associated with a haplotype block<sup>168</sup>; the correlation may

stem from a nearby allele which has been inherited together with the miRNA binding site SNP. However, the SNPs which I explored in this thesis which are associated with changes to TGW are in genes which are expressed in wheat grains <sup>22,155</sup>, as are the miRNAs which target them <sup>153</sup>. Therefore, these may represent exciting avenues for future research. A more detailed analysis of the miRNA binding site SNPs that are correlated with phenotypic changes may elucidate further promising candidates.

The variation identified in Watkins and TILLING lines is non-genetically modified (GM) so would not be subject to any restrictive legislation regarding its growth or commercial use. This is important in the current legislative landscape, as the release and marketing of genetically modified organisms (GMOs) remains under strict control <sup>169</sup>. The regulation surrounding gene edited crops has been relaxed over the past five years in the UK <sup>170</sup>, providing scope to develop and commercialise gene-edited crops. There is potential to use gene editing to alter miRNAs and their target sites, although this is limited by the small target region (just 20-24 nucleotides for miRNA target sites) and whether there's a nearby protospacer adjacent motif (PAM). Also, any non-synonymous changes to miRNA binding sites in open reading frames (ORFs) may also impact the amino acid sequence as well as miRNA binding. This is of particular relevance in plants as it has been suggested that, unlike in mammals where the majority of miRNAs bind to 3' UTRs <sup>171</sup>, miRNA binding sites in plants are distributed across genes <sup>172</sup>. However, this appears to be mostly anecdotal knowledge. A more comprehensive analysis of miRNA target site locations using this dataset would be beneficial to assess how much consideration would need to be given to the impact on non-synonymous mutations in miRNA binding sites. AlphaFold could also be used to predict and minimise the impact of non-synonymous SNPs on protein structure <sup>173</sup>. Unintended effects could also be minimised by using prime editing (reviewed in Chen and Liu <sup>174</sup>) to make precise single nucleotide changes to miRNA target sites without affecting the rest of the mRNA.

One advantage of exploiting SNPs in miRNA binding sites is that the effect of specific SNPs can be easily tested using a dual luciferase assay. I will describe how I used a dual luciferase assay to test a range of miRNA binding sites in Chapter 3. This method can be used to quickly estimate changes in target mRNA and protein level for different miRNA target sites. This assay does not account for the mRNA context of the binding site (and therefore binding site accessibility), however it provides a good initial indication of whether particular SNPs in target sites affect miRNA binding before taking lines forward for further fundamental study or into pre-breeding pipelines.

### 2.5.5 Future perspectives

Novel sources of variation are essential to facilitate crop improvement and ensure future food security in the context of an increasing global population. Dominant and easily selectable variation, such as that often conferred by changes to miRNA binding sites, is of particular use to wheat breeders. This strategy was highlighted as an effective avenue for plant breeding in Brassica in a recent review by Rani, *et al.*<sup>175</sup>; the potential highlighted in this review is also applicable to wheat.

New tools and resources, including this one, allow us to carry out developmental biology experiments directly in crop species, rather than transferring knowledge from model species such as Arabidopsis. This has the potential to accelerate the acquisition of knowledge. Future work that was beyond the scope of this thesis could include the development of a miRNA expression atlas to complement that published by Ramírez-González, *et al.*<sup>22</sup> for mRNA. A more comprehensive resource such as this would be valuable for studies in wheat developmental biology and for wheat breeding.

# 3 How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by microRNA172?

## 3.1 Chapter summary

In this chapter, I deepened our understanding of a single genetic network involving a miRNA which plays a key role during wheat spike development. The miR172-*AP2* interaction has been studied previously in wheat<sup>32,42,66</sup>, however I have taken a nuanced approach to disentangle the distinct functions of miR172 family members and of the *AP2*-like genes *AP2-2* and *AP2-5*. I found that *AP2-2* and *AP2-5* have opposite expression profiles during early wheat spike development (increasing and decreasing, respectively) and I generated hypotheses to explain these patterns in the context of their repression by miR172. I found that miR172a and miR172b repress *AP2-2* and *AP2-5* equally in a dual luciferase assay, so different miR172 family members do not target different *AP2*-like binding sites. I have begun to generate transgenic lines with modified *AP2-5A* miR172 binding sites to test further test this hypothesis in the context of *AP2-5A* mRNA. Using MERFISH (multiplexed error-robust fluorescence *in situ* hybridization), I found that there were relatively few cells in a W5.0 wheat spike where *AP2-2* transcripts were present, but *AP2-5* transcripts were not. This suggests that the opposite expression profiles cannot be explained by miR172 being absent from domains where *AP2-2* is expressed. I have described ways in which I could test my final hypothesis, that *AP2-2* and *AP2-5* have significantly different transcription rates. By taking this detailed approach, I have begun to shed light on the distinct functions of closely related miRNAs and genes, as has been shown in *Arabidopsis*.

## 3.2 Introduction

### 3.2.1 The interaction between miR172 and *AP2*-like genes is a cornerstone of wheat spike development

Small changes to miRNA-mRNA interactions can result in outsized phenotypic effects. One of the key miRNA-mRNA interactions that regulates wheat spike development is the interaction between *AP2*-like genes and miR172.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

Alteration of the miR172-AP2-5 interaction has been identified as an important step during wheat domestication<sup>32,66</sup>. AP2-5A, also known as *Q*, is a member of the AP2/ERF transcription factor family. The crucial role of this single locus in the formation of modern wheat spikes has been known since 1954<sup>176</sup>. The *Q* allele of AP2-5A confers a free-threshing phenotype which is essential for modern wheat cultivation, as well as other useful phenotypes such as reduced rachis fragility, a compact spike (higher spikelet density) and free-threshing grains<sup>177,178</sup>. The ancestral *q* allele of AP2-5A confers a speltoid spike (lower spikelet density) with tenacious glumes, fragile rachises and non-free-threshing seed<sup>177</sup>. AP2-5A, like several AP2-like genes in Arabidopsis, is negatively regulated by miR172 in wheat<sup>32</sup>. The *Q* allele of AP2-5A differs from the *q* allele by a SNP in the miR172 binding site, which subtly alleviates miR172-mediated repression (Figure 3-1)<sup>32,177</sup>. It does so by changing the binding at this position from a Watson-Crick base pair (a C:G pair) to a G:U wobble. G:U base pairs (reviewed by Varani and McClain<sup>179</sup>) are capable of binding and have a level of thermodynamic stability between a Watson-Crick base pair and other mismatches.

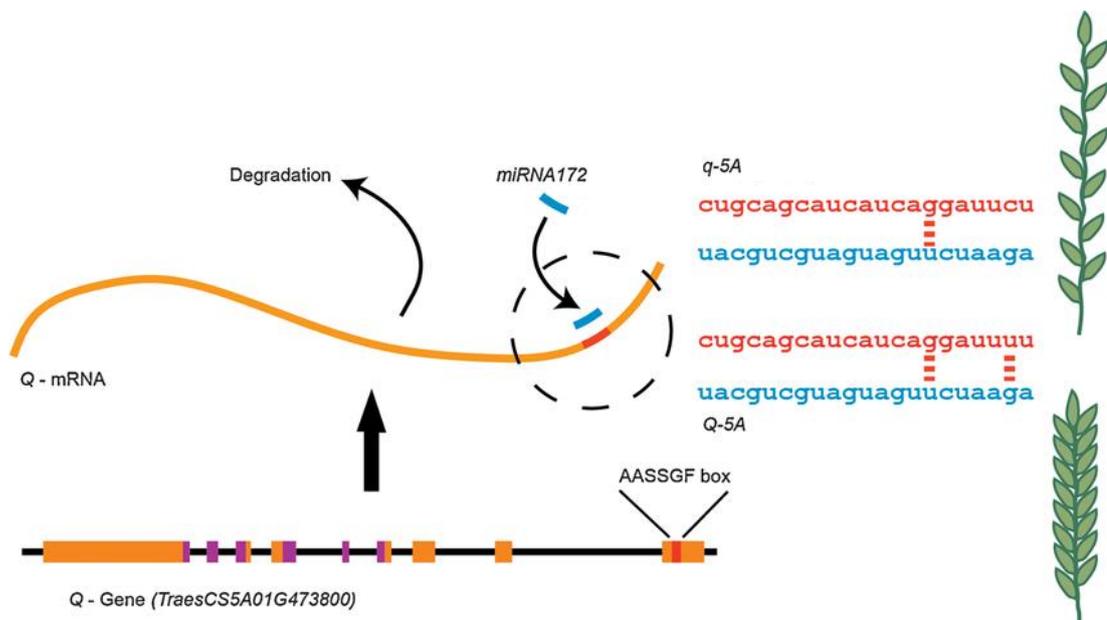


Figure 3-1: The effect of *q* and *Q* alleles of APETALA2-5A (AP2-5A) on microRNA172 (miR172) binding and wheat spike phenotypes. The *Q* allele contains an additional G:U wobble in the miR172 binding site compared to the ancestral *q* allele, reducing miR172-mediated degradation, conferring a compact spike phenotype with an increased number of spikelets per cm along the rachis. AP2 domains are shown in purple and the 'AASSGF' box (or miR172 binding site) is shown in red. Adapted from Gauley and Boden<sup>180</sup>.

Genotypes which confer more extreme changes in AP2-5A expression have more severe phenotypic effects<sup>32,66</sup>. *ap2-5* and Ubi::miR172 (constitutive overexpression of miR172) plants have lower levels of AP2-5 expression and protein level, respectively, than the wildtype

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

and this results in phenotypes such as sham ramification (a high number of florets per spikelet) and glume-like bracts between the glumes and the first florets<sup>32</sup>. MIM172 plants (which contain a MIM construct which sequesters miR172) have a higher level of *AP2-5* expression than the wildtype. MIM172 plants have opposing phenotypes to *ap2-5* and *Ubi::miR172*<sup>32</sup>: a lower number of florets per spikelet and glumes appear lemma-like with a reduced keel and longer awns<sup>32</sup>.

Studies in *Arabidopsis* have shown that the orthologue of *AP2-5A* (*TARGET OF EATI* (*AtTOE1*), also known as *RELATED TO AP2.7* (*AtRAP27*)) delays flowering<sup>181</sup>. There are two orthologues of *AP2-5* in rice<sup>42</sup>: *INDETERMINATE SPIKELET1* (*OsIDS1*) and *SUPERNUMERARY BRACT* (*OsSNB*). Knockout mutations in either of these genes results in a similar phenotype to *ap2-5* in that spikelets produce additional bracts (reminiscent of sham ramification)<sup>182,183</sup>. In barley, knockouts of the *AP2-5* orthologue (*HvAPETALA2Like-H5* (*HvAP2L2-H5*)) result in higher numbers of florets per spikelet and the formation of a terminal spikelet which is absent in wildtype barley spikes<sup>40</sup>.

This evidence suggests that *AP2-5* expression represses branching along the rachilla by increasing spikelet meristem determinacy. If *AP2-5* is absent, a higher number of floral meristems form along the rachilla, resulting in sham ramification. This indicates that there is an increased level of indeterminacy in the spikelet meristem.

*HvAP2L-H5* has been shown to be expressed in barley lemma and palea<sup>40</sup>, suggesting that *AP2-5* may promote a lemma-like identity. This is supported by the fact that in wheat low *AP2-5* expression leads to the formation of glume-like empty bracts, while high *AP2-5* expression leads to lemma-like glumes<sup>42</sup>. I hypothesize that *AP2-5* promotes a transition from glume to lemma identity. The release of floral meristem repression in spikelet phytomers (see Chapter 1 for a review of spikelet development) may be prevented by a strong glume identity (or promoted by a strong lemma identity), which is why in plants with low *AP2-5* expression there are additional glume-like bracts which do not contain floral organs.

### 3.2.2 *AP2-2* and *AP2-5* have partially redundant roles in wheat spike development

Although *AP2-5* is the most well-known, there are three other *AP2*-like genes in wheat with miR172 binding sites: *AP2-1*, *AP2-2*, and *AP2-7*<sup>42</sup>. *AP2-7* is not expressed in any tissues nor conditions sampled to date, suggesting it may be a pseudogene<sup>22,42</sup>. *AP2-5B* has also been shown to be a pseudogene<sup>184</sup>. *AP2-1* is expressed in the roots and vegetative tissues but not in the spike or grain<sup>22,42</sup>. *AP2-2* has a very similar expression profile to *AP2-5*, and it has been shown that these genes have partially redundant functions<sup>22,42</sup>. *AP2-2* is orthologous to *AP2*

**Chapter 3: How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by *miR172*?**

in Arabidopsis <sup>42</sup> and *SHATTERING ABORTION1* (*OsSHAT1*) in rice <sup>42</sup>. Plants with an *rAp2l-A2* allele which is resistant to miR172 (resulting in higher *AP2-2* expression) have increased spikelet density and lemma-like glumes, as well as reduced lodicule swelling, showing that *AP2-2* is actively regulated by miR172 <sup>42</sup>. *ap2-2* plants have wildtype spike architecture, however have larger lodicules <sup>42</sup>. Increasing *AP2-2* expression has similar effects as increasing *AP2-5* (except for the *AP2-2*-specific lodicule phenotype), while decreasing *AP2-2* expression only affects lodicules, which suggests partial functional redundancy.

*ap2-2 ap2-5* double mutants have a stronger phenotype compared to either single mutant, with spikes made up almost entirely with empty lemmas which seem to have lemma- and glume-like characteristics <sup>42</sup>. This suggests that they may be acting via the same genetic pathway(s). I hypothesise that *AP2-2* may also promote the transition from glume to lemma identity in a redundant manner with *AP2-5*, meaning that in the double mutant the bracts have a strong glume identity, repressing differentiation of the axillary meristem into a floral meristem. This generates plants with numerous empty glume-like bracts. This phenotype was similar to that seen in a strong *Ubi::miR172* line, suggesting that of the genes targeted by miR172, *AP2-2* and *AP2-5* are likely to have the largest effect on spike architecture <sup>32,42</sup>.

Homeotic conversions were also observed in *ap2-2 ap2-5* double mutants; the two lodicules and the adjacent stamen were replaced by a second carpel <sup>42</sup>. Although homeotic conversions are a fundamental part of the ABCDE model <sup>45,50</sup>, this conversion is not consistent with the model based on Arabidopsis. Therefore, although some elements of the ABCDE model are conserved between dicots such as Arabidopsis and monocots such as wheat, there is evidence of divergence in gene function.

### 3.2.3 Knowledge from other species shows we must properly disentangle the functions of miRNA family members

I have shown in Chapter 2 that miRNAs play a critical role during inflorescence development. Hence, to disentangle the complex interactions occurring during this key period, we must fully understand the role that miRNAs play. Thus far, studies on the miR172-*AP2* interaction in wheat have treated miR172 as a single molecule <sup>32,42</sup>. However, at least three different mature miR172 sequences have been proposed, originating from five different loci in the wheat genome <sup>32</sup>. There is evidence in Arabidopsis that members of the same miRNA family can have unique functions. miR164c (but not miR164a or miR164b) has been shown to control petal number <sup>163</sup>. A second study also found divergent functions for miR164 family members in Arabidopsis: *MIR164b* is required to specify the shoot and flower boundary while *MIR164a* is required for fruit growth <sup>102</sup>. In rice, *MIR396e* and *MIR396f* were found to affect shoot architecture and grain size, but not *MIR396a-d* or *MIR396g-h* <sup>185</sup>. miR172 family

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

members have been shown to have highly divergent expression patterns and functions in *Arabidopsis* and *Solanum lycopersicum* (tomato) <sup>164,186</sup>. Therefore, it is critical that where possible we investigate the characteristics of miRNA family members independently to fully understand their functions. Due to recent advances in genetic resources and protocols, it is now possible for this to be done in wheat.

### 3.2.4 Aims and hypotheses

My initial aim was to further characterise how mir172 regulated AP2-like genes during spike development. Early results presented an unexpected finding whereby AP2-2 and AP2-5 have opposite expression patterns despite previous studies reporting similar expression profiles <sup>32</sup>. Therefore, I honed my question towards understanding how AP2-2 and AP2-5 have opposite expression profiles while being regulated by the same miRNA.

## 3.3 Methods

### 3.3.1 Phylogenetic analysis

I downloaded all amino acid sequences from Ensembl Plants release 58 apart from for maize for which I used release 59 <sup>187</sup>. I included all homoeologues of AP2-2 and AP2-5 from *T. aestivum* and orthologues of these genes in *O. sativa* ssp. *japonica*, maize and *Arabidopsis* as reported by Ensembl Plants <sup>187</sup>. Finally, I used the amino acid sequence of *TaSEPALLATA1-A1* (*TaSEPI-A1*) (TraesCS4A02G078700) as an outgroup.

I used MEGAX <sup>188</sup> for phylogenetic analysis according to the guidance in Hall <sup>189</sup>. I aligned the amino acid sequences using MUSCLE <sup>190</sup> with default parameters. I computed a Maximum Likelihood tree using the JTT (Jones-Taylor-Thornton) matrix-based model <sup>191</sup> and used the 95 positions with more than 95% coverage for tree construction. I used a discrete Gamma distribution to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.1237)) and 100 bootstrap replications to estimate the confidence of each clade.

### 3.3.2 Paragon RNA-Seq

This dataset is an internal Uauy group resource generated by Dr Nikolai Adamski, Dr Anna Backhaus and Max Jones.

*T. aestivum* plants of cv ‘Paragon’ NILs carrying the *VRT-A2a* allele ( $PI^{WT}$ ) and *VRT-A2b* allele ( $PI^{POL}$ ) as described in Adamski, *et al.* <sup>56</sup> were grown in 24-cell trays containing “John Innes Cereal Mix” (described in Section 2.3.1). All plants were grown in growth chambers at 65% humidity with 16 h photoperiods and 20 °C/15 °C day/night temperatures. Basal and

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

central sections of W2.5, W4.0 and W5.0 developing spikes were dissected as described in Section 2.3.1 on dry ice then flash-frozen in liquid nitrogen.

Total RNA was extracted using the Direct-zol RNA Microprep Kit (Zymo Research) according to the manufacturer's instructions <sup>109</sup>, including on-column DNase I treatment. mRNA was sequenced by Novogene (Cambridge, UK) and approximately 50 million 150 bp PE reads were sequenced per sample. Four biological replicates were sequenced per timepoint and genotype.

Data quality was checked using FastQC v0.11.7 (Andrews <sup>110</sup>). Reads were trimmed using Trim Galore v0.4.2 and Cutadapt v1.9.1 (Martin <sup>113</sup>, Krueger <sup>192</sup>) with a minimum length after trimming of 50 bp (`--length 50`). The Kallisto v0.44.0 (Bray, *et al.* <sup>193</sup>) `quant` command was used to quantify transcripts against the IWGSC v1.1 annotation <sup>98</sup> with the number of bootstrap samples set to 30 (`-b 30`).

I carried out statistical analyses using R v4.3.2 <sup>123</sup>. I calculated a mean TPM (transcripts per million) value for each replicate by averaging the TPM values for the central and basal sections. I used one-way ANOVAs with Tukey post-hoc tests to test differences between the expression of homoeologues at each timepoint.

### 3.3.3 Chinese Spring RNA-Seq

This dataset is an internal Uauy group resource generated by Max Jones, Isabel Faci, Katie Long and Dr Neil McKenzie.

#### 3.3.3.1 RNA Sequencing

*T. aestivum* cv 'Chinese Spring' seeds were germinated in Petri dishes on filter paper wetted with distilled water. The seeds were kept at 4 °C for two days, then at room temperature for two days. Germinated seeds were sown in 24-cell trays of "John Innes Cereal Mix" (described in Section 2.3.1). W1.5, W2.0, W2.5 and W3.0 samples were collected from seedlings in 24-cell trays, later stages were sampled from plants that were potted on at 21 days into 1 L pots containing John Innes Cereal Mix. All plants were grown in growth chambers at 60% humidity with 16 h photoperiods and 20 °C/15 °C day/night temperatures. All samples originate from 14 batches of plants grown and sampled over 143 days.

Spikes were harvested into 2 mL microcentrifuge tubes on dry ice and flash-frozen in liquid nitrogen within 20 min, then stored at -70 °C. Tissue was homogenised using a TissueLyser II bead mill (QIAGEN). The adapter blocks were supercooled by surrounding them with dry ice and pouring liquid nitrogen over. Samples were homogenised at 27.5 Hz for 30 s twice. Samples for W1.5, W2.0 and W2.5 were derived from samples split into two tubes after the

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

first round of homogenisation (one was used to generate this RNA-Seq dataset; the other was used for an experiment not described here). Homogenised samples were stored at -70 °C.

Total RNA was extracted from all samples within two weeks of homogenisation. 800 µL TRIzol Reagent was added to each sample and mixed gently by inversion for 2-3 min at RT. The samples were incubated with the TRIzol reagent for 5 min at RT with occasional agitation by inversion. The samples were centrifuged at 12,000 RCF for 10 min at 4 °C and the supernatant was transferred to a new 1.5 mL microcentrifuge tube. 160 µL chloroform was added and mixed thoroughly by shaking for 15 s. The samples were incubated for 2-3 min at RT. The samples were centrifuged at 12,000 RCF for 10 min at 4 °C and the aqueous phase was transferred to a new microcentrifuge tube. The RNA Clean and Concentrator Kit (Zymo) was used to clean up the extracted RNA. One volume of ethanol was added to the aqueous phase and mixed. The sample was transferred to a Zymo-Spin IC column and centrifuged at 14,000 RCF for 30 s before discarding the flow-through. The rest of the protocol was carried out according to the Zymo kit manufacturer's instructions from step 4, including on-column Dnase I treatment <sup>194</sup>. RNA was eluted twice, in 14 µL nuclease-free water each time, into a single microcentrifuge tube.

mRNA sequencing was carried out by Novogene UK using the Illumina NovaSeq 6000 system. Approximately 50 million 150 bp PE reads were sequenced per sample.

### 3.3.3.2 Bioinformatic analysis

Max Jones (JIC, UK) performed all bioinformatic analysis of this RNA-Seq dataset. Reads were trimmed using Fastp v0.23.1 (Chen, *et al.* <sup>195</sup>) using the `--detect_adapter_for_pe` option. Kallisto v0.44.0 (Bray, *et al.* <sup>193</sup>) was used for transcript quantification using a Kallisto index (k=31) of the IWGSC v1.1 genome annotation <sup>98</sup>. The `--bias` and `--bootstrap-samples=30` options were used.

### 3.3.4 miR172 meta-analysis

To generate a list of miR172 loci, I used information from databases:

- MIRBASE <sup>94</sup>
- sRNAanno <sup>96</sup>

and publications:

- Sun, *et al.* <sup>196</sup>
- Debernardi, *et al.* <sup>32</sup>
- Tang, *et al.* <sup>133</sup>

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

MIRBASE is the most commonly used database for miRNAs, however, the most recent release was published in 2019 and there are only 122 loci entries for *T. aestivum*<sup>94</sup>. These 122 entries do not include miR172<sup>94</sup>. However, the database does include some miR172 family members from the D-genome progenitor of wheat, *A. tauschii*, which I included in my analysis. sRNAanno is a newer database for small RNAs<sup>96</sup> which contains a larger number of loci for *T. aestivum* miRNAs (616) than MIRBASE. In Figure S1A of Debernardi, *et al.*<sup>32</sup> the authors outline five mature sequences for miR172. However, the source of these sequences is not reported, only that they are orthologs of miR172 in rice. The papers they reference related to wheat miRNAs are Tang, *et al.*<sup>133</sup> and Sun, *et al.*<sup>196</sup>, which I included in my analysis.

I collated a list of miR172 family members reported by the above sources. I checked if the mature sequences in my list appeared in three small RNA-Seq datasets containing grain, spike, seedling, and leaf tissue at various time points<sup>132-134</sup>. I excluded any mature sequences absent from all the RNA-Seq datasets. I further excluded a single mature sequence present in only one dataset in very low numbers as the reported pri-miRNA sequences did not produce any BLAST hits against the IWGSC v1.0 genome assembly<sup>15,114</sup>.

### 3.3.5 Syntenic analysis

To establish whether the *MIR172* loci were homoeologous, I examined the synteny of these loci. I took the closest upstream and downstream IWGSC v1.0 annotations. If I located putative homoeologous pre-miRNA loci between the homoeologues of these genes according to Ensembl Plants (release 40)<sup>197</sup>, I deemed the loci to be syntenic and therefore homoeologues.

If the closest annotations had no homoeologues according to Ensembl Plants, I took the closest annotation that had reported homoeologues.

### 3.3.6 MERFISH

MERFISH was carried out on young wheat spikes by Katie Long (JIC, UK) and Ashleigh Lister (Earlham Institute, UK). The MERFISH protocol was first described by Chen, *et al.*<sup>198</sup>.

#### 3.3.6.1 Plant tissue

*PI<sup>WT</sup>* and *PI<sup>POL</sup>* NILs (described in Section 3.3.2) seed was germinated as described in Simmonds, *et al.*<sup>106</sup>. Plants were grown in 15-cell trays containing “F2 starter and grit” (90% peat, 10% grit, 4 kg/m<sup>3</sup> dolomitic limestone, 1.2 kg/m<sup>3</sup> osmocote start). All plants were grown in growth chambers at 65% humidity with 16 h photoperiods and 20 °C/16 °C day/night temperatures.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

$PI^{WT}$  and  $PI^{POL}$  plants were grown as described in Section 2.3.1. Whole W4.0 (terminal spikelet) and W5.0 ( $0.5\text{ cm} < \text{spike length} < 1.5\text{ cm}$ ) spikes were dissected using microdissection tools and a dissection microscope (Leica, Germany) which had been cleaned using 70% ethanol and Blitz RNase Spray (Thistle Scientific).

### 3.3.6.2 Sample preparation

Spikes were placed in 4% paraformaldehyde (PFA) in 1X phosphate-buffered saline (PBS) for up to 30 min, then vacuum infiltrated until the spikes sunk (approximately 10 min at 200 mbar). Spikes were washed three times with 1X PBS then immersed in 15% sucrose in 1X PBS at 4 °C for 6-12 h. The spikes were transferred to 30% sucrose in 1X PBS at 4 °C until the spikes sunk (approximately overnight).

Spikes were placed in an optimal cutting temperature compound (OCT) bath and coated in OCT by gentle agitation. Tissue was trimmed as required and transferred to an OCT-filled cryomould. The tissue was frozen in OCT by placing the mould on a petri dish floating in liquid nitrogen. Samples were stored at -80 °C until required.

Samples were allowed to warm in the cryostat for 30 min before sectioning. Sectioning was done using a CryoStar NX70 cryostat (Thermo Scientific) with a chuck temperature of -20 °C, and a blade temperature of -18 °C. Sections were adhered to poly-L-lysine coated glass slides.

### 3.3.6.3 MERFISH imaging

MERFISH imaging was performed on prepared samples using a Vizgen MERSCOPE® according to the manufacturer's instructions <sup>199</sup>.

### 3.3.6.4 Data analysis

Cell segmentation and transcript assignment was carried out using cellpose2 <sup>200</sup> and Vizgen Post-processing Tool (VPT) <sup>201</sup>.

Sample concatenation was performed using scanorama <sup>202</sup>. Unsupervised clustering was performed using squidpy <sup>203</sup> and scanpy <sup>204</sup>. Normalised transcript counts were generated using the normalisation and log transformation functions in scanpy. Transcript counts were generated for filtered cells (excluding those with a low volume and/or total transcript count) and normalised so that every cell had the same total transcript count.

## 3.3.7 sRNA-Seq

Details of sRNA-Seq methods can be found in Sections 2.3.1 and 2.3.2.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

### 3.3.7.1 sRNA-Seq statistical analysis

I carried out statistical analyses of RPM data using R v4.3.2 (R Core Team <sup>123</sup>). miR172a corresponds to miRNA candidate-30 and miR172b corresponds to miRNA candidate-190.

I carried out a one-way ANOVA (using the base R `anova()` function) to test for differences in expression between miR172a and miR172b at W4.0. These samples were all dissected on the same day.

For W2.5 and W5.0 stages samples were collected in batches, so I performed a different analysis. In the PCA plot (Figure 2-3) I observed a batch effect, particularly between W5.0 samples collected on different days. I used a linear mixed model (LMM) to test for differences in expression between miR172a and miR172b at W2.5 and W5.0 to include sampling date as a random effect. I calculated an LMM using the R package ‘lme4’ (Bates, *et al.* <sup>205</sup>). The LMM equation used was  $\text{RPM} \sim \text{mirna\_id} + (1|\text{batch})$ . The models fit the assumptions of an LMM (normal distribution of residuals and homogeneity of variances). I used the `anova()` function to run ANOVAs on the LMMs.

### 3.3.8 Identifying miRNA target sites

Details of the methods I used to identify miRNA target sites can be found in Section 2.3.2.4.

### 3.3.9 Dual luciferase assay

I carried out this assay with help from Oscar Carey-Fung (University of Melbourne, Australia) and Max Jones (JIC, UK).

#### 3.3.9.1 Preparation of electro-competent *A. tumefaciens*

Electro-competent *Agrobacterium tumefaciens* was prepared with the help of Oscar Carey-Fung. We prepared competent *A. tumefaciens* stocks based on an adaptation of the Wen-jun and Forde <sup>206</sup> protocol. *A. tumefaciens* GV3101::pMP90 was a gift from Matthew Downie (JIC, UK). We inoculated 100 mL 25 µg/mL gentamycin, 50 µg/mL rifampicin LB (Lennox Broth) (1% w/v peptone from casein (tryptone), 0.5% w/v yeast extract, 0.5% w/v sodium chloride, 0.1% glucose) from a glycerol stock of GV3101::pMP90 and incubated it overnight at 28 °C, 100 RPM until the OD<sub>600</sub> was 0.5-0.75 (measured using an Eppendorf Biophotometer). We split the culture into four 50 mL Falcon tubes and cooled on ice for 15 min. We centrifuged the cells at 5,841 RCF for 10 min at 4 °C. We poured off the supernatant and resuspended in 50 mL 10% glycerol. We centrifuged cells at 4,000 RCF for 10 min at 4 °C. We poured off the supernatant and resuspended in 25 mL 10% glycerol. We centrifuged the cells at 4,000 RCF for 10 min at 4 °C. We poured off the supernatant and resuspended in 1 mL 10% glycerol. We centrifuged the cells at 4,000 RCF for 10 min at 4 °C.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

We poured off the supernatant and resuspended in 1 mL 10% glycerol. We centrifuged the cells at 4,000 RCF for 10 min at 4 °C. We poured off the supernatant and resuspended in 500 µL 10% glycerol. We froze 50 µL aliquots in liquid nitrogen and stored at -80 °C.

**3.3.9.2 Plasmid construction**

pGreen\_dualuc\_3'UTR\_sensor (Addgene plasmid #55206) and pGreen\_GUS\_competitor (Addgene plasmid # 55208) were gifts from Michael Axtell (Penn State University, USA) <sup>207</sup>. I streaked bacterial stabs on 30 mg/mL kanamycin LB agar plates (1% w/v peptone from casein (Tryptone), 0.5% w/v yeast extract, 0.1% w/v sodium chloride, 1.1% w/v agar) and incubated overnight at 37 °C. I picked single colonies and used them to inoculate 5 mL 30 mg/mL kanamycin LB starter cultures which I incubated overnight at 37 °C, 200 RPM. I made glycerol stocks from these cultures by storing 500 µL culture with 500 µl 40 % glycerol at -80 °C. I extracted plasmid from the remaining culture using a QIAprep spin miniprep kit (QIAGEN) according to the manufacturer's instructions <sup>208</sup>. I quantified plasmid yield using a Qubit dsDNA HS assay (Thermo Fisher) according to the manufacturer's instructions <sup>209</sup>. Plasmids were sequenced by Plasmidsaurus. The pGreen\_dualuc\_3'UTR\_sensor sequence was as expected, however the pGreen\_GUS\_competitor plasmid contained an unexpected 1,338 bp insertion between the *KanR* and *ori* sequences. Nucleotides 10-1,338 of this insertion have 100% homology with *IS10* sequences from the *Tn10* transposon (nucleotides 7,819-9,147 of GenBank entry AF162223.1 <sup>210,211</sup> and nucleotides 867-2,195 of GenBank entry AH003348.2 <sup>210,212</sup>). This sequence has been identified as a common contaminant in *E. coli* plasmids <sup>213</sup>. As the insertion was located between plasmid elements, I decided to continue cloning with the insertion.

**3.3.9.2.1 miRNA overexpression plasmids**

To generate the miRNA overexpression plasmids, miR172a and miR172b precursor sequences (Section A.2) were synthesised by GENEWIZ in a pUC-GW-Kan vector. I transformed the plasmids into Library Efficiency DH5α *Escherichia coli* (Invitrogen). I transformed the cells according to the manufacturer's instructions <sup>214</sup> using 10 ng of plasmid, except I used 50 µL of cells per transformation, I added 450 µL of SOC (super optimal broth with catabolite repression) medium for recovery, and the recovery period was altered to 1.5 h at 200 RPM. After the recovery period, I centrifuged the cells at 300 RCF for 3 min at room temperature. I removed 400 µL of supernatant and spread the remaining 100 µL on two 30 mg/mL kanamycin LB agar plates. I incubated the plates overnight at 37 °C. I created 10 mL starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.9.2. I quantified plasmid yield using a Nanodrop 1000 (Thermo). The insert was confirmed by Sanger sequencing using M13F(-21) and M13R primers (Section A.1, Table A-1) by GENEWIZ.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

I used the pGreen\_GUS\_competitor, pUC-GW-Kan-miR172a and pUC-GW-Kan-miR172b plasmids in a double restriction digest as in Table 3-1.

*Table 3-1: Reaction setup for double digestion of the miRNA overexpression plasmid inserts and vector, later ligated and used in the dual luciferase assay. The volumes and final amounts given are per reaction.*

<b>Reagent</b>	<b>Volume (µL)</b>	<b>Final amount</b>
Plasmid DNA	Variable	1 µg
NEB 10X CutSmart buffer	5	NA
NEB EcoRI-HF	0.5	10 units
NEB XhoI	0.5	10 units

I ran the digestion reaction at 37 °C for 1 h, then I heat inactivated the enzymes at 65 °C for 20 min. I ran 60 µL each of the pUC-GW-Kan-miR172a and pUC-GW-Kan-miR172b restriction digest reactions (20 µL per well) on a 1% agarose gel at 100 V for 50 min. I excised the required bands and extracted the DNA from the gel slices using the Promega Wizard SV Gel and PCR Clean-Up System according to the manufacturer's instructions <sup>215</sup>. I quantified and checked the quality of the DNA using a Nanodrop 1000 (Thermo) and Qubit dsDNA HS assay (Thermo Fisher). I phosphorylated the pGreen\_GUS\_competitor restriction digest reaction using rAPid alkaline phosphatase (Roche) according to the manufacturer's instructions <sup>216</sup>.

I set up 5:1 and 3:1 (insert:vector) ligation reactions to make pGreen\_GUS\_competitor-miR172a and pGreen\_GUS\_competitor-miR172b. I used T4 DNA ligase (New England Biolabs) according to the manufacturer's instructions <sup>217</sup>. I used 75 ng vector DNA for the 5:1 reactions, and 100 ng vector DNA for the 3:1 reactions. I incubated the reactions overnight at 4 °C.

I transformed 5-alpha competent *E. coli* (high efficiency) (New England Biolabs) with the ligation reaction product. I transformed the cells according to the manufacturer's instructions <sup>218</sup> using 5 µL ligation reaction. I spread 100 µL on a 30 mg/mL kanamycin LB agar plate. I spun down the remaining cells at 300 RCF for 3 min and removed 800 µL of supernatant. I resuspended the pellet in the remaining 100 µL medium and then spread it on a second 30 mg/mL kanamycin LB agar plate. I incubated the plates overnight at 37 °C.

I created 10 mL starter cultures from single colonies and extracted plasmid DNA as described in Section 3.3.9.2, except the starter colonies were incubated at 250 RPM. I quantified plasmid yield using a Qubit dsDNA HS assay (Thermo Fisher) according to the manufacturer's instructions <sup>209</sup>. To validate the pGreen\_GUS\_competitor-miR172a and

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

pGreen\_GUS\_competitor-miR172b plasmids (Figure 3-2 and Figure 3-3, respectively), I submitted the two plasmids to Plasmidsaurus for whole plasmid sequencing using ONT (Oxford Nanopore Technology) with custom analysis and annotation.

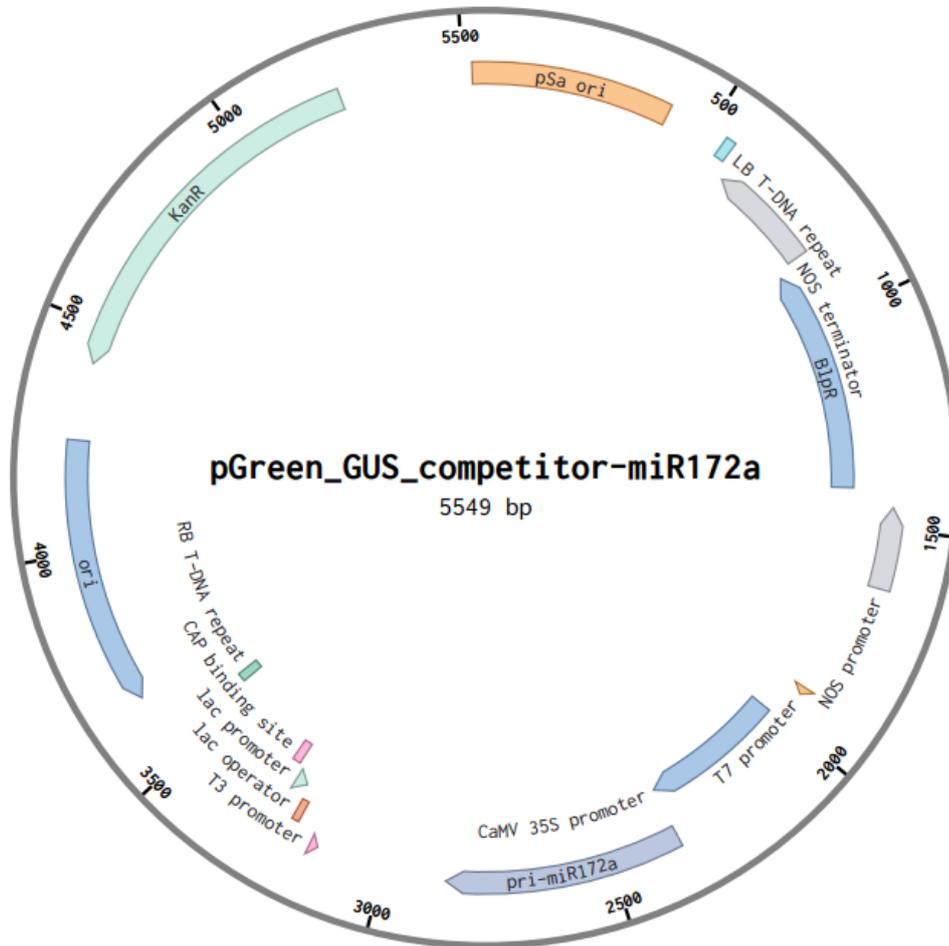


Figure 3-2: Plasmid map of pGreen\_GUS\_competitor-miR172a (total size: 5549 base pairs (bp)). The plasmid contains the following features which are annotated on the map: pSa ori (origin of replication from the bacterial pSa plasmid); LB T-DNA repeat (truncated left border of the nopaline C58 T-DNA repeat); NOS terminator (nopaline synthase terminator and nopaline synthase poly(A) signal); BLP (phosphinothricin acetyltransferase gene, confers resistance to bialophos and phosphinothricin); NOS promoter (nopaline synthase promoter); T7 promoter (bacteriophage T7 RNA polymerase promoter); CaMV 35S promoter (Cauliflower Mosaic Virus 35S promoter); pri-miR172a (primary microRNA172a sequence); T3 promoter (bacteriophage T3 RNA polymerase promoter); lac operator (from *E. coli*); lac promoter (from *E. coli*); CAP binding site (cyclic AMP receptor protein binding site from *E. coli*); RB T-DNA repeat (right border of the nopaline C58 T-DNA repeat); ori (origin of replication); KanR (Aminoglycoside 3'-phosphotransferase, confers resistance to kanamycin).

**Chapter 3:** How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?

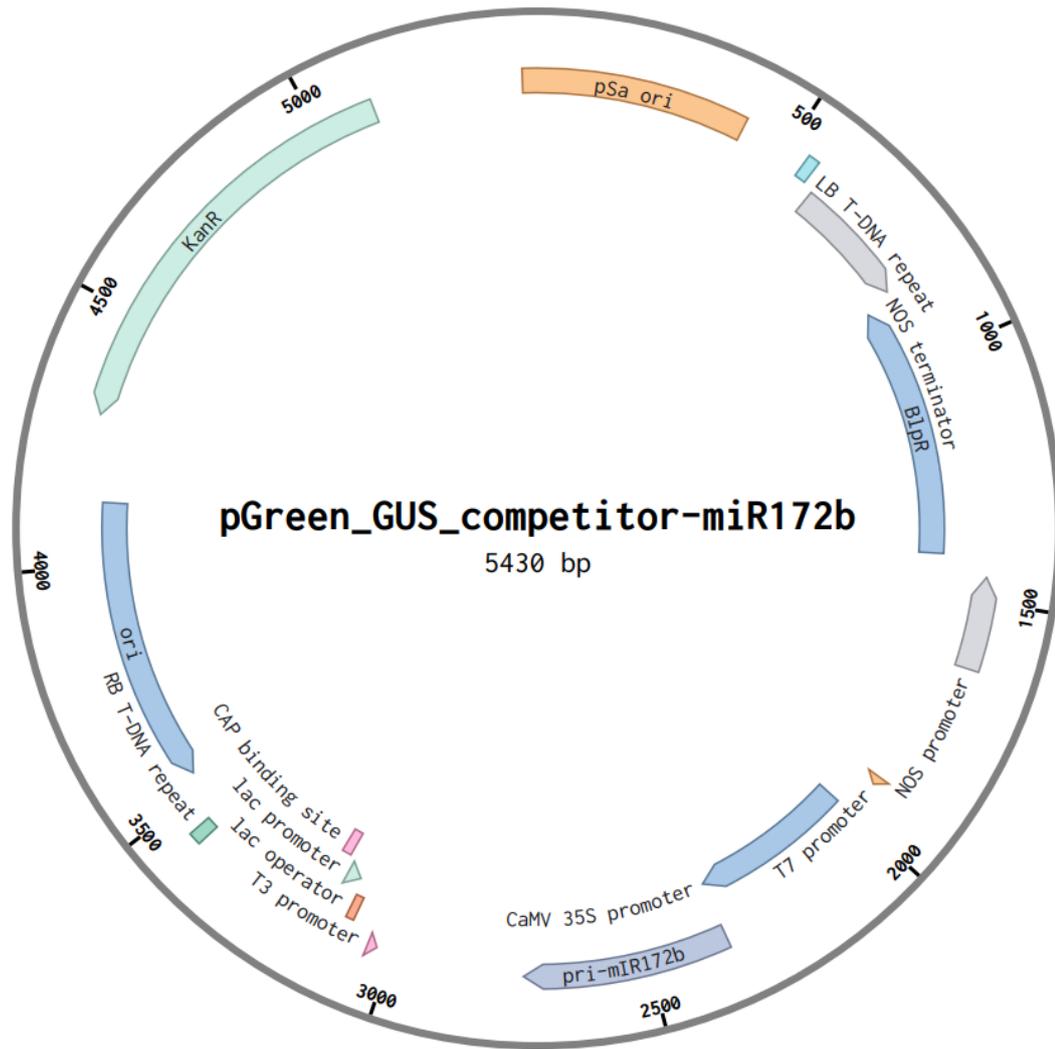


Figure 3-3: Plasmid map of pGreen\_GUS\_competitor-miR172b (total size: 5430 base pairs (bp)). The plasmid contains the following features which are annotated on the map: pSa ori (origin of replication from the bacterial pSa plasmid); LB T-DNA repeat (truncated left border of the nopaline C58 T-DNA repeat); NOS terminator (nopaline synthase terminator and nopaline synthase poly(A) signal); BLPB (phosphinothricin acetyltransferase gene, confers resistance to bialaphos and phosphinothricin); NOS promoter (nopaline synthase promoter); T7 promoter (bacteriophage T7 RNA polymerase promoter); CaMV 35S promoter (Cauliflower Mosaic Virus 35S promoter); pri-miR172b (primary microRNA172b sequence); T3 promoter (bacteriophage T3 RNA polymerase promoter); lac operator (from *E. coli*); lac promoter (from *E. coli*); CAP binding site (cyclic AMP receptor protein binding site from *E. coli*); RB T-DNA repeat (right border of the nopaline C58 T-DNA repeat); ori (origin of replication); KanR (Aminoglycoside 3'-phosphotransferase, confers resistance to kanamycin).

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

**3.3.9.2.2 Dual luciferase target plasmids**

I designed oligos for target sites which were synthesised by Merck (Table 3-2).

*Table 3-2: Oligos used to generate target sites used in the dual luciferase assay. The oligo sequence is provided in the 5' to 3' orientation.*

<b>Target site name</b>	<b>Oligo name</b>	<b>Oligo sequence</b>
5AQ	5AQ sense	CTAGGCTGCAGCATCATCAGGATTTTA
	5AQ antisense	CCGGTAAAATCCTGATGATGCTGCAGC
5Aq/B/D	5Aq/B/D sense	CTAGGCTGCAGCATCATCAGGATTCTA
	5Aq/B/D antisense	CCGGTAGAATCCTGATGATGCTGCAGC
2A/D	2A/D sense	CTAGGCTGCAGCATCATCACGATTCCA
	2A/D antisense	CCGGTGGAATCGTGATGATGCTGCAGC
2B	2B sense	CTAGGCCGCAGCATCATCACGATTCCA
	2B antisense	CCGGTGGAATCGTGATGATGCTGCGGC
miR172a perfect	miR172a perfect sense	CTAGGATGCAGCATCATCAAGATTCTA
	miR172a perfect antisense	CCGGTAGAATCTTGATGATGCTGCATC
miR172b perfect	miR172b perfect sense	CTAGGATGCAGCATCATCAAGATTCCA
	miR172b perfect antisense	CCGGTGGAATCTTGATGATGCTGCATC
miR172a scrambled	miR172a scrambled sense	CTAGGGCTCTTCAGAAACATGTAATCA
	miR172a scrambled antisense	CCGGTGATTACATGTTTCTGAAGAGCC
miR172b scrambled	miR172b scrambled sense	CTAGGGGAAACTACGCATTACCCTATA
	miR172b scrambled antisense	CCGGTATAGGGTAATGCGTAGTTTCCC

I annealed the oligo pairs according to the protocol in Liu and Axtell <sup>219</sup>. I used the pGreen\_dualuc\_3'UTR\_sensor plasmid (Figure 3-4) in a double restriction digest as in Table 3-3.

**Chapter 3:** How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?

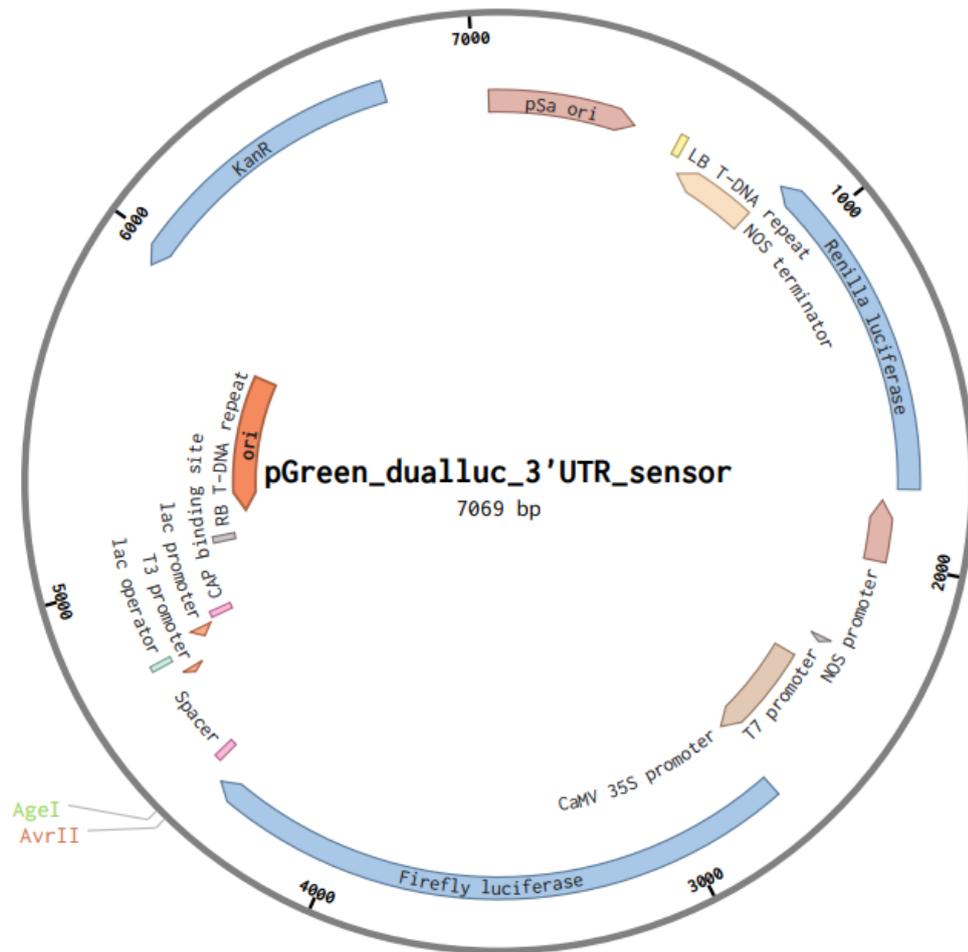


Figure 3-4: Plasmid map of pGreen\_dualuc\_3' UTR\_sensor (total size: 7069 base pairs (bp)). The plasmid contains the following features which are annotated on the map: pSa ori (origin of replication from the bacterial pSa plasmid); LB T-DNA repeat (truncated left border of the nopaline C58 T-DNA repeat); NOS terminator (nopaline synthase terminator and nopaline synthase poly(A) signal); Renilla luciferase gene; NOS promoter (nopaline synthase promoter); T7 promoter (bacteriophage T7 RNA polymerase promoter); CaMV 35S promoter (Cauliflower Mosaic Virus 35S promoter); Firefly luciferase gene; AvrII (AvrII restriction site); AgeI (AgeI restriction site); Spacer (the sequence in the 3' UTR of the Firefly luciferase gene which is replaced by miRNA target sites to be tested); T3 promoter (bacteriophage T3 RNA polymerase promoter); lac operator (from *E. coli*); lac promoter (from *E. coli*); CAP binding site (cyclic AMP receptor protein binding site from *E. coli*); RB T-DNA repeat (right border of the nopaline C58 T-DNA repeat); ori (origin of replication); KanR (Aminoglycoside 3'-phosphotransferase, confers resistance to kanamycin).

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

Table 3-3: Reaction setup for double digestion of the pGreen\_dualLuc\_3'UTR\_sensor plasmid. The volumes and final amounts given are per reaction.

Reagent	Volume ( $\mu$ L)	Final amount
Plasmid DNA	3.64	531.44 ng
NEB 10X CutSmart buffer	5	NA
NEB AvrII	1	5 units
NEB AgeI-HF	0.5	10 units

I ran the digestion reaction at 37 °C for 1 h and then ran the product on a 1% agarose gel at 100 V for 45 min. I excised the required bands, extracted the DNA, and quantified and checked the quality as described in Section 3.3.9.2.1.

I set up 5:1 and 3:1 (insert:vector) ligation reactions for the 5AQ, 2A/D and miR172a perfect target sites to insert the target sites of interest into the pGreen\_dualLuc\_3'UTR\_sensor vector as described in Section 3.3.9.2.1 using 100 ng vector DNA. I set up 500:1 and 300:1 (insert:vector) ligation reactions for the 5Aq/B/D, 2B, miR172b perfect, miR172a scrambled and miR172b scrambled target sites to insert the target sites of interest into the pGreen\_dualLuc\_3'UTR\_sensor vector as described in Section 3.3.9.2.1 using 100 ng vector DNA.

I transformed 5-alpha competent *E. coli* (high efficiency) (New England Biolabs) with the ligation reaction product. I transformed the cells according to the manufacturer's instructions<sup>218</sup> using 10  $\mu$ L ligation reaction, except I used 100  $\mu$ L cells per reaction, added 450  $\mu$ L of SOC medium for recovery, and I altered the recovery period to 1.5 h. I spread 100  $\mu$ L on a 30 mg/mL kanamycin LB agar plate. I incubated the plates overnight at 37 °C.

I created 10 mL (5AQ target) or 5 mL (miR172a perfect, miR172a scrambled, miR172b perfect, miR172b scrambled, 5Aq/B/D, 2A/D and 2B targets) starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.9.2, except the 2A/D and miR172a perfect target starter cultures were incubated at 250 RPM. I validated the plasmids using ONT sequencing as described in Section 3.3.9.2.1.

### 3.3.9.2.3 Transformation into *A. tumefaciens*

Transformation of the constructs into *A. tumefaciens* was performed with the help of Oscar Carey-Fung (University of Melbourne, Australia).

We transformed all plasmids into electro-competent cells prepared in section 3.3.9.1. The protocol used was from Eugenio Butelli (John Innes Centre, UK) adapted by Oscar Carey-Fung (University of Melbourne, Australia). We thawed 50  $\mu$ L aliquots of electrocompetent *A. tumefaciens* cells on ice. Once thawed, we added 100 ng of the relevant plasmid and 100

**Chapter 3:** *How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

ng pSOUP<sup>220</sup> (Figure 3-5, a gift from Matthew Downie (JIC, UK)) and mixed gently. We transferred the cell/DNA mix to the bottom of a chilled 2 mm electroporation cuvette (Geneflow), tapping gently to remove bubbles. We transferred the cuvette to a Gene Pulser (BIO-RAD) and applied a single 2.5 kV electrical pulse (25  $\mu$ F capacitor, 400 ohms, field strength = 12.5 kV/cm). The time reading was > 9.3. We added 1 mL SOC and mixed the cells by gently pipetting up and down. We transferred the cells to a 1.7 mL microcentrifuge tube and incubated them at 28 °C with shaking for 2 h.

We plated 100  $\mu$ L culture on RTG (50  $\mu$ g/mL rifampicin, 5  $\mu$ g/mL tetracyclin, 25  $\mu$ g/mL gentamycin) (pSOUP only culture) and RTGK (50  $\mu$ g/mL rifampicin, 5  $\mu$ g/mL tetracyclin, 25  $\mu$ g/mL gentamycin, 25  $\mu$ g/mL kanamycin) (for all other cultures) LB agar plates. We incubated the plates at 28 °C for 4 d.

**Chapter 3:** How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by microRNA172?

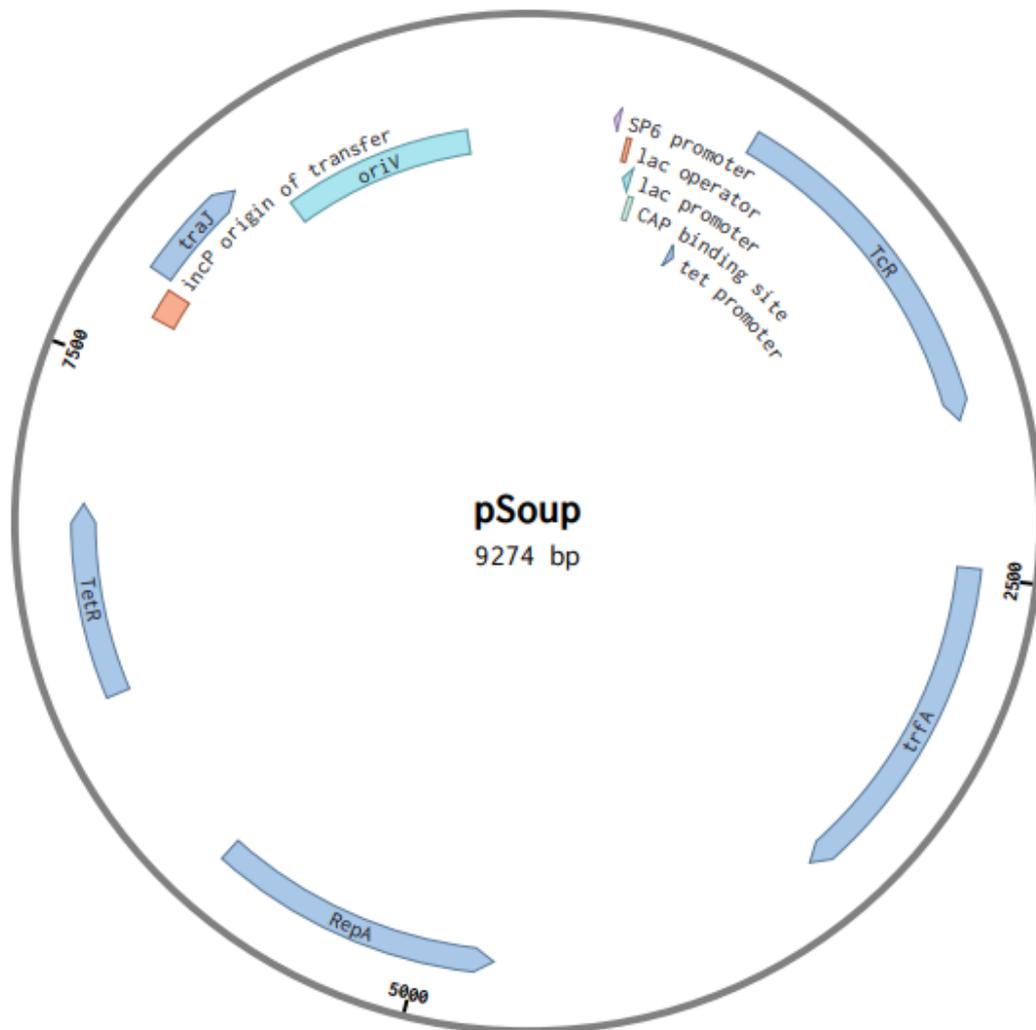


Figure 3-5: Plasmid map of pSoup (total size: 9274 base pairs (bp)). The plasmid contains the following features which are annotated on the map: SP6 promoter (bacteriophage SP6 RNA polymerase promoter); lac operator (from *E. coli*); lac promoter (from *E. coli*); CAP binding site (cyclic AMP receptor protein binding site from *E. coli*); tet promoter (tetracycline efflux protein promoter from *E. coli*); TcR (tetracycline resistance gene); trfA (trans-acting replication protein); RepA (plasmid pSa replication protein); TetR (regulatory protein for tetracycline resistance); incP origin of transfer; TraJ (transfer regulator); oriV (origin of replication for incP plasmids).

We picked single colonies and used them to inoculate 10 mL RTG (50 µg/mL rifampicin, 5 µg/mL tetracyclin, 25 µg/mL gentamycin) (pSOUP only culture) and RTGK (50 µg/mL rifampicin, 5 µg/mL tetracyclin, 25 µg/mL gentamycin, 25 µg/mL kanamycin) (for all other

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

cultures) LB starter cultures which we then incubated overnight at 28 °C 280 RPM. We made glycerol stocks from these cultures as described in Section 3.3.9.2.1.

### 3.3.9.3 Infiltrations

Infiltrations were done with the help of Oscar Carey-Fung (replicate 1) and Max Jones (replicate 2).

The assay was based on that published in Liu and Axtell <sup>219</sup>. Two repetitions of two experiments were carried out as laid out in Table 3-4 and Table 3-5.

*Table 3-4: List of co-infiltration combinations used in dual luciferase experiment 1. Nicotiana benthamiana leaves were co-infiltrated with two cultures as described in the Culture 1 and Culture 2 columns. Each culture contained Agrobacterium tumefaciens transformed with either one or two plasmids (separated by a '+' in the culture description). A. tumefaciens in culture 1 of each co-infiltration contained the pSOUP plasmid, and some contained an additional plasmid which constitutively overexpresses a microRNA, in this case miR172a. A. tumefaciens in culture 2 of each co-infiltration contained the pSOUP plasmid, and some contained an additional dual luciferase reporter plasmid which constitutively overexpresses Renilla luciferase and Firefly luciferase containing a miRNA binding site.*

Co-infiltration ID	Culture 1	Culture 2
1	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::5AQ + pSOUP
2	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::5Aq/B/D + pSOUP
3	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::2A/D + pSOUP
4	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::2B + pSOUP
5	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172a perfect + pSOUP
6	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172a scrambled + pSOUP
7	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172b perfect + pSOUP
8	pGreen_GUS_competitor-miR172a + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172b scrambled + pSOUP
9	pSOUP	pSOUP
10	pGreen_GUS_competitor-miR172a + pSOUP	pSOUP
11	pSOUP	pGreen_dualuc_3'UTR_sensor + pSOUP

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

Table 3-5: List of co-infiltration combinations used in dual luciferase experiment 2. *Nicotiana benthamiana* leaves were co-infiltrated with two cultures as described in the Culture 1 and Culture 2 columns. Each culture contained *Agrobacterium tumefaciens* transformed with either one or two plasmids (separated by a '+' in the culture description). *A. tumefaciens* in culture 1 of each co-infiltration contained the pSOUP plasmid, and some contained an additional plasmid which constitutively overexpresses a microRNA, in this case miR172b. *A. tumefaciens* in culture 2 of each co-infiltration contained the pSOUP plasmid, and some contained an additional dual luciferase reporter plasmid which constitutively overexpresses *Renilla luciferase* and *Firefly luciferase* containing a miRNA binding site.

Co-infiltration ID	Culture 1	Culture 2
1	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::5AQ + pSOUP
2	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::5Aq/B/D + pSOUP
3	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::2A/D + pSOUP
4	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::2B + pSOUP
5	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172a perfect + pSOUP
6	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172a scrambled + pSOUP
7	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172b perfect + pSOUP
8	pGreen_GUS_competitor-miR172b + pSOUP	pGreen_dualuc_3'UTR_sensor::miR172b scrambled + pSOUP
9	pSOUP	pSOUP
10	pGreen_GUS_competitor-miR172b + pSOUP	pSOUP
11	pSOUP	pGreen_dualuc_3'UTR_sensor + pSOUP

We inoculated 10 mL RTG (50 µg/mL rifampicin, 5 µg/mL tetracyclin, 25 µg/mL gentamycin) (pSOUP only culture) and RTGK (50 µg/mL rifampicin, 5 µg/mL tetracyclin, 25 µg/mL gentamycin, 25 µg/mL kanamycin) (for all other cultures) LB starter cultures with the relevant glycerol stocks and incubated them overnight at 28 °C, 280 RPM.

We quantified the OD<sub>600</sub> for the starter cultures using an Eppendorf Biophotometer and inoculated 50 mL (pGreen\_GUS\_competitor plasmids), 12/20 mL (pSOUP only plasmid, replicate 1/replicate 2) and 5/10 mL (pGreen\_dualuc\_3'UTR\_sensor plasmids, replicate

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

1/replicate 2) RTG/RTGK LB secondary cultures with the starter cultures, normalising the OD<sub>600</sub> to 0.01 or 0.02. We incubated the secondary cultures overnight at 28 °C, 200 RPM.

We centrifuged secondary cultures for 15 min, 3,738 RCF. We removed the supernatant and resuspended cells in 15 mL per culture of agroinfiltration solution (10 mM MgCl<sub>2</sub>, 10 mM MES (pH 5.5) by vortexing. We centrifuged the cells for 15 min, 3,738 RCF. We removed the supernatant and resuspended the cells in 50 mL (pGreen\_GUS\_competitor plasmids), 20 mL (pSOUP only plasmid) and 10 mL (pGreen\_dualuc\_3'UTR\_sensor plasmids) agroinfiltration solution with 200 mM acetosyringone. We incubated the cultures at room temperature for 2 h with gentle shaking (approximately 100 RPM) in the dark. We quantified the OD<sub>600</sub> of the cultures using an Eppendorf Biophotometer and normalised the cultures to an OD<sub>600</sub> of 0.5. We combined equal volumes of cultures to form co-infiltration solutions as in Table 3-4 and Table 3-5.

We grew wildtype *Nicotiana benthamiana* plants in 100% peat in 9 cm pots. Plants were grown at 22 °C and 70% humidity in 16 h photoperiods for 30 d before infiltration. We infiltrated the co-infiltration solutions into the abaxial side of *N. benthamiana* leaves using a 1 mL syringe. We infiltrated three plants (biological replicates) for co-infiltrations 1-8 and 10. Only one co-infiltration combination was used per plant. We infiltrated one plant for co-infiltrations 9 and 11. Per plant, we infiltrated three leaves four times each.

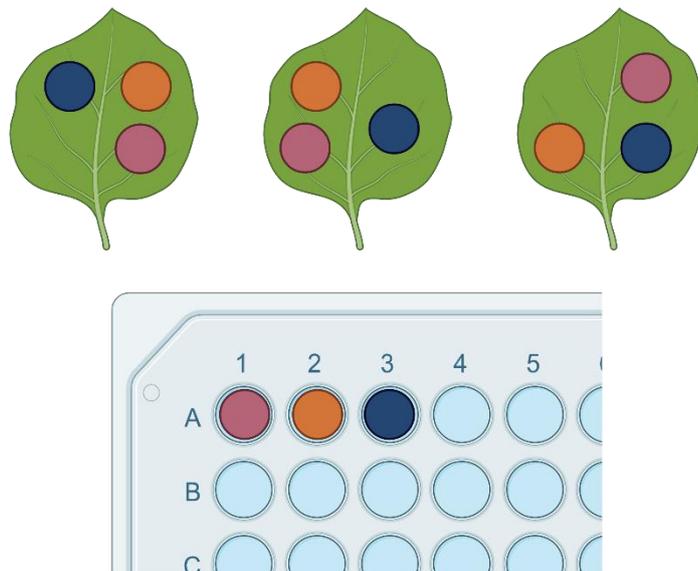
#### 3.3.9.4 Sample collection

Sample collection was done with the help of Oscar Carey-Fung (replicate 1) and Max Jones (replicate 2).

Three days after infiltration, we took samples from the *N. benthamiana* plants. We took three leaf discs from each leaf and pooled them to form three technical replicates for protein analysis

**Chapter 3:** *How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

(Figure 3-6). We placed the samples in wells of a 96-well collection plate containing 200  $\mu$ L PBS and a 70 % ethanol-cleaned 3 mm tungsten carbide bead per well.



*Figure 3-6: Sampling plan for the dual luciferase assay. Three leaves were co-infiltrated per plant (biological replicate). Three leaf discs were taken from each leaf and leaf discs from all three leaves were combined in single wells to create three technical replicates, shown here in pink, orange, and blue respectively. Created in BioRender. Carpenter, S. (2024) BioRender.com/x20m263.*

### 3.3.9.5 Protein assay

I sealed the plate and ground the tissue in a Geno/Grinder (SPEX SamplePrep) at 1000 RPM for 2-4.5 min. I briefly spun down the plate in a centrifuge. I transferred the PBS buffer to a 96-well NBS Microplate (flat bottom clear, white polystyrene) (Corning).

I used the Dual-Glo luciferase assay system (Promega) to quantify Firefly and *Renilla* luciferase activity. I took a background reading for the plate using a GloMax-Multi+ Detection System (Promega) with Instinct Software (one read per well, 1 s reading). I added 75  $\mu$ L Dual-Glo reagent (Promega) to each well and incubated the plate in the dark for 10-15 min with gentle shaking. I used the GloMax-Multi+ Detection System (Promega) to measure Firefly luciferase activity (one read per well, 1 s reading). I added 75  $\mu$ L Dual-Glo Stop & Glo reagent (Promega) to each well and incubated the plate in the dark for 10-15 min with gentle shaking.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

I used the GloMax-Multi+ Detection System (Promega) to measure *Renilla* luciferase activity (one read per well, 1 s reading).

### 3.3.9.6 Statistical analysis

I subtracted the background reading from the Firefly and *Renilla* luciferase readings. I calculated a Firefly/*Renilla* (F-luc/R-luc) ratio from these adjusted readings to control for infiltration efficiency and construct expression. I normalised each miR172 mature sequence F-luc/R-luc ratios by dividing them by the mean spacer F-luc/R-luc ratio (which did not have a miRNA overexpression culture co-infiltrated).

I carried out statistical analyses of normalised ratios using R v4.3.2 R Core Team <sup>123</sup>). I calculated an LMM using the R package ‘lme4’ (Bates, *et al.* <sup>205</sup>). The LMM equation used was  $\log(\text{F-luc/R-luc ratio}) \sim \text{treatment} + (1|\text{replicate})$ . There was a high level of variation between the two experimental replicates, so I included replicate as a random effect in my model. A log transformation of the data was necessary to fit the assumptions of an LMM (normal distribution of residuals and homogeneity of variances). I used the R packages ‘emmeans’ <sup>221</sup>, ‘multcomp’ <sup>222</sup> and ‘multcompView’ <sup>223</sup> to test for differences between treatments and calculate compact letter displays (CLDs).

## 3.3.10 Generating AP2-5A / AP2-2A miR172 target site swap transgenic lines

### 3.3.10.1 Plasmid construction

#### 3.3.10.1.1 Constructing the L0 plasmids

##### 3.3.10.1.1.1 pUC-GW-Kan-AP2-5A PROM+5'UTR L0

I identified upstream sequence of putative regulatory importance for the *T. aestivum* cv ‘Kronos’ AP2-5A gene using an in-house map of conserved cis-regulatory elements in cereals (Jones *et al.*, unpublished data). I used the IWGSC v1.0 cv ‘Chinese Spring’ AP2-5A gene model to BLAST the Kronos assembly from GrainGenes for the AP2-5A sequence <sup>224</sup>. I identified a 2,558 bp region upstream of the gene which I had synthesized by GENEWIZ including Golden Gate adapter sequences which would allow for scarless cloning (sequence in Section A.3). I domesticated the sequence *in silico* before synthesis to remove endogenous BsaI and BpiI recognition sites. The synthesised gene was cloned into a pUC-GW-Kan vector. I received 4 µg lyophilised plasmid which I resuspended in nuclease-free water to make a 10 ng/µL stock solution. I transformed the plasmid into Library Efficiency™ DH5α Competent *E. coli* (ThermoFisher) according to the manufacturer’s instructions <sup>214</sup> (except I used 50 µL cells and 5 ng plasmid and recovered the reaction with 450 µL SOC medium). I

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

inoculated a 30 mg/mL kanamycin LB agar plates with 100  $\mu$ L cells. I incubated the plate overnight at 37 °C.

I created starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.9.2, except the starter cultures were incubated at 250 RPM. I quantified plasmid DNA using a QUBIT dsDNA HS kit (Thermo Fisher) according to the manufacturer's instructions <sup>209</sup>.

**3.3.10.1.1.2 pUC-GW-Kan-AP2-5A CDS L0**

GENEWIZ synthesised the *T. aestivum* cv 'Kronos' AP2-5A CDS (coding sequence), including Golden Gate adapters which would allow for scarless cloning (sequence in Section A.3). I used the IWGSC v1.0 cv 'Chinese Spring' AP2-5A gene model to BLAST the Kronos assembly from GrainGenes for the AP2-5A sequence <sup>224</sup>. I domesticated the sequence *in silico* before synthesis to remove endogenous BsaI and BpiI recognition sites. The synthesised gene was cloned into a pUC-GW-Kan vector. I received 4  $\mu$ g lyophilised plasmid which I transformed into *E. coli* as described in Section 3.3.10.1.1.1. I centrifuged the cells at 800 RCF for 5 min at RT and removed 900  $\mu$ L supernatant I inoculated a 30 mg/mL kanamycin LB agar plates with 100  $\mu$ L cells. I incubated the plate overnight at 37 °C.

I created starter cultures from single colonies, made glycerol stocks, extracted and quantified plasmid DNA as described in Section 3.3.10.1.1.1.

**3.3.10.1.1.3 AP2-5A CDS+AP2-2A L0**

I used the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) to convert the endogenous miR172 binding site in the pUC-GW-Kan-AP2-5A CDS L0 plasmid to the AP2-2A miR172 target site using the AP2-5A CDS L0 MIR172 F and AP2-5A CDS L0 MIR172 R primers (Section A.1, Table A-1). I conducted the mutagenesis according to the kit manufacturer's instructions <sup>225</sup>. I spun down the transformed cells at 800 RCF for 5 min. I removed 900  $\mu$ L supernatant and resuspended the cells by gentle inversion. I inoculated 30  $\mu$ g/mL kanamycin LB plates the mutagenesis reaction and incubated the plates overnight at 37 °C. I created starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.10.1.1.1.

I quantified and quality checked plasmid DNA using a Nanodrop 1000 (Thermo). I carried out a diagnostic digest of the mutagenized plasmid using *EcoR* V (Roche) according to the manufacturer's instructions <sup>226</sup> (but using 0.5  $\mu$ L enzyme) and ran the restriction digest products on a 1% agarose gel. Plasmids which had the correct banding pattern were Sanger sequenced by GENEWIZ using the M13R primer (Section A.1, Table A-1). The sequencing

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

showed that the first mutagenesis reaction was successful however the resulting plasmid was missing a single thymine from the final sequence.

I carried out a second round of site-directed mutagenesis on the AP2-5A CDS+AP2-2A L0 plasmid using the SDM round 2 Fwd and SDM round 2 Rev primers. I used the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's instructions<sup>225</sup>. I inoculated a 30 µg/mL kanamycin LB plate with 100 µL of the mutagenesis reaction, then spun down the remainder at 300 RCF, removed 800 µL supernatant and resuspended the cells, then inoculated a second 30 µg/mL kanamycin LB plate with the remaining 100 µL.

I created starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.10.1.1.1. I quantified and quality checked plasmid DNA using a Nanodrop 1000 (Thermo). I verified the plasmid using a diagnostic digest and Sanger sequencing as described above.

**3.3.10.1.1.4 pUC-GW-Kan-AP2-5A 3'UTR L0**

As for the upstream sequence, I identified downstream sequence of putative regulatory importance for the *T. aestivum* cv 'Kronos' AP2-5A gene using an in-house map of conserved cis-regulatory elements in cereals (Jones *et al.*, unpublished data). I used the IWGSC v1.0 cv 'Chinese Spring' AP2-5A gene model to BLAST the Kronos assembly from GrainGenes for the AP2-5A sequence<sup>224</sup>. I identified a 1,561 bp region downstream of the gene which I had synthesized by GENEWIZ including Golden Gate adapter sequences as recommended by TSL SynBio<sup>227</sup> (sequence in Section A.3). I domesticated the sequence *in silico* before synthesis to remove endogenous BsaI and BpiI recognition sites. The synthesised gene was cloned into a pUC-GW-Kan vector. I received 4 µg lyophilised plasmid which I transformed into *E. coli* as described in Section 3.3.10.1.1.1. I spun down the cells at 800 RCF for 5 min at RT and removed 900 µL supernatant and resuspended the cells. I inoculated two 30 mg/mL kanamycin LB agar plates with 50 µL cells each. I incubated the plates overnight at 37 °C. I created starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.10.1.1.1. I quantified and quality checked plasmid DNA using a Nanodrop 1000 (Thermo).

I used the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) to convert the Golden Gate adapters to ones which would result in scarless cloning using the AP2-5A 3'UTR L0 F and AP2-5A 3'UTR L0 R primers (Section A.1, Table A-1). I conducted the mutagenesis according to the kit manufacturer's instructions<sup>225</sup>. I spun down the transformed cells at 800 RCF for 5 min. I removed 900 µL supernatant and resuspended the cells by gentle inversion. I inoculated 30 µg/mL kanamycin LB plates with 50 µL mutagenesis reaction each

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

and incubated the plates overnight at 37 °C. I created starter cultures from single colonies, made glycerol stocks and extracted plasmid DNA as described in Section 3.3.9.2.1. I quantified and quality checked plasmid DNA using a Nanodrop 1000 (Thermo). I carried out a diagnostic digest of the mutagenized plasmid using PvuII-HF (New England Biolabs) according to the manufacturer's instructions <sup>228</sup> (using 10 units of enzyme) and ran the restriction digest products on a 1% agarose gel. Plasmids which had the correct banding pattern were Sanger sequenced by GENEWIZ using the M13R primer (Section A.1, Table A-1). The sequencing showed that the mutagenesis reaction was successful.

### 3.3.10.1.1 Constructing the L1 plasmids

#### 3.3.10.1.1.1 pICH47742-AP2-5A L1

I constructed the level 1 plasmid using a Golden Gate-based protocol. I combined the following L0 plasmids: 200 ng pICH47742 acceptor (TSL SynBio) <sup>229</sup>, 418.8 ng pUC-GW-Kan-AP2-5A PROM+5'UTR L0, 473 ng pUC-GW-Kan-AP2-5A CDS L0, 338.6 ng pUC-GW-Kan-AP2-5A 3'UTR L0 and 202.4 ng NOS terminator pICSL62001 L0 (TSL SynBio). I used 1.5 µL T4 ligase buffer (New England Biolabs), 1.5 µL 10X BSA, 0.5 µL 400 U/µL T4 DNA ligase (New England Biolabs) and 0.5 µL 10 U/µL BsaI (New England Biolabs). The reaction was cycled through the following programme: 20 s 37 °C, 26 cycles of 3 min at 37 °C and 4 min at 16 °C, 5 min at 50 °C, 5 min at 80 °C, hold at 16 °C. I used 5 µL of the reaction to transform 5-alpha competent *E. coli* (high efficiency) (New England Biolabs) according to the manufacturer's instructions <sup>218</sup>. I used 100 µL of cells to inoculate a 100 µg/mL carbenicillin, 0.1 mM IPTG, 60 µg/mL X-Gal LB agar plate. I spun down the remaining cells, removed 800 µL of supernatant and resuspended the cells and used them to inoculate a second 100 µg/mL carbenicillin, 0.1 mM IPTG, 60 µg/mL X-Gal LB agar plate. The plates were incubated overnight at 37 °C. I picked single white colonies and used them to inoculate 100 µg/mL carbenicillin LB starter cultures which I incubated overnight at 37 °C, 250 RPM. I minipreped the starter cultures using a QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions <sup>208</sup>. I quantified and quality checked plasmid DNA using a QUBIT dsDNA HS assay (Thermo Fisher) according to the manufacturer's instructions <sup>209</sup>. Putative L1 plasmids were sequenced by Plasmidsaurus.

#### 3.3.10.1.1.2 pICH47742-AP2-5A-AP2-2A miR172 binding site L1

This plasmid was constructed as described in Section 3.3.10.1.1.1, however 200 ng of AP2-5A CDS+AP2-2A L0 plasmid was used instead of AP2-5A CDS L0.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

**3.3.10.1.2 Constructing the L2 plasmids**

The AP2-5A and AP2-5A-AP2-2A miR172 binding site constructs were cloned into level 2 pGoldenGreenGate plasmids<sup>230</sup> by Mark Smedley using Golden Gate assembly as described by Weber, *et al.*<sup>229</sup> (Figure 3-7 and Figure 3-8, respectively).

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

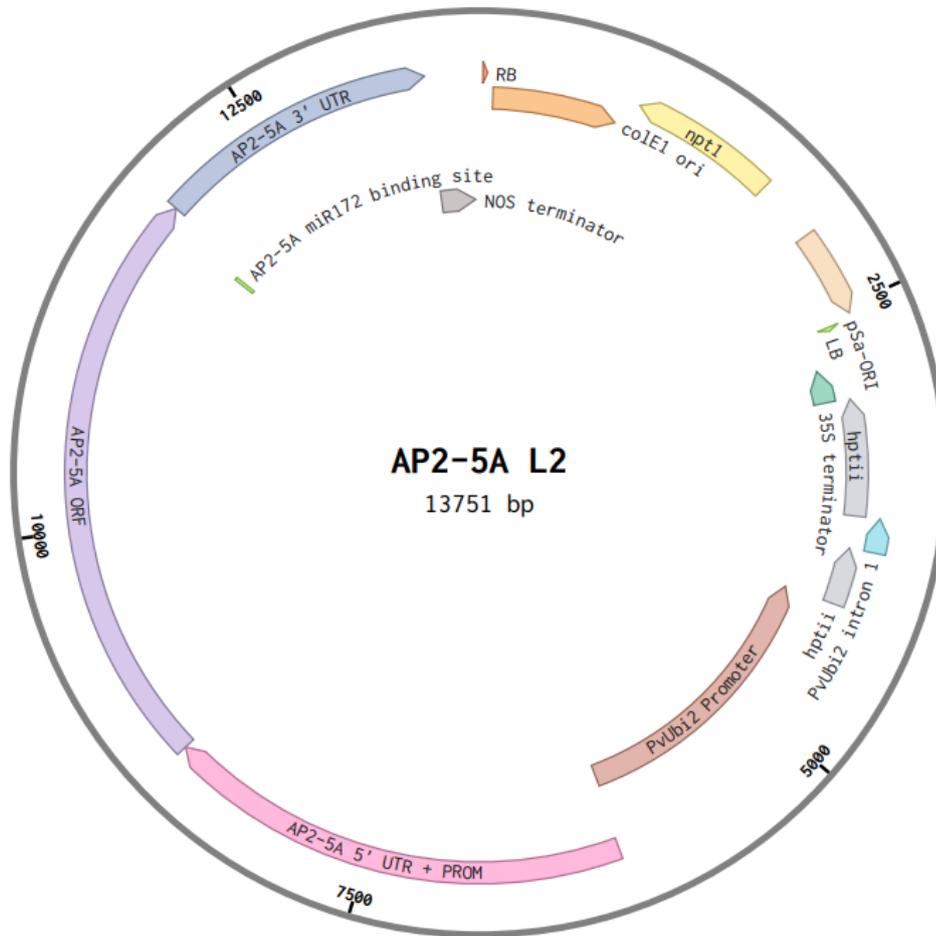


Figure 3-7: Plasmid map of AP2-5A Level 2 (total size: 13, 751 base pairs (bp)). The plasmid contains the following features which are annotated on the map: RB (right border of T-DNA region); colE1 ori (origin of replication for ColE1 plasmids); psa-ORI (origin of replication from the bacterial pSa plasmid); LB (left border of T-DNA region); 35S terminator (Cauliflower Mosaic Virus 35S terminator); hptii (hygromycin phosphotransferase II, confers resistance to hygromycin B); pvUbi2 intron (intron from the ubiquitin2 gene from switchgrass (*Panicum virgatum* L.)); PvUbi2 promoter (promoter from the ubiquitin2 gene from switchgrass); AP2-5A 5' UTR + PROM (5' untranslated region and promoter of the APETALA2-5A (AP2-5A) gene from *Triticum turgidum* cv 'Kronos'); AP2-5A ORF (open reading frame of the AP2-5A gene from *T. turgidum* cv 'Kronos', including introns); AP2-5A mir172 binding site (microRNA172 binding site from the AP2-5A gene from *T. turgidum* cv 'Kronos'); AP2-5A 3' UTR (3' untranslated region of the AP2-5A gene from *T. turgidum* cv 'Kronos'); NOS terminator (nopaline synthase terminator and nopaline synthase poly(A) signal).

**Chapter 3:** How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by microRNA172?

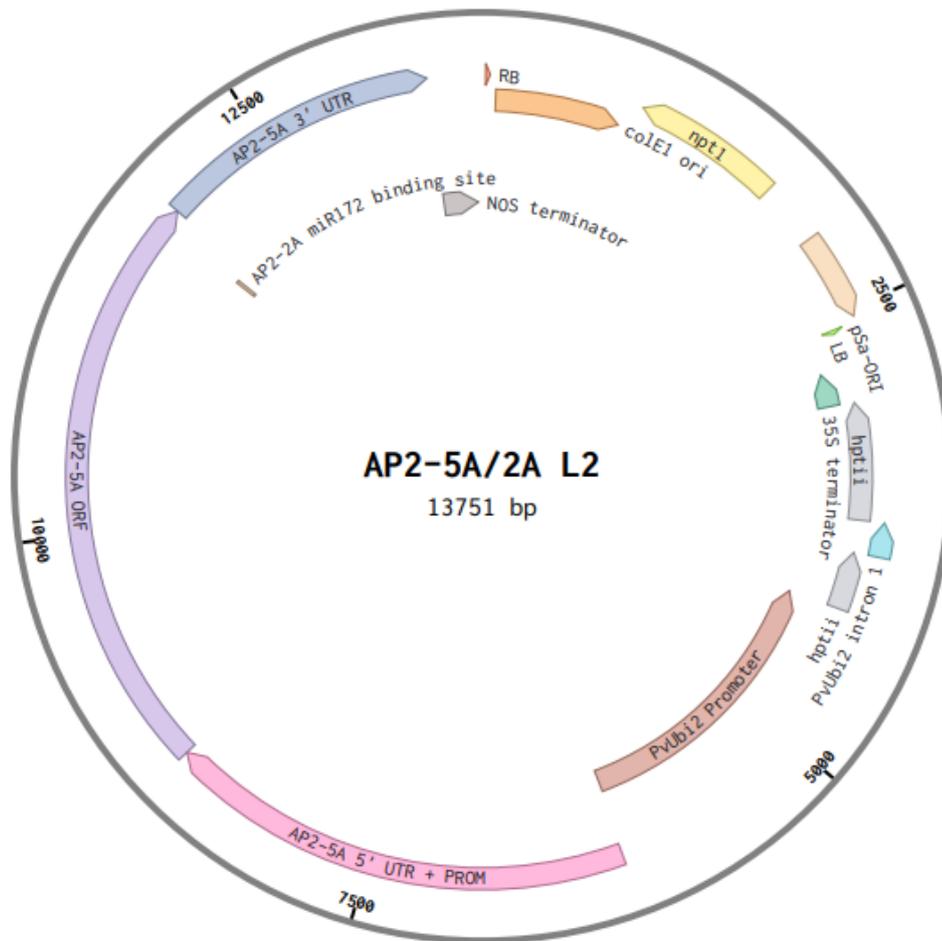


Figure 3-8: Plasmid map of AP2-5A/2A Level 2 (total size: 13, 751 base pairs (bp)). The plasmid contains the following features which are annotated on the map: RB (right border of T-DNA region); colE1 ori (origin of replication for ColE1 plasmids); psa-ORI (origin of replication from the bacterial pSa plasmid); LB (left border of T-DNA region); 35S terminator (Cauliflower Mosaic Virus 35S terminator); hptii (hygromycin phosphotransferase II, confers resistance to hygromycin B); pvUbi2 intron (intron from the ubiquitin2 gene from switchgrass (*Panicum virgatum* L.)); PvUbi2 promoter (promoter from the ubiquitin2 gene from switchgrass); AP2-5A 5' UTR + PROM (5' untranslated region and promoter of the *APETALA2-5A* (*AP2-5A*) gene from *Triticum turgidum* cv 'Kronos'); AP2-5A ORF (open reading frame of the *AP2-5A* gene from *T. turgidum* cv 'Kronos', including introns); AP2-2A mir172 binding site (microRNA172 binding site from the *APETALA2-2A* gene from *T. turgidum* cv 'Kronos'); AP2-5A 3' UTR (3' untranslated region of the *AP2-5A* gene from *T. turgidum* cv 'Kronos'); NOS terminator (nopaline synthase terminator and nopaline synthase poly(A) signal).

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

### 3.3.10.2 Genotyping donor material

I designed this experiment so that the *AP2-5A* constructs with either the endogenous miR172 target site or the *AP2-2A* miR172 target site would be used to complement a knockout of *AP2-5A*. I used tetraploid Kronos TILLING line <sup>31</sup> K3946 which contains a premature stop codon in the fourth exon and has been shown to confer sham ramification and empty lemma phenotypes<sup>32</sup>.

I germinated K3946 seeds on wet filter paper in Petri dishes for two days at 4 °C and two days at RT. The seeds were sown by the JIC Horticultural Services team in 11 cm pots containing John Innes Cereal Mix. I took 2 cm leaf samples approximately two weeks after sowing. The JIC Genotyping Platform extracted DNA from these samples and used the K3946\_Q5A\_01\_FAM, K3946\_Q5A\_01\_HEX and K3946\_Q5A\_01\_COM primers (Section A.1, Table A-1) for KASP genotyping as described in Chen, *et al.* <sup>231</sup>. Plants which were homozygous for the premature stop codon in *AP2-5A* were taken forward for transformation.

## 3.4 Results

### 3.4.1 AP2-2 and AP2-5 form separate phylogenetic clades

Debernardi, *et al.* <sup>42</sup> showed that *AP2-2* and *AP2-5* have partially redundant functions during wheat spike development <sup>42</sup>. To begin to disentangle their functions, I constructed a phylogenetic tree to define their sequence-based relationship (Figure 3-9). This phylogenetic tree was based on the amino acid sequences of AP2-2 and AP2-5 and their orthologues from *O. sativa* ssp. *japonica* and Arabidopsis. AP2-2 and AP2-5 form separate clades.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

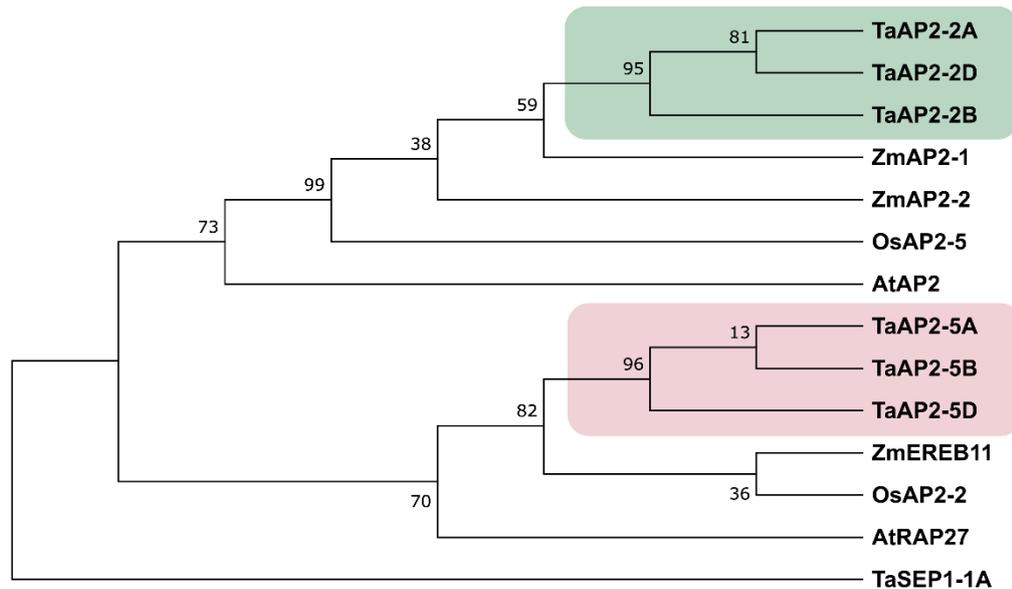


Figure 3-9: Phylogenetic tree of APETALA2-2 (AP2-2) and APETALA2-5 (AP2-5) proteins in wheat and their orthologues in rice, maize and Arabidopsis thaliana. The tree is based on amino acid sequences and is rooted on the *Triticum aestivum* SEPALLATA1-A1 protein. Numbers on branch points indicate confidence values based on bootstrap analysis (100 replicates). A two-letter prefix before the protein name indicates the species: Ta = *T. aestivum* (wheat), Os = *Oryza sativa ssp. japonica* (rice), Zm = *Zea mays*, At = *A. thaliana*. Wheat AP2-2 proteins are highlighted in green; wheat AP2-5 proteins are highlighted in pink.

The distinct AP2-2 and AP2-5 clades have a high level of support ( $\geq 70\%$  bootstrap support, which corresponds to approximately 95% confidence<sup>232,233</sup>), providing strong evidence that they have different orthologues in Arabidopsis, rice and maize. As described in Section 3.2.2, *AtAP2* and *AtRAP27* have different functions, giving rise to the idea that AP2-2 and AP2-5 may not be completely redundant.

### 3.4.2 AP2-2 and AP2-5 have opposite patterns of expression during wheat spike development

Previously, the expression profiles of AP2-2 and AP2-5 have been described as similar, although this was specifically for early stages of spike development up to W3.5 (florete primordium present)<sup>41,42</sup>. Given the results of the phylogenetic analysis above I utilised existing datasets to look at their expression patterns in more granular detail, which may provide some insight into their functions. I analysed an RNA-Seq dataset generated by Dr Nikolai Adamski, Dr Anna Backhaus and Max Jones (JIC, UK). This dataset contains samples from the basal and central sections of developing *T. aestivum* cv ‘Paragon’ spikes during key

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

stages (double ridge (W2.5), terminal spikelet (W4.0) and approximately one week after terminal spikelet (W5.0) <sup>41</sup>). This is a critical period during spike development as important yield-related traits, such as spikelet number, are defined during this time and when floral organs are differentiating. Throughout this time *AP2-5* expression is higher than *AP2-2* expression. These data revealed opposite expression profiles for the two *AP2*-like genes. From W2.5 to W5.0 expression of *AP2-5* decreases, while expression of *AP2-2* increases (Figure 3-10). The three homoeologues of *AP2-2* have significantly different levels of expression at each stage (one-way ANOVA:  $F(2, 9) = 29.55, p < 0.001$  at W2.5;  $F(2, 9) = 81.94, p < 0.001$  at W4.0;  $F(2, 9) = 10.08, p = 0.005$  at W5.0) although show the same expression profile; the TPM values for all *AP2-2* homoeologues increase from W2.5 to W4.0 and from W4.0 to W5.0. The A-genome copy of *AP2-5* is dominant, having the highest level of expression amongst the three homoeologues at all three timepoints. *AP2-5A* expression decreases the most of the three homoeologues in absolute terms from 168.64 TPM at W2.5 to 58.42 TPM at W5.0, while *AP2-5D* expression decreases the most in relative terms, by 77.75%. These opposite *AP2-2* and *AP2-5* expression patterns contradict literature which has suggested they have similar expression profiles <sup>42</sup>.

As this Paragon dataset contradicted previous literature and only includes samples from the base and centre of the spike, I analysed a second dataset generated by Isabel Faci, Max Jones, Dr Neil McKenzie and Katie Long (JIC, UK) to confirm the observed expression patterns. This second RNA-Seq dataset was generated from whole *T. aestivum* cv ‘Chinese Spring’ spikes at seven timepoints from W1.5 (transition apex) to W4.0 (terminal spikelet) <sup>41</sup>. The same gene expression pattern can be seen for *AP2-2* and *AP2-5* as in the Paragon dataset; *AP2-2* expression increases from W1.5 to W4.0, while *AP2-5* expression decreases (Figure 3-11). There are some differences in absolute TPM values between the two datasets. For example, in Chinese Spring at W2.5, *AP2-5A* expression is 58.9 TPM, while in Paragon *AP2-5A* has a higher TPM value of 168.6. *AP2-5A* expression is dominant compared to its two homoeologues, as seen in the Paragon dataset. The balance of expression between the *AP2-2* homoeologues differs between the Paragon and Chinese Spring datasets.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

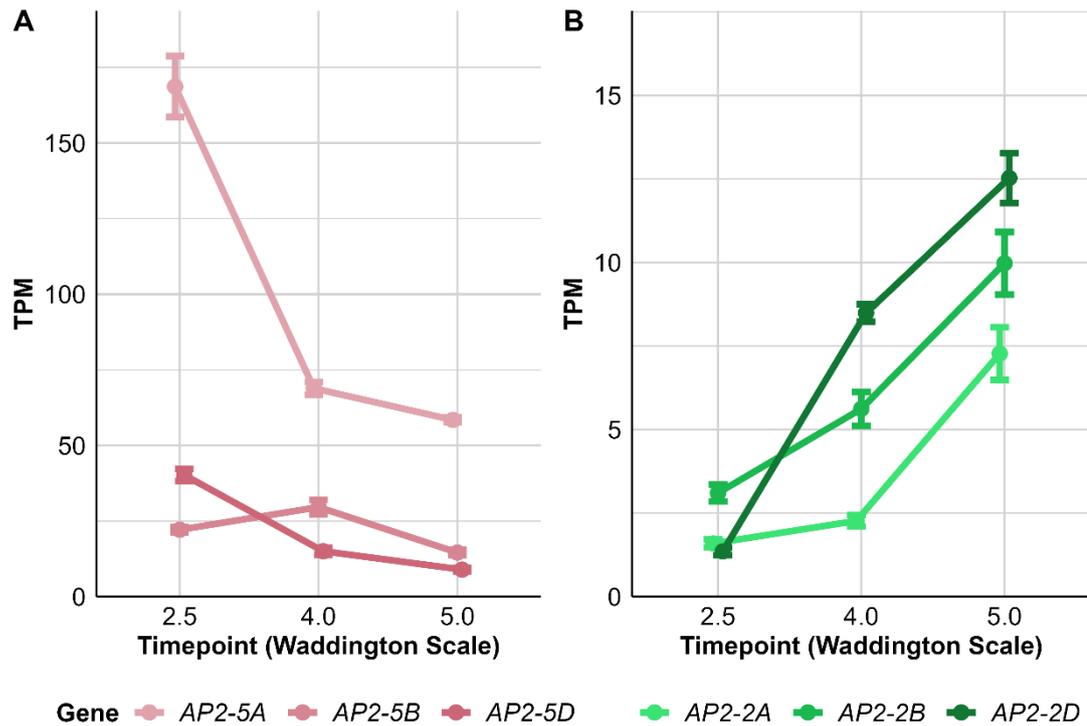


Figure 3-10: Expression of (A) APETALA2-5 (AP2-5) and (B) APETALA2-2 (AP2-2) homoeologues during wheat spike development in Transcripts Per Million (TPM) quantified using mRNA RNA-Seq. RNA-Seq libraries were prepared from basal and central sections of *Triticum aestivum* cv 'Paragon' spikes at W2.5, W4.0 and W5.0 according to the Waddington scale <sup>41</sup>. W2.5 = double ridge stage, W4.0 = stamen primordium present, W5.0 = extending round three sides of ovule, approximated by measuring spike length which is 0.5-1.5 cm during this stage. TPM values are average values from four central and four basal sections for W2.5 and W5.0, and three central and four basal sections for W4.0. Mean TPM values are plotted, and error bars represent the standard error of the mean.

The differences between these two datasets may be the result of the contrasting sampling techniques; the Paragon dataset only includes tissue from the base and centre of the spike therefore does not capture any expression in the apex, while the Chinese Spring dataset is based on whole spike samples. There may also be some differences between the two cultivars which are reflected in this data, for example there may be epigenetic variation within these loci, or there may be small differences in the expression of upstream regulators.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

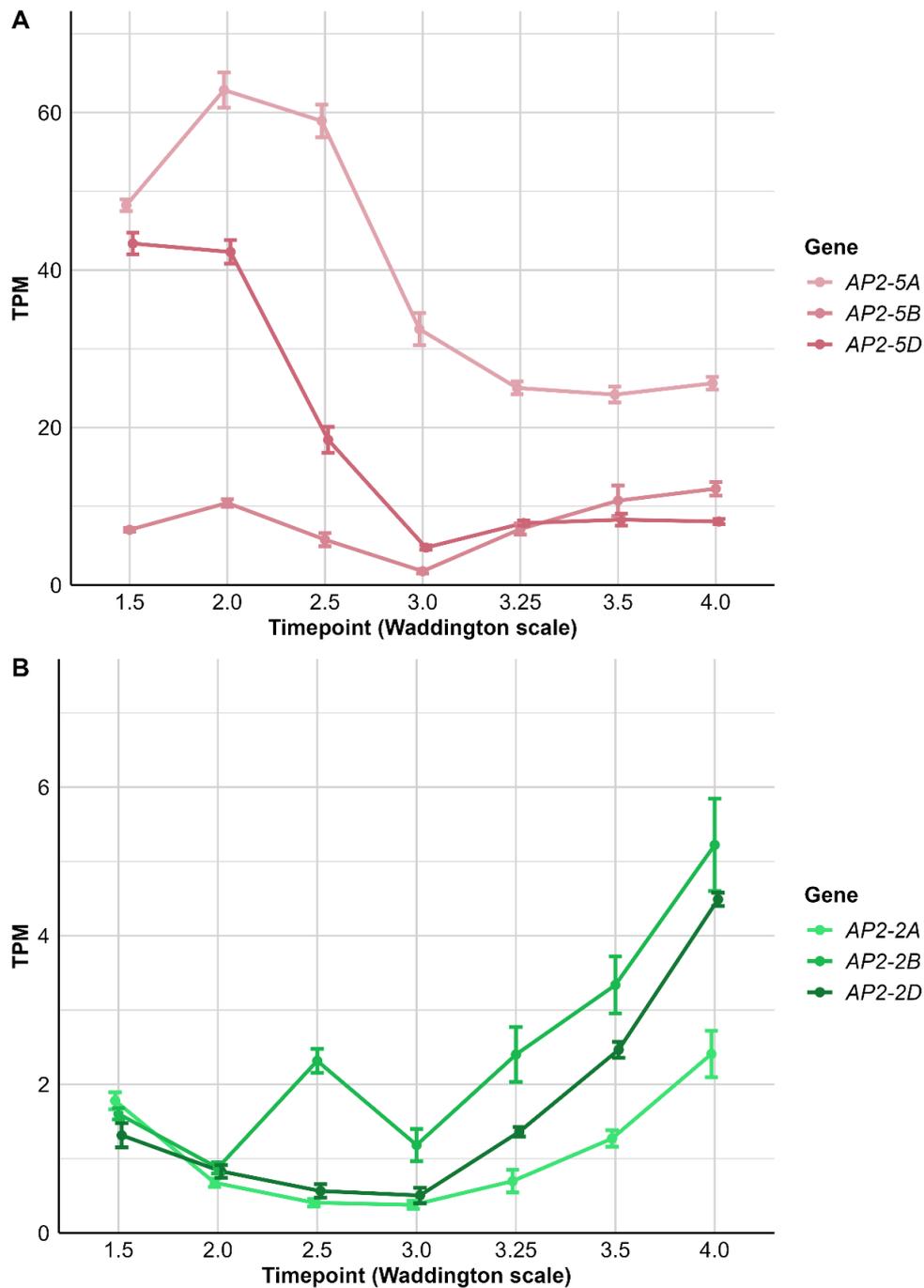


Figure 3-11: Expression of (A) APETALA2-5 (AP2-5) and (B) APETALA2-2 (AP2-2) homoeologues during wheat spike development in Transcripts Per Million (TPM) quantified using mRNA RNA-Seq. RNA-Seq data was generated from whole *Triticum aestivum* cv ‘Chinese Spring’ spikes at seven stages from W1.5 to W4.0 according to the Waddington scale<sup>41</sup>. W1.5 = transition apex, W2.0 = early double ridge stage, W2.5 = double ridge stage, W3.0 = glume primordium present, W3.25 = lemma primordium present, W3.5 = floret primordium present, W4.0 = stamen primordium present. n = 4 biological replicates for all timepoints. Mean TPM values are plotted, and error bars represent the standard error of the mean.

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

The opposing expression patterns of *AP2-2* and *AP2-5* are unexpected as both genes contain miR172 binding sites<sup>42</sup>. Mutations in the miR172 binding sites of either *AP2-2* or *AP2-5* have been shown to lead to phenotypic effects during spike development<sup>32,42,66</sup>. This suggests that both *AP2-2* and *AP2-5* are regulated by miR172 and that this regulation is essential for proper spike development. This observation led me to focus my work on understanding how *AP2-2* and *AP2-5* have opposite expression patterns despite being regulated by the same miRNA.

### 3.4.3 There are three mature miR172 family members in *T. aestivum*

To understand the interaction between AP2 genes and miR172, I sought a better understanding of miR172 in wheat. There are conflicting lists of miR172 family members in the literature<sup>32,94,96,133,196</sup>. The IWGSC v1.0 annotation includes six *MIR172* loci and one miR172 mature sequence (a 5p sequence)<sup>15</sup>. No updated lists of miRNAs in wheat have been released in subsequent IWGSC v1.1 and v2.1 annotations<sup>97,98</sup>. I worked to consolidate and validate these putative miR172 family members. I compiled a list of 15 reported miR172 loci from a comprehensive review of the literature and miRNA databases<sup>32,94,96,132-134,196,234-239</sup>. I attempted to verify these loci by using BLASTn<sup>136</sup> to search for the pre-miRNA sequence (or mature sequence if a pre-miRNA sequence was not available) against the IWGSC v1.0 genome assembly<sup>15</sup>. Three of the putative loci produced no good quality hits against the IWGSC v1.0 assembly (e-value < 0.0001, percentage identify > 90%). I searched for reads corresponding to the putative mature miR172 sequences in publicly available sRNA-Seq datasets<sup>132-134</sup>. There was no evidence of expression of two of the putative miR172 mature sequences in any of the sRNA-Seq datasets. I removed all putative miR172 loci which either produced no good BLASTn hits against the IWGSC v1.0 genome assembly or had no evidence of expression in the sRNA-Seq datasets. From this analysis I compiled a list of eleven *MIR172* loci in the wheat genome which are processed into three unique mature miR172 sequences: miR172a, miR172b and miR172c (Table 3-6). These are produced by seven, three, and one *MIR172* loci, respectively. None of the *MIR172* loci overlap with any IWGSC v1.0 annotations<sup>15</sup> *i.e.*, they are all intergenic. *MIR172B-1-3* and *MIR172A-4-6* are present in the IWGSC v1.0 annotation<sup>15</sup> but *MIR172A1-3* and *MIR172C* are novel loci.

By examining the genes upstream and downstream of the *MIR172* loci, I established that six of the loci which produce miR172a form two homoeologous groups, on chromosome groups 1 and 6, as they are syntenic. The three *MIR172* loci which produce miR172b form a single homoeologous group. There are *MIR172* loci on chromosomes 7A and 7D encoding for miR172a and miR172c mature sequences, respectively. The gene downstream of *MIR172A-7* is homoeologous to the gene downstream of *MIR172C* suggesting these loci are likely to be homoeologues. The fact that the eleven *MIR172* loci fall within four homoeologous groups

***Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?***

suggests that they are highly conserved as they would have been present in the diploid progenitors of wheat. This suggests that they have important functions in developmental processes of evolutionary relevance.

Interestingly, the single copy of *MIR172C* is present on the D-genome, so it would not be present in tetraploid wheat. It also had extremely low levels of abundance (fewer than five reads per sample) in all the sRNA-Seq datasets I included in my analysis. This suggests that miR172c is not necessary for proper wheat spike development. Therefore, I will be focusing on miR172a and miR172b for the remainder of this chapter.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

Table 3-6: microRNA172 (miR172) family members in the *Triticum aestivum* genome. The mature sequence is the most prevalent mature sequence from each MIRNA locus. Underlined nucleotides vary between the three miR172 copies. The MIRNA loci and strands listed are locations within the IWGSC v1.0 assembly<sup>15</sup>. The 'IWGSC v1.0 annotations' column indicates whether a particular MIRNA locus overlaps with any IWGSC v1.0 gene annotations; if it doesn't, it is labelled as 'intergenic'.

Mature miRNA name	Mature miRNA sequence (5' → 3')	MIRNA name	MIRNA loci	MIRNA loci strand	IWGSC v1.0 annotations
miR172a	<u>A</u> GAAUCUUGAUGAUGCUGCA <u>U</u>	MIR172A-1	1A:515766436-515766592	+	Intergenic
		MIR172A-2	1B:566097611-566097811	+	Intergenic
		MIR172A-3	1D:418305926-418306197	+	Intergenic
		MIR172A-4	6A:607452471-607452582	-	Intergenic
		MIR172A-5	6B:702433181-702433290	-	Intergenic
		MIR172A-6	6D:460873268-460873377	-	Intergenic
		MIR172A-7	7A:8252392-8252531	+	Intergenic
miR172b	<u>G</u> GAAUCUUGAUGAUGCUGCA <u>U</u>	MIR172B-1	3A:741660243-741660435	-	Intergenic
		MIR172B-2	3B:818985880-818986071	-	Intergenic
		MIR172B-3	3D:608542517-608542712	-	Intergenic
miR172c	<u>A</u> GAAUCUUGAUGAUGCUGC <u>G</u> <u>U</u>	MIR172C	7D:7295646-7295790	+	Intergenic



*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

mismatch in *AP2-2A* was shown by Debernardi, *et al.*<sup>42</sup> to confer a 14% reduction in plant height. Therefore, miR172 is able to bind and repress *AP2* expression in spite of the presence of these mismatches and G:U wobble base pairs and miRNAs do not require perfect complementarity in order to repress target genes.

The SNP between the ‘*Q*’ and ‘*q*’ alleles of *AP2-5A* is shown in Figure 3-12 in the penultimate nucleotide position (5’ to 3’) in the target site. This single additional G:U wobble confers the modern compact spike phenotype and free-threshing grains<sup>32,66</sup>. This suggests that single nucleotide changes to miRNA-target complementary can have significant effects on miRNA-mediated repression.

A SNP between *AP2-5* and *AP2-2* at the 3’ end of the miR172 binding sites reflects the SNP which differentiates miR172a from miR172b. I hypothesised that due to the ability of SNPs in miRNA binding sites to have large phenotypic effects, miR172a may preferentially target *AP2-5* and miR172b may preferentially target *AP2-2*.

The literature consistently shows that complementary base pairing at the 5’ end of a miRNA to its target is more important than complementary base pairing at the 3’ end<sup>141-143,207</sup>. In mammals, a seed region between nucleotides 2 and 8 on the miRNA is critical for effective binding and regulation<sup>142</sup>. In plants, this seed region may be located slightly differently, perhaps at nucleotides 3-10<sup>141</sup> or 2-12<sup>143</sup>. Seed regions generally do not include the first miRNA nucleotide; in a study of Arabidopsis miRNAs and their targets, mismatches were most commonly located at nucleotide positions 1, 2, 20 and 21<sup>141</sup>. However, in a separate study carried out in *N. benthamiana*, it was shown that a single mismatch at the 5’ end of the miRNA had a significant effect on the RNA and protein abundance of the target compared to a perfectly complementary site<sup>207</sup>. There is some evidence that SNP position affects the strength of phenotypic consequences, for example Houston, *et al.*<sup>240</sup> showed that different alleles of *HvAP2* with SNPs in different positions had different effects on spike density; SNPs in some positions conferred a stronger phenotypic change compared to others. Interestingly, in this allelic series SNPs located within two base pairs of one another have phenotypic consequences of very different magnitudes<sup>240</sup>. This suggests that the effect of specific additional mismatches in any one position in the miR172 binding site of *AP2* genes in wheat must be empirically tested to determine its consequences.

### 3.4.5 Hypotheses to explain the opposite expression patterns of *AP2-2* and *AP2-5*

My aim outlined at the beginning of this chapter was to elucidate a more nuanced understanding of the regulatory relationship between miR172 and *AP2*-like genes in wheat.

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

My initial analyses outlined in sections 3.4.1 to 3.4.4 led me to focus my question on how *AP2-2* and *AP2-5* have opposite expression profiles during wheat spike development despite previous studies showing that they have partially redundant functions and are both regulated by miR172<sup>42</sup>.

I generated three hypotheses to explain the opposite *AP2-2* and *AP2-5* expression profiles observed during wheat spike development:

1. *The expression domains of AP2-like genes and miR172 do not overlap spatially or temporally*

For miR172 to bind to and repress *AP2*-like genes, miR172 and *AP2*-like transcripts must be present in the same cells at the same time. If the miR172 expression domain has greater overlap with *AP2-5* transcripts than *AP2-2* transcripts, miR172 may be able to regulate *AP2-5* to a greater extent, allowing *AP2-2* expression to increase while miR172 is still present.

2. *Different mature miR172 family members target different AP2 genes*

As a result of the analysis laid out in Section 3.4.4, where I showed that different miR172 mature sequences have different levels of complementarity to different *AP2*-like genes, I hypothesise that miR172a may preferentially target *AP2-5* and miR172b may preferentially target *AP2-2*.

3. *AP2-2 and AP2-5 have different baseline levels of transcription*

In Debernardi, *et al.*<sup>32</sup>, they showed that miR172 levels remain approximately constant through the floret primordia, awn elongation and ‘young spike’ stages, while *AP2-5* expression decreases. In MIM172 lines where miR172 abundance is knocked-down, *AP2-5* levels stay high through to at least the awn elongation stage, by which time *AP2-5* expression in wildtype Kronos spikes has decreased<sup>32</sup>. This suggests that miR172 plays an instrumental role in the observed decrease in *AP2-5* expression during spike development. It may be the case, however, that *AP2-2* transcription increases significantly during wheat spike development, and miR172 simply acts to dampen the abundance of *AP2-2* mRNA.

Due to time constraints, I focused my experimental work on hypotheses 1 and 2. I will discuss how this hypothesis could be tested in Section 3.6.

### 3.4.6 Testing hypothesis 1: The expression domains of AP2-like genes and miR172 do not overlap spatially or temporally

I hypothesised that the miR172 expression domain has greater overlap with *AP2-5* transcripts than *AP2-2* transcripts. This would mean that miR172 would be able to regulate *AP2-5* to a greater extent, allowing *AP2-2* expression to increase while miR172 is still present.

To test this hypothesis, I established the spatial expression pattern of *AP2-2* and *AP2-5*. If there are domains in the developing wheat spike where *AP2-2* is expressed and *AP2-5* is not, this would allow the possibility that miR172 may not overlap with *AP2-2* transcripts to the same extent as *AP2-5* transcripts and therefore represses *AP2-5* to a greater extent than *AP2-2*.

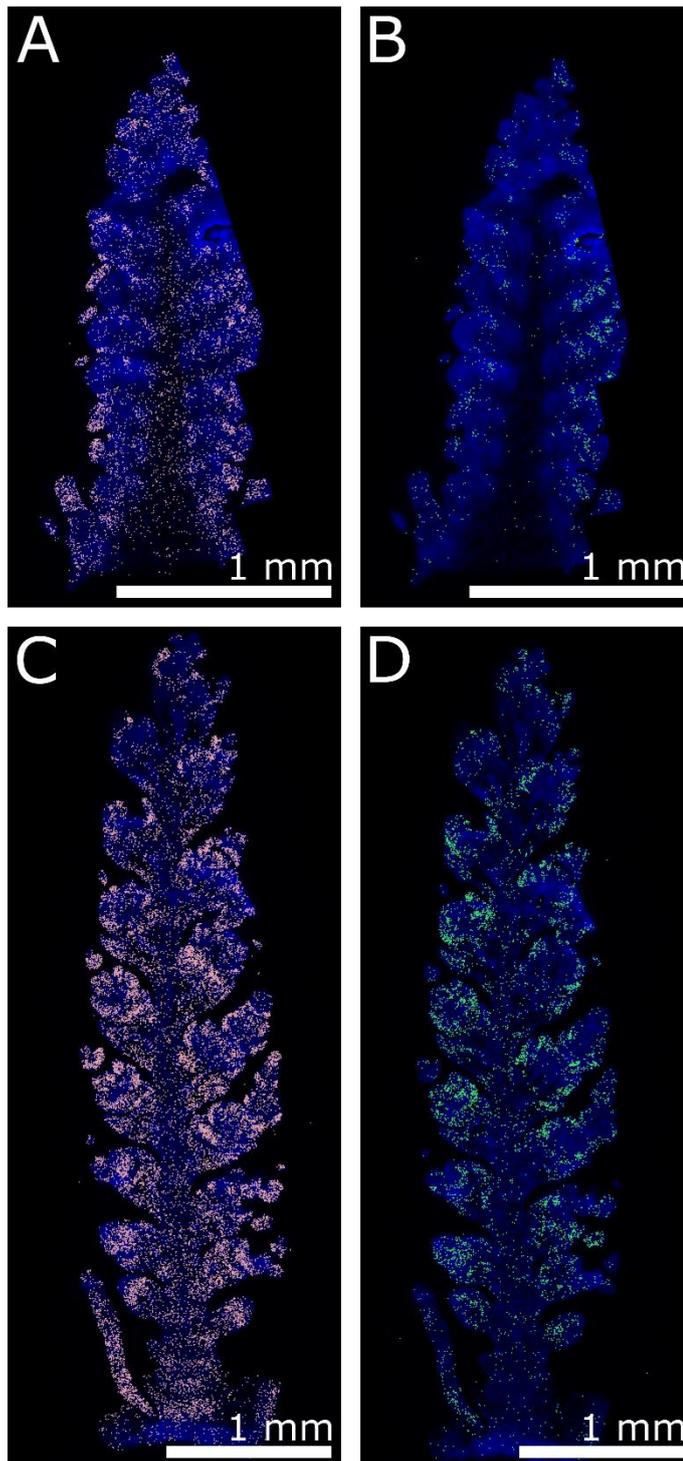
I also looked at changes in miR172 abundance during spike development. The expression profiles in Figure 3-10 and Figure 3-11 show that the decrease in *AP2-5* expression occurs mostly before W3.0, while the increase in *AP2-2* expression occurs from W3.0 onwards<sup>41</sup>. Therefore, if miR172 abundance is lower after W3.0 it would have a lesser impact on *AP2-2* expression which would allow it to increase.

#### 3.4.6.1 The expression domains of AP2-2 and AP2-5 partially overlap

To establish the expression domains of *AP2-2* and *AP2-5* in developing wheat spikes, I analysed data generated by Katie Long (JIC, UK) and Ashleigh Lister (Earlham Institute, UK) who used MERFISH to spatially resolve the expression of over 200 wheat genes (including *AP2-2* and *AP2-5*) in wheat spikes of Paragon at W5.0 (Figure 3-13).

In brief, MERFISH is a spatial transcriptomics technique based on smFISH (single molecule fluorescence *in situ* hybridization). It uses error-robust barcodes to label mRNA transcripts in fixed tissue sections for a panel of up to hundreds of genes. This allowed us to generate a map of gene expression across a wheat spike with sub-cellular resolution.

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*



*Figure 3-13: Images from a MERFISH (Multiplexed Error-Robust Fluorescence in situ Hybridization) assay showing the expression of APETALA2-2 (AP2-2) and APETALA2-5 (AP2-5) in sectioned  $P1^{WT}$  spikes (a *Triticum aestivum* cv ‘Paragon’ near isogenic line containing the ancestral ‘a’ allele of VEGETATIVE TO REPRODUCTIVE TRANSITION A2). Each pink and green dot represents an individual AP2-5 and AP2-2 mRNA transcript, respectively. (A) AP2-5 transcripts in a W4.0  $P1^{WT}$  spike. (B) AP2-2 transcripts in the same W4.0  $P1^{WT}$  spike. (C) AP2-5 transcripts in a W5.0  $P1^{WT}$  spike. (D) AP2-2 transcripts in the*

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

same W5.0 P1<sup>WT</sup> spike. DAPI staining of dsDNA (double-stranded DNA) is shown in blue. W4.0 = stamen primordium present, W5.0 = carpel extending round three sides of ovule, approximated using spike length (0.5-1.5 cm at this stage).

I further analysed the results of analysis Katie Long (JIC, UK) carried out on the MERFISH dataset. Katie Long carried out cell segmentation and UMAP (uniform manifold approximation and projection) clustering (Figure 3-14) to define expression domains within the wheat spike which approximately correspond to different tissue types. 17 cell clusters were generated from this analysis which approximately correspond to tissue types in the developing wheat spike.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

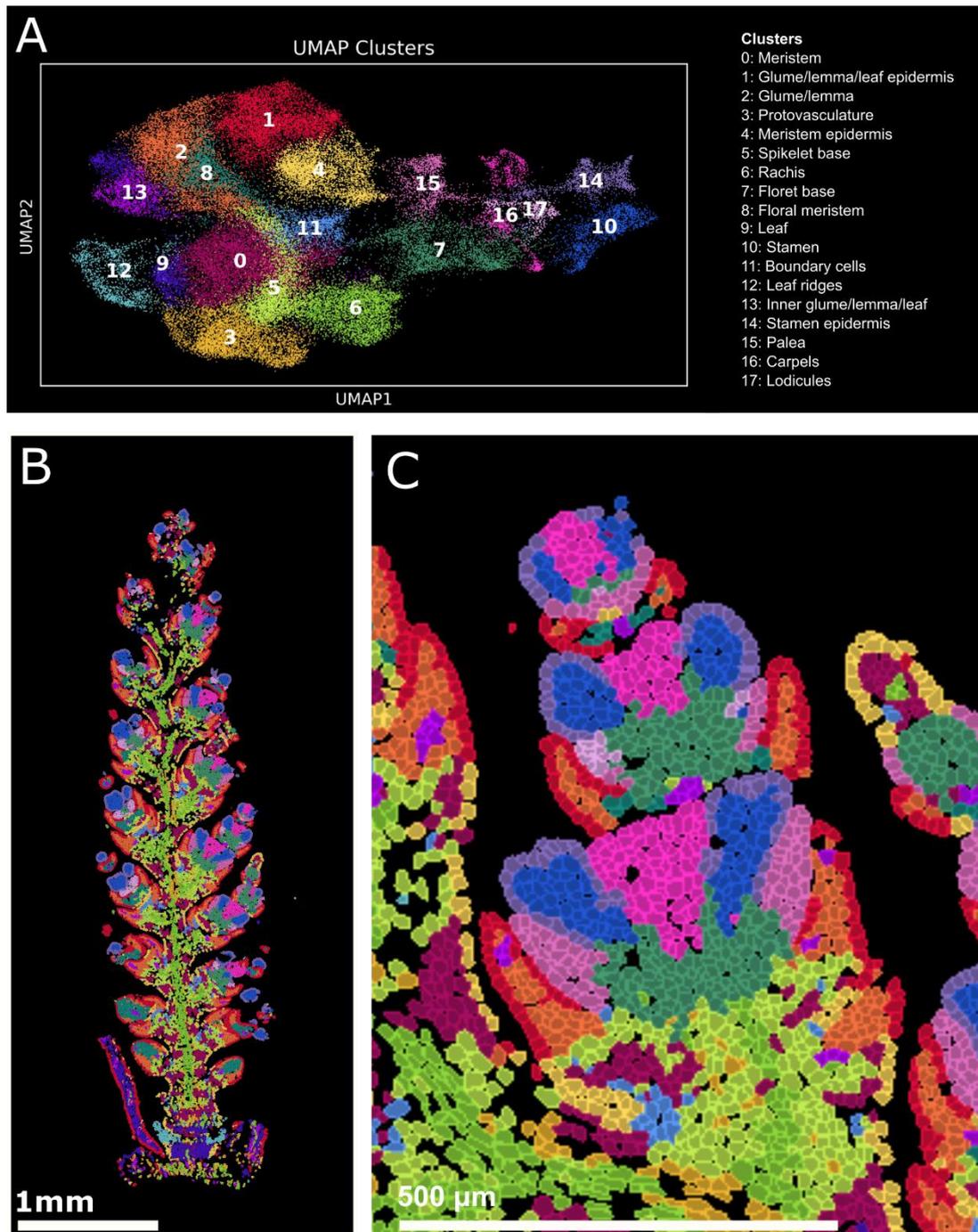


Figure 3-14: UMAP (Uniform Manifold Approximation and Projection) clustering of cells in a *W5.0 P1<sup>WT</sup>* wheat spike (*Triticum aestivum* cv 'Paragon') using data generated by a MERFISH assay ( $n = 1$  biological replicate). *W5.0* = carpel extending round three sides of ovule, approximated using spike length (0.5-1.5 cm at this stage). (A) UMAP projection of cell clusters 0-17 with putative tissue types for each cluster. (B) Wheat spike and (C) spikelet with false colouring of cells according to their assigned cluster (according to the colour key in (A)).

**Chapter 3:** *How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

I calculated the percentage of cells in each cluster which contained both *AP2-2* and *AP2-5* transcripts (Figure 3-15). Over 20% of cells in clusters 2 (glume/lemma), 8 (floral meristem), 10 (stamen) and 13 (inner glume/lemma/leaf) contain both *AP2-2* and *AP2-5* transcripts.

I calculated the mean log normalised transcript count per cell for each cluster (Figure 3-16). *AP2-5* expression was higher in all clusters except for cluster 17 (lodicules). *AP2-5* expression was highest in vegetative tissues: clusters 0 (meristem), 9 (leaf) and 12 (leaf ridges).

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

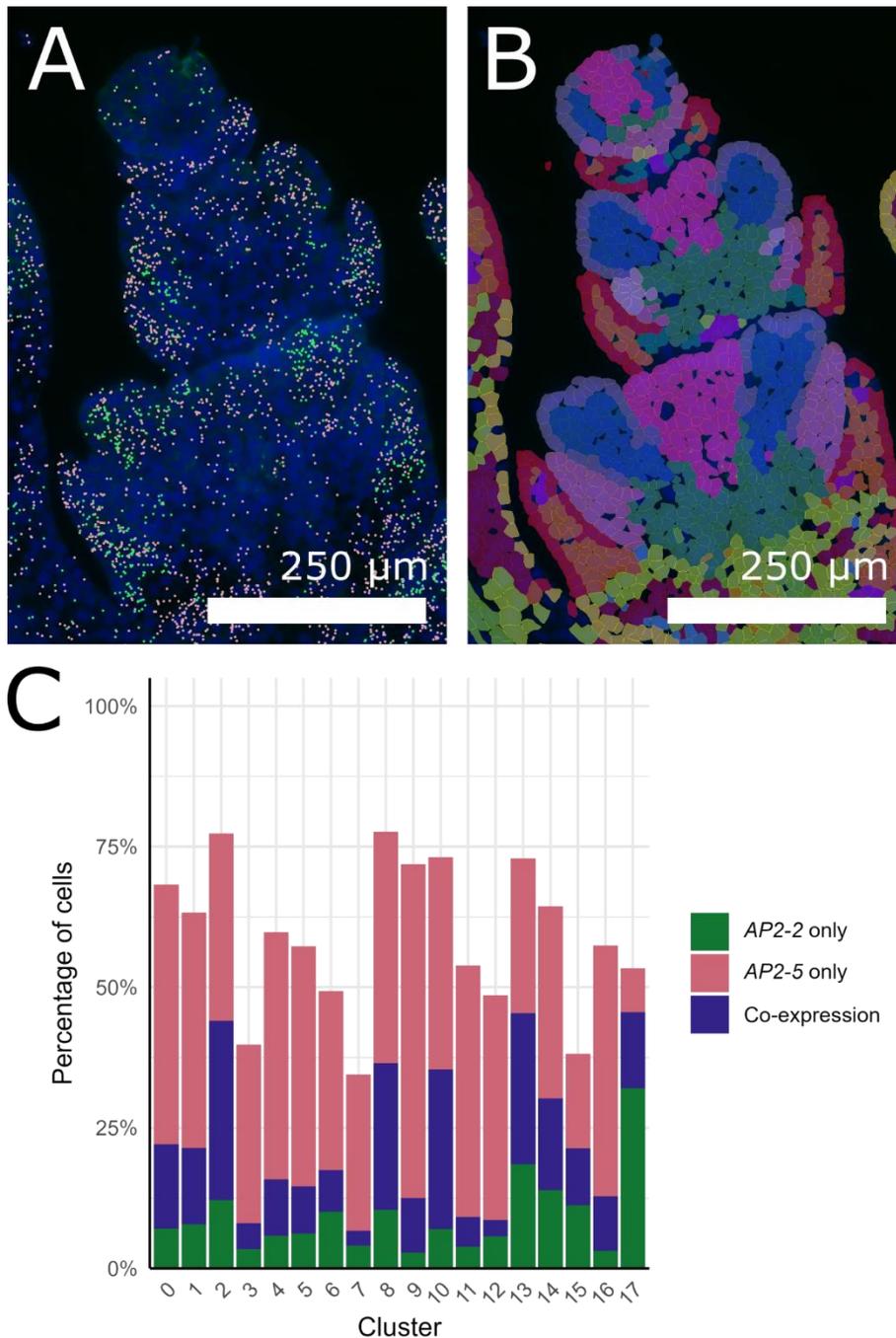


Figure 3-15: Overlap between APETALA2-2 (AP2-2) and APETALA2-5 (AP2-5) transcripts. (A) AP2-2 transcripts in green and AP2-5 transcripts in pink in a W5.0 P1<sup>WT</sup> spikelet using data generated by a MERFISH assay ( $n = 1$  biological replicate). P1<sup>WT</sup> is a *Triticum aestivum* cv 'Paragon' near isogenic line containing the ancestral 'a' allele of VEGETATIVE TO REPRODUCTIVE TRANSITION A2. W5.0 = carpel extending round three sides of ovule, approximated using spike length (0.5-1.5 cm at this stage). (B) Cells in the same W5.0 P1<sup>WT</sup> spikelet with false colouring according to the colour key in Figure 3-14A. (C) The percentage of cells in each cluster of a W5.0 P1<sup>WT</sup> wheat spike which contain AP2-2 transcripts only, AP2-5 transcripts only, or both AP2-2 and AP2-5 transcripts.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

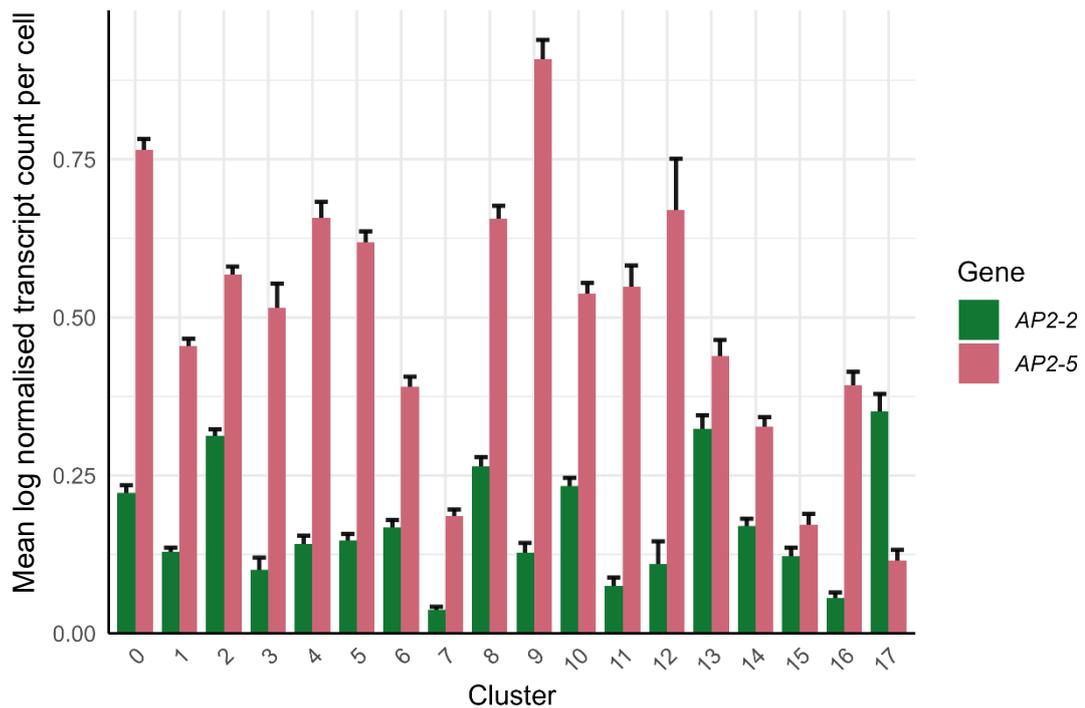


Figure 3-16: Mean log-normalised APETALA2-2 (AP2-2) and APETALA2-5 (AP2-5) transcript count per cell for cell clusters of a W5.0 P1<sup>WT</sup> spike using data generated by a MERFISH assay (n = 1 biological replicate). P1<sup>WT</sup> is a *Triticum aestivum* cv ‘Paragon’ near isogenic line containing the ancestral ‘a’ allele of VEGETATIVE TO REPRODUCTIVE TRANSITION A2. W5.0 = carpel extending round three sides of ovule, approximated using spike length (0.5-1.5 cm at this stage). Error bars depict standard error of the mean.

AP2-2 expression was highest (and higher than AP2-5 expression) in cluster 17 (lodicules). This is consistent with literature which has shown that changes to AP2-2 expression in wheat and barley confer changes to lodicule phenotypes<sup>42,57</sup>. This is the only cluster in which AP2-2 expression is higher than AP2-5 expression. In cluster 17, 32.0% of cells contain AP2-2 transcripts but not AP2-5 transcripts. From this, I could hypothesise that miR172 is absent from lodicules, so allowing AP2-2 expression to increase in this tissue. However, 13.5% of cells still contain both AP2-2 and AP2-5 transcripts. Just 2% of the cells containing AP2-2 transcripts in the W5.0 spike shown in Figure 3-13 are found in cluster 17, which I cannot conclude is enough to claim that miR172 being absent from these regions could result in AP2-2 expression increasing despite the presence of miR172 in the rest of the spike. I cannot conclusively say whether miR172 is co-expressed with AP2-2 without understanding the spatial expression pattern of miR172.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

**3.4.6.2 miR172a and miR172b abundance diverges after terminal spikelet stage during wheat spike development**

To interrogate the temporal expression profile of miR172, I used the sRNA-Seq described in Chapter 2 to quantify the abundance of the different miR172 family members from W2.5 to W5.0<sup>41</sup>. Using sRNA-Seq, as opposed to RT-qPCR which has been used to quantify miR172 abundance previously<sup>32</sup>, allowed me to distinguish between miR172a and miR172b mature sequences. miR172a and miR172b were both identified in this dataset by the miRNA prediction analysis outlined in Section 2.3.2.2. Mature miR172c was also present in the dataset however it was not identified as a miRNA by any of the three prediction pipelines, possibly due to a very low level of abundance.

At W2.5, both mature miR172 sequences show a very low level of abundance (< 8 RPM) (Figure 3-17). miR172a and miR172b abundance increases at W4.0 to 148.14 and 169.98 RPM, respectively. After this, at W5.0 less than a week later, the abundance of miR172a and miR172b has diverged and are significantly different (LMM followed by one-way ANOVA:  $F(1, 5) = 10.37$ ,  $p = 0.02$ ). miR172a remains high at 166.39 RPM, while miR172b abundance decreases to 63.61 RPM.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

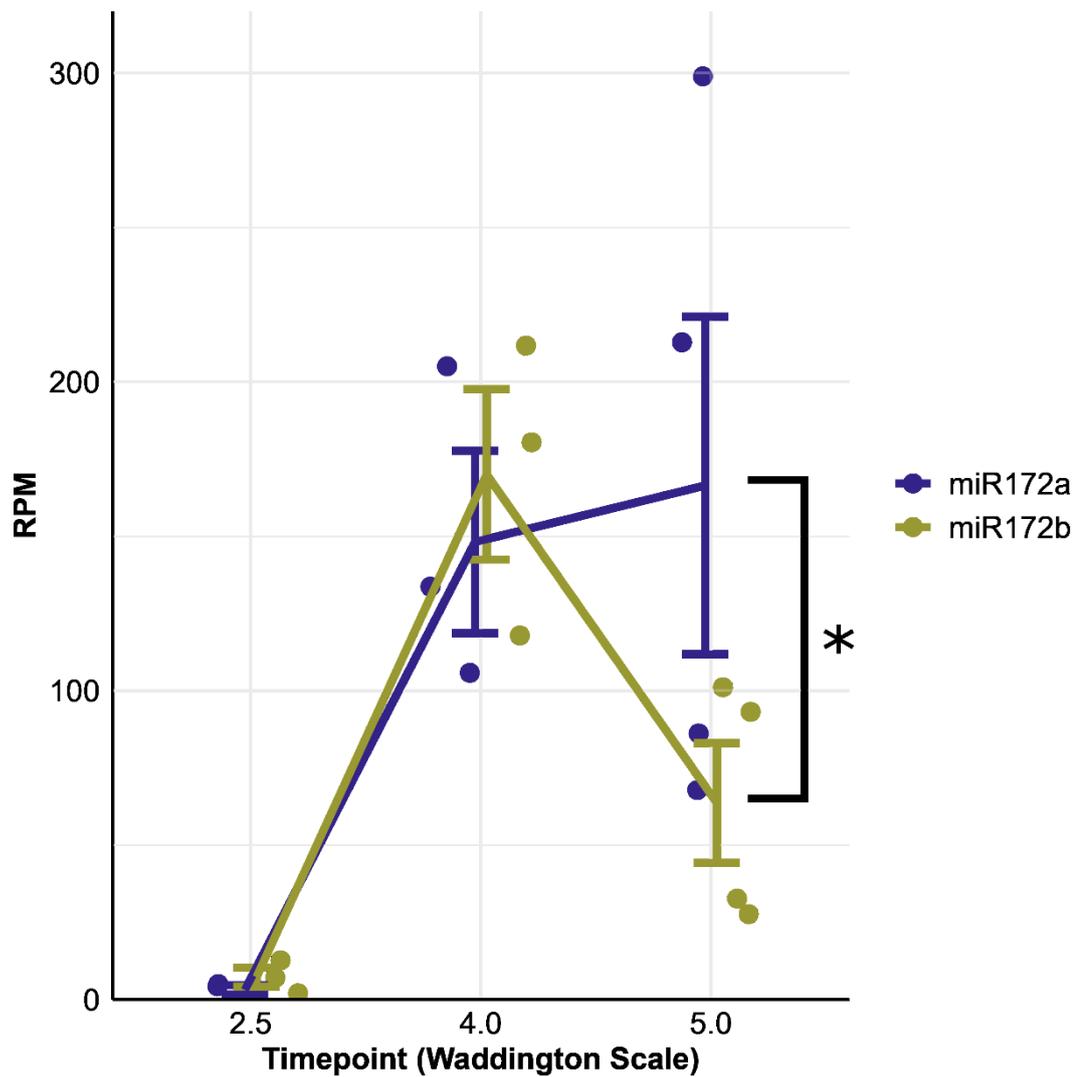


Figure 3-17: Abundance of mature microRNA172 (miR172) family members in developing  $P1^{WT}$  spikes in reads per million (RPM) as quantified using smallRNA-Seq.  $P1^{WT}$  is a *Triticum aestivum* cv ‘Paragon’ near isogenic line containing the ancestral ‘a’ allele of *VEGETATIVE TO REPRODUCTIVE TRANSITION A2*. The timepoints given are according to the Waddington scale <sup>41</sup>. miR172a (purple) and miR172b (green) abundance is shown for each biological replicate by a coloured dot. The lines are plotted against mean abundances and error bars depict the standard error of the mean. W2.5 = double ridge stage, W4.0 = stamen primordium present, W5.0 = extending round three sides of ovule, approximated by measuring spike length which is 0.5-1.5 cm during this stage.  $n = 3$  biological replicates for W2.5 and W4.0,  $n = 4$  biological replicates for W5.0. \*  $p < 0.05$ .

Hypothesis 2 was based on the idea that *AP2-5* expression decreases from W1.5 to W3.0, while *AP2-2* expression increases from W3.0 to W5.0. Therefore, if miR172 expression was higher before W3.0 compared to after, this might explain how *AP2-2* expression increases

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

despite the presence of miR172. However, this sRNA-Seq data appears to show the opposite. Expression at W4.0 of both miR172a and miR172b is at least 24 times higher than at W2.5.

The fact that miR172a and miR172b have distinct expression patterns (miR172a remaining highly expressed and miR172b abundance decreasing at W5.0) lends weight to the idea that these two mature miR172 sequences may have different functions during wheat spike development.

**3.4.6.3 sRNA-Seq reveals an additional miR172 family member which has predicted targets in *T. aestivum***

In addition to the three miR172 mature sequences I identified in section 3.4.3, I identified an additional miR172 family member in the sRNA-seq dataset (Figure 3-18). This mature sequence, miR172d (miRNA candidate-180), can be produced by *MIR172A-4*, *MIR172A-5* and *MIR172A-6* which are also capable of producing miR172a. According to TargetFinder predictions<sup>124-126</sup> generated in Section 2.4.2, miR172d targets a range of genes from different families (Table 3-7). It has no sequence homology to *AP2-2* or *AP2-5* and is not predicted by TargetFinder<sup>124-126</sup> to target these genes. It designated as a member of the miR172 family because the *MIRNA* loci which produce miR172d contain miR172 sequences.

*MIR172A-4* was used by Debernardi, *et al.*<sup>32</sup> in a dual luciferase assay to quantify repression of *AP2-5* by miR172 and they showed that the miRNA produced by *MIR172A-4* was capable of repressing a target with perfect complementarity to miR172a. Therefore, I can conclude that *MIR172A-4* produces miR172a, however I cannot say whether it is this *MIRNA* locus or *MIR172A-5* and/or *MIR172A-6* that produce miR172d. The expression pattern of miR172d is similar to that of miR172b in that it is low at W2.5, increases at W4.0 before decreasing at W5.0.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

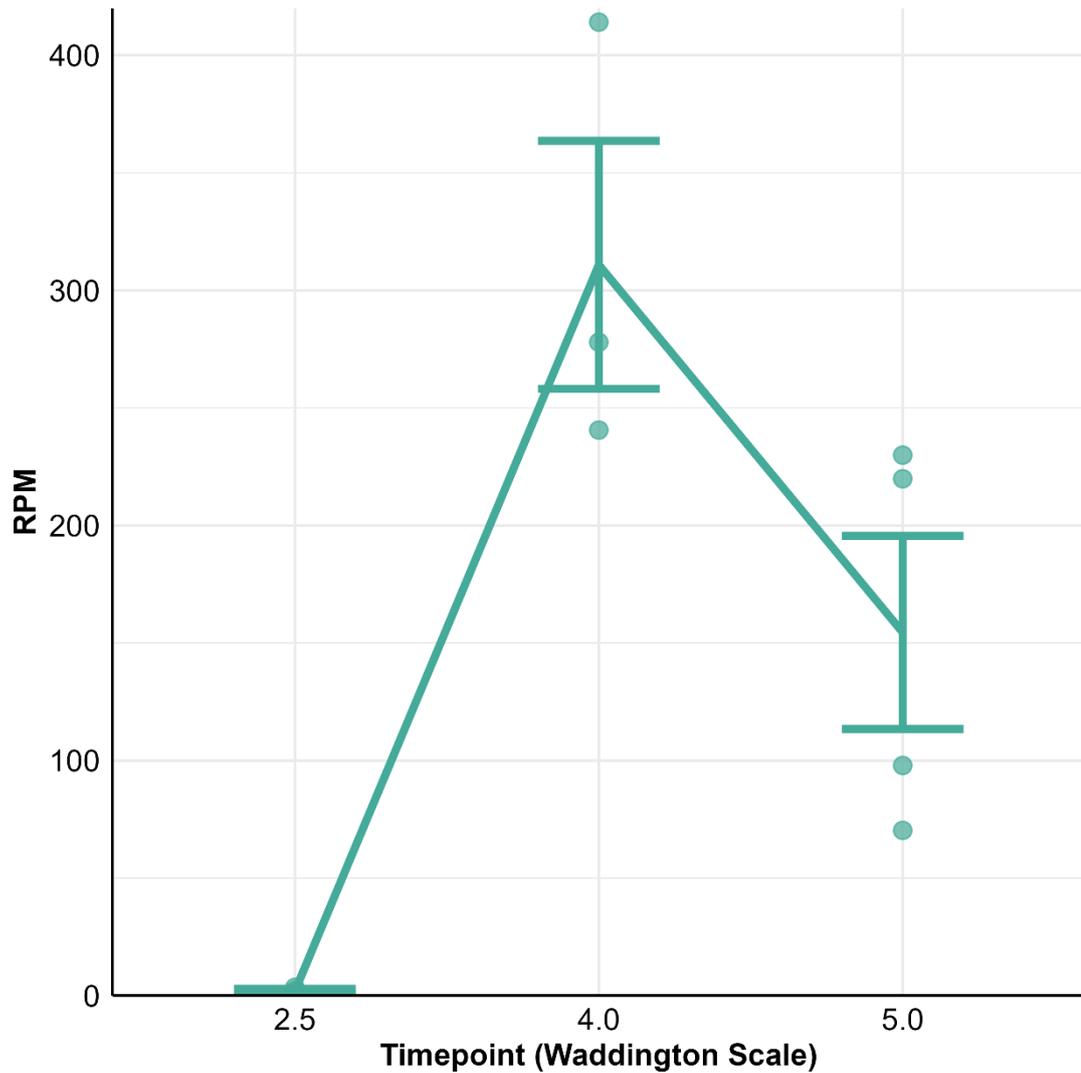


Figure 3-18: Abundance of microRNA172d (miR172d) in developing  $PI^{WT}$  spikes in reads per million (RPM) as quantified using smallRNA-Seq.  $PI^{WT}$  is a *Triticum aestivum* cv ‘Paragon’ near isogenic line containing the ancestral ‘a’ allele of VEGETATIVE TO REPRODUCTIVE TRANSITION A2. The timepoints given are according to the Waddington scale <sup>41</sup>. miR172d abundance is shown for each biological replicate by a dot, the line is plotted against the mean abundance, and error bars depict the standard error of the mean. W2.5 = double ridge stage, W4.0 = stamen primordium present, W5.0 = extending round three sides of ovule, approximated by measuring spike length which is 0.5-1.5 cm during this stage.  $n = 3$  biological replicates for W2.5 and W4.0,  $n = 4$  biological replicates for W5.0.

#### 3.4.6.4 Hypothesis 1 conclusions

The spatial expression patterns of AP2-2 and AP2-5 show that there is only a small section of the developing wheat spike where AP2-2 expression is dominant. Additionally, only a small percentage of AP2-2 transcripts are located in this domain (cluster 17, lodicules). Therefore,

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

it is unlikely that miR172 being less abundant in cluster 17 would produce the differential expression patterns seen for *AP2-2* and *AP2-5*.

miR172 is more abundant at W4.0 than at W2.5 (12.11 RPM at W2.5 compared to 629.00 RPM at W4.0 (mean of combined miR172a, miR172b and miR172d RPM for three biological replicates)), meaning that the differential expression patterns of *AP2-2* and *AP2-5* cannot be explained by a higher level of miR172 while *AP2-5* expression is decreasing and a lower level of miR172 while *AP2-2* is increasing. Interestingly, miR172a and miR172b have different expression profiles. The abundance of both increases from W2.5 to W4.0, and at W5.0 miR172a levels remain high while miR172b abundance decreases. This suggests that these mature miR172 copies may have different functions during wheat spike development, a hypothesis I will test in the next section

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

Table 3-7: Putative targets of microRNA172d (miR172d) in *Triticum aestivum* in the IWGSC v1.0 annotation . Target predictions were generated by TargetFinder<sup>124-126</sup>. The TargetFinder score represents the confidence of the prediction; a lower score confers a higher level of confidence, and the cutoff used was 4. *Oryza sativa* spp. *japonica* (IRGSO-1.0/RAPDB annotation<sup>241</sup>) and *Arabidopsis thaliana* (Araport11 annotation<sup>242</sup>) orthologues according to Ensembl Plants<sup>128</sup> are included for each *T. aestivum* target. The PANTHER<sup>243</sup> protein family is shown for each *T. aestivum* target.

miR172d target in wheat	<i>O. sativa</i> ssp. <i>japonica</i> orthologue	<i>A. thaliana</i> orthologue	PANTHER protein family	TargetFinder score
TraesCS6D01G295000	<i>OsUCL6</i> (Os02g0758800)	N/A	Phytocyanin	3
TraesCS4A01G104100	Os08g0390100	AT2G20010 AT2G25800	Protein unc-13 homologue	3
TraesCS4A01G218900	None	AT1G44120 AT1G77460	Protein CELLULOSE SYNTHASE INTERACTIVE 1/2/3	3.5
TraesCS6A01G195500	<i>OsE2F1</i> (Os02g0537500)	<i>E2F3</i> (AT2G36010) <i>E2F1</i> (AT5G22220)	E2F Family	3.5
TraesCS5D01G090100	<i>OsPAP1d</i> (Os12g0576600) <i>OsPAP1c</i> (Os12g0576700)	AT1G13750	None	3.5
TraesCS5B01G406800	<i>OsSigP2</i> (Os03g0765200)	<i>Plsp2A</i> (AT1G06870) <i>TPP</i> (AT2G30440)	Peptidase S26A, signal peptidase I	3.5
TraesCS7B01G501900	Os03g0709200	<i>LOV1</i> (AT1G10920) AT1G50180 AT1G53350 AT1G58390 AT1G58400 AT1G58410 AT1G58602 AT1G58848	Disease resistance protein, plants	4

**Chapter 3:** How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by microRNA172?

		AT1G59218 <i>CW9</i> (AT1G59620) AT1G59780 AT5G35450 <i>RPP8</i> (AT5G43470) AT5G48620		
TraesCS5D01G030200	<i>OsCCX4</i> (Os12g0624200)	AT5G17850	None	4
TraesCS5D01G030700	<i>OsCCX4</i> (Os12g0624200)	AT5G17850	None	4
TraesCS6A01G375400 (homoeologue to TraesCS6D01G359900)	Os02g0806600	AT2G43160 AT3G59290	None	4
TraesCS6D01G359900 (homoeologue to TraesCS6A01G375400)	Os02g0806600	AT2G43160 AT3G59290	None	4
TraesCS7B01G260800	<i>OsABCA2</i> (Os06g0589300)	N/A	ABC transporter A	4

### 3.4.7 Testing hypothesis 2: Different mature miR172 family members target different AP2 genes

There is a SNP between the mature sequences of miR172a and miR172b which corresponds to a SNP between AP2-2 and AP2-5 miRNA target regions. I hypothesised that miR172a binds preferentially to AP2-5, while AP2-2 binds preferentially to AP2-2.

#### 3.4.7.1 TargetFinder predicts that miR172 targets AP2-5, but not AP2-2

To assess how miR172a and miR172b might bind to different targets, I used the TargetFinder pipeline as outlined in Section 2.3.2.4 to predict their targets *in silico*. This analysis predicted that both miR172a and miR172b target AP2-5B and AP2-5D only, but not AP2-5A or any AP2-2 homoeologues.

This result suggests that the SNP between AP2-2 and AP2-5 at nucleotide 15 of the target site, which confers a G:U wobble between AP2-5 and miR172 and a mismatch between AP2-2 and miR172, may be more important for miR172 binding than the SNP at the last position (Figure 3-12).

#### 3.4.7.2 miR172a represses AP2-2 and AP2-5 target sites with equal efficiency

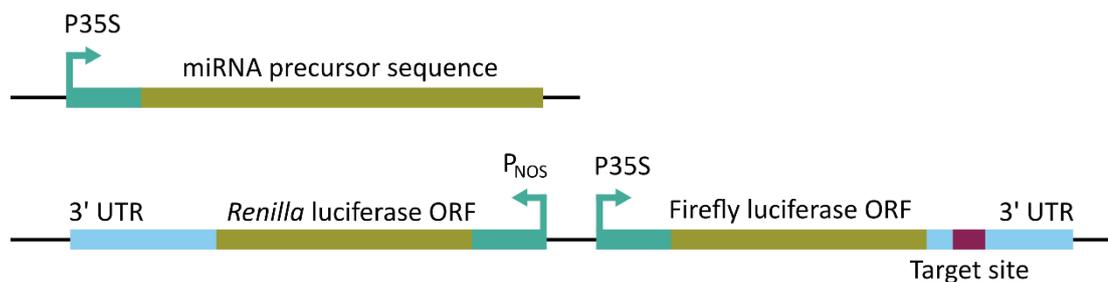
To test the *in silico* predictions made by TargetFinder, I led a project with Max Jones (JIC, UK) and Oscar Carey-Fung (University of Melbourne, Australia) to use a dual luciferase assay to evaluate repression of different target sites by miR172 mature sequences *in vivo*. This experiment was based on that published by Liu and Axtell <sup>219</sup>.

In brief, two constructs from Liu and Axtell <sup>219</sup> (Figure 3-19) were co-infiltrated into *N. benthamiana* leaves. An F-Luc/R-Luc ratio is used to evaluate miRNA repression efficiency; a lower ratio indicates a higher level of repression. I used this system to quantify repression by mature miR172 sequences of the target sites found in AP2-2 and AP2-5 (Figure 3-20).

I included multiple negative and positive controls to verify the validity of the results. The first negative control treatment was a single infiltration of *N. benthamiana* leaves with a dual luciferase construct bearing a spacer target site. This demonstrates the theoretical maximum F-Luc/R-Luc ratio, as no miRNA has been co-infiltrated with the dual-luciferase construct. F-Luc/R-Luc ratios were normalised against the spacer only F-Luc/R-Luc ratio. The positive control was a dual luciferase construct containing target sites with perfect complementarity to either miR172a or miR172b. These should demonstrate the theoretical minimum F-Luc/R-Luc ratio for this experiment. The F-Luc/R-Luc ratios of miR172a/b target sites were significantly

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

lower than the spacer only F-Luc/R-Luc ratio in the presence of constitutively overexpressed miR172a or miR172b (LMM followed by one-way ANOVA:  $t(44) = -7.93$ ,  $p < 0.001$  for miR172a;  $t(44) = -8.89$ ,  $p < 0.001$  for miR172b) (Figure 3-20). There was no significant difference between miR172a and miR172b perfect target site ratios in the presence of miR172a (LMM followed by one-way ANOVA:  $t(44) = -1.26$ ,  $p = 0.94$ ) or miR172b ( $t(44) = 0.40$ ,  $p = 1.00$ ). The second set of negative controls were scrambled target sites, which had the same nucleotide composition as the perfect target sites but in a randomised order. The F-Luc/R-Luc ratios for the scrambled target sites were not significantly different from the spacer target site ratio in the presence of either miR172a (LMM followed by one-way ANOVA: miR172a scrambled target site  $t(44) = -0.75$ ,  $p = 1.00$ ; miR172b scrambled target site  $t(44) = 0.28$ ,  $p = 1.00$ ) or miR172b (LMM followed by one-way ANOVA: miR172a scrambled target site  $t(44) = 0.33$ ,  $p = 1.00$ ; miR172b scrambled target site  $t(44) = -0.03$ ,  $p = 1.00$ ), showing that miR172 does not bind effectively to these sites and their repression is sequence-dependent.



*Figure 3-19: Simplified diagrams of the constructs used in the dual luciferase assay described in Section 3.4.7.2. In the first construct, the gene encoding a miRNA precursor sequence is driven by a constitutive P35S promoter. In the second construct, The Renilla luciferase gene is driven by a constitutive P<sub>NOS</sub> promoter. The 3' untranslated region (UTR) of the Renilla luciferase gene is also included in the construct. In the opposite orientation, the Firefly luciferase gene is driven by a constitutive P35S construct. The 3' UTR of the Firefly luciferase gene is included in the construct and contains a 'miRNA target site' which is flanked by restriction sites, meaning it can easily be modified using traditional cloning techniques.*

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

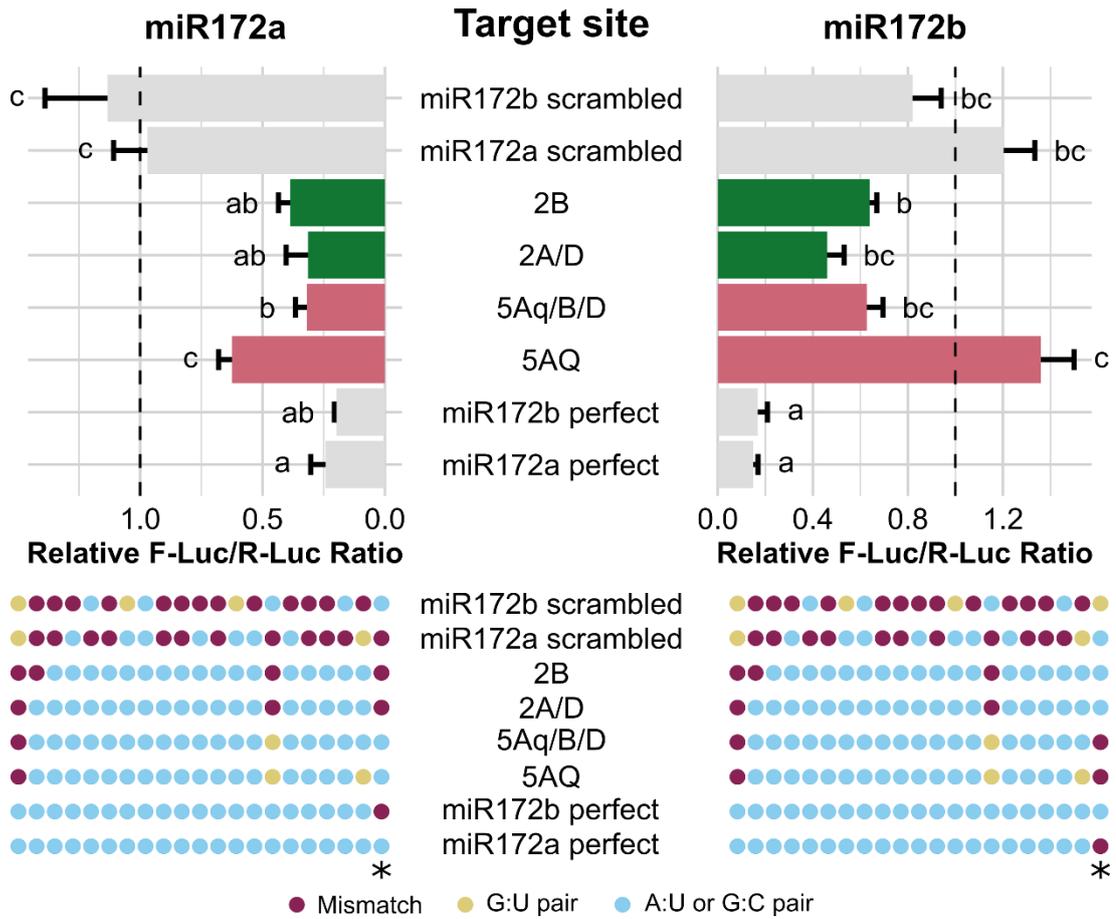


Figure 3-20: Ratio of Firefly luciferase (F-luc) to Renilla luciferase (R-luc) activity for dual luciferase constructs containing different miRNA target sites in the 3' UTR of the Firefly luciferase gene in the presence of microRNA172a (miR172a) or microRNA172b (miR172b). Ratios are normalised against the activity ratio of a dual luciferase construct containing a spacer sequence (shown as a dotted vertical line). Bars and error bars are for the second experimental replicate only. Error bars show standard error of the mean. Bars which do not share any letters are significantly different from one other according to a Tukey test at the 0.05 significance level for both experimental replicates. Complementarity between the target site and miR172a/b is shown below bar charts. Blue circles represent a Watson-Crick base pair (A:U or G:C), yellow circles represent a G:U wobble base pair, and red circles represent a mismatch. The SNP between miR172a and miR172b is denoted with an asterisk. The perfect target sites have perfect complementarity to miR172a or miR172b. The scrambled target sites have the same nucleotide composition as the perfect target sites, but the order has been randomised.

In hexaploid wheat, there are two distinct target sites found in AP2-5 genes. AP2-5B and AP2-5D have identical miR172 binding sites. The ancestral AP2-5A 'q' allele has the same miR172 target site as AP2-5B and AP2-5D<sup>32,66</sup>. The AP2-5Aq/B/D target site produces a

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

significantly different ratio to the spacer target site (LMM followed by one-way ANOVA:  $t(44) = -4.64, p = 0.001$ ), miR172a scrambled target site (LMM followed by one-way ANOVA:  $t(44) = -3.89, p = 0.01$ ) and miR172b scrambled target site (LMM followed by one-way ANOVA:  $t(44) = -4.93, p < 0.001$ ) in the presence of miR172a. The *AP2-5Aq/B/D* target site ratio was not significantly different to the spacer (LMM followed by one-way ANOVA:  $t(44) = -1.82, p = 0.67$ ), miR172a scrambled (LMM followed by one-way ANOVA:  $t(44) = -2.16, p = 0.45$ ), or miR172b scrambled (LMM followed by one-way ANOVA:  $t(44) = -1.80, p = 0.69$ ) target sites in the presence of miR172b. This suggests that transcripts of these genes are repressed by miR172a, but not miR172b.

The *AP2-5A* 'Q' allele has a SNP in the target site compared to the ancestral 'q' allele<sup>32,66</sup>. The ratio for this target site was not significantly different from the spacer (LMM followed by one-way ANOVA: in the presence of miR172a  $t(44) = -1.33, p = 0.92$ , in the presence of miR172b  $t(44) = 1.24, p = 0.94$ ), miR172a scrambled (LMM followed by one-way ANOVA: in the presence of miR172a  $t(44) = -0.58, p = 1.00$ , in the presence of miR172b  $t(44) = 0.90, p = 0.99$ ), or miR172b scrambled target sites (LMM followed by one-way ANOVA: in the presence of miR172a  $t(44) = -1.62, p = 0.79$ , in the presence of miR172b  $t(44) = 1.26, p = 0.94$ ). This demonstrates that the SNP that distinguishes the Q allele from the q allele prevents miR172 from significantly repressing *AP2-5A* expression, which is consistent with the literature<sup>32,66</sup> and supports the idea that this *in vivo* technique reflects *in planta* phenomena.

*AP2-2A* and *AP2-2D* have identical miR172 target sites. In the presence of miR172a this target site produces a significantly different ratio to the spacer (LMM followed by one-way ANOVA:  $t(44) = -4.87, p < 0.001$ ), miR172a scrambled (LMM followed by one-way ANOVA:  $t(44) = -4.12, p = 0.005$ ), or miR172b scrambled (LMM followed by one-way ANOVA:  $t(44) = -5.15, p < 0.001$ ) target sites. In the presence of miR172b, the ratio for this target site is not significantly different to the spacer (LMM followed by one-way ANOVA:  $t(44) = -2.01, p = 0.54$ ), miR172a scrambled (LMM followed by one-way ANOVA:  $t(44) = -2.35, p = 0.34$ ), or miR172b scrambled (LMM followed by one-way ANOVA:  $t(44) = -1.98, p = 0.56$ ) target site ratios. This suggests that *AP2-2A* and *AP2-2D* are significantly repressed by miR172a, but not miR172b.

*AP2-2B* has a unique miR172 target site among the *AP2-2* homoeologues, with an additional SNP at the 5' end of the target site. In the presence of miR172a this target site produces a significantly different ratio to the spacer (LMM followed by one-way ANOVA:  $t(44) = -5.13, p < 0.001$ ), miR172a scrambled (LMM followed by one-way ANOVA:  $t(44) = -4.37, p = 0.002$ , or miR172b scrambled (LMM followed by one-way ANOVA:  $t(44) = -5.41, p < 0.001$ ) target sites. In the presence of miR172b, the ratio for this target site is not

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

significantly different to the spacer (LMM followed by one-way ANOVA:  $t(44) = -2.20$ ,  $p = 0.42$ ), miR172a scrambled (LMM followed by one-way ANOVA:  $t(44) = -2.53$ ,  $p = 0.25$ ), or miR172b scrambled (LMM followed by one-way ANOVA:  $t(44) = -2.17$ ,  $p = 0.44$ ) target site ratios. This suggests that *AP2-2B* is significantly repressed by miR172a, but not miR172b.

I can conclude that the target sites of *AP2-5B*, *AP2-5D*, *AP2-5A* 'q' and all homoeologues of *AP2-2* are targeted by miR172a, but not miR172b. The *AP2-5A* 'Q' allele is not targeted by miR172a or miR172b. In Figure 3-20 it appears that *AP2*-like gene target sites except the *AP2-5A* 'Q' allele target site have lower F-Luc/R-Luc ratios than the scrambled controls. There may be too much variation in this assay to detect a perhaps weaker level of repression by miR172b.

There is no significant difference between the repression of *AP2-2* homoeologues and non-'Q' *AP2-5* homoeologues by miR172a (LMM followed by one-way ANOVA: *AP2-2A/D* target site  $t(44) = -0.23$ ,  $p = 1.00$ ; *AP2-2B* target site  $t(44) = -0.49$ ,  $p = 1.00$ ). This is evidence against the hypothesis that different mature miR172 copies target different *AP2* genes. This is also inconsistent with the *in silico* TargetFinder prediction in Section 3.4.7.1, as the dual luciferase assay shows that miR172a is capable of repressing *AP2-2*.

#### 3.4.7.3 miR172 target site swap transgenics

To interrogate the miR172-*AP2* interaction *in planta*, I developed a strategy to evaluate the effect of swapping the *AP2-5A* miR172 binding site for the *AP2-2A* miR172 binding site. These constructs are currently being transformed into wheat (Figure 3-21A); I will show results to date, however unfortunately due to time constraints we were unable to obtain T<sub>0</sub> plants to evaluate before the submission of this thesis. The constructs contain the full *AP2-5A* genomic sequence, upstream and downstream sequence to ensure any *cis*-regulatory elements are included. The *AP2-5A* construct contains the endogenous *AP2-5A* miR172 target site, while the *AP2-5A-AP2-2A* construct contains the miR172 target site from *AP2-2A*. These constructs will be transformed into the K3946 TILLING line<sup>31</sup> (Figure 3-21B) which contains a premature stop codon which leads to a loss of *AP2-5A* function. K3946 plants have striking phenotypes with one or two bracts between the glumes and first floret and sham ramification. The *AP2-5A* construct should complement the *ap2-5A* loss-of-function mutation in K3946, meaning the transgenic loses the sham ramification and bract phenotypes. The phenotypic implications of transforming the *AP2-5A-AP2-2A* construct into K3946 germplasm should help to establish the effect that the SNPs between the target sites have on miR172 repression in the context of the gene including mRNA folding.

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

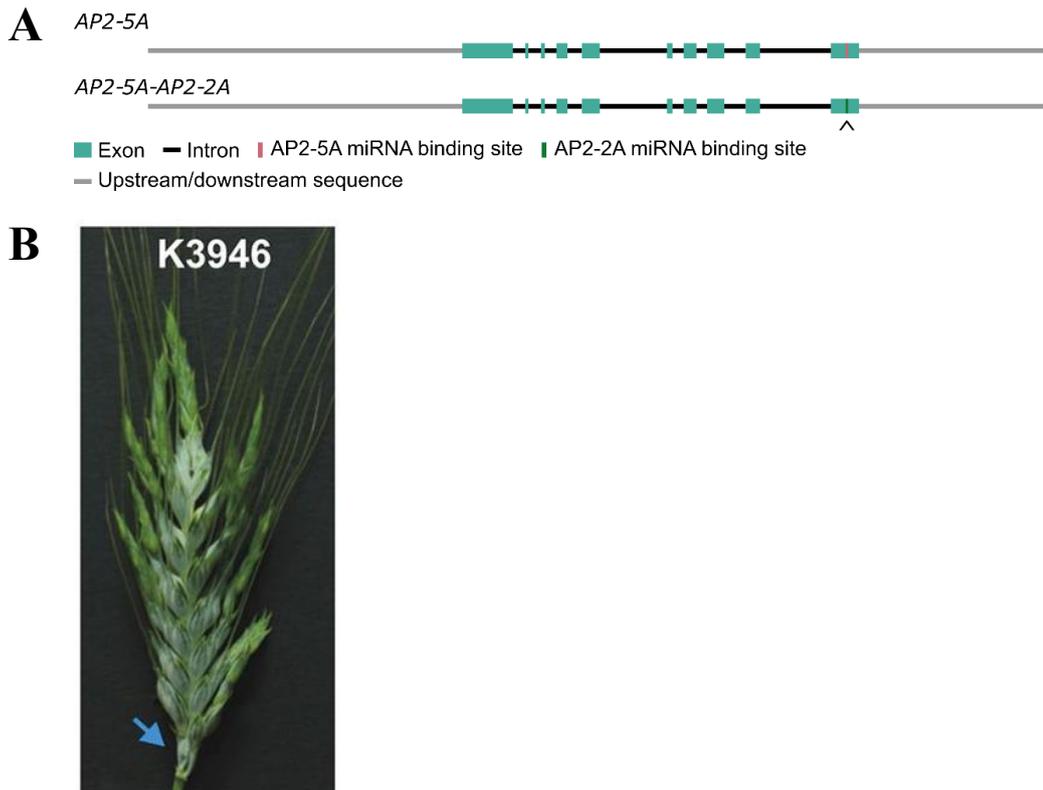


Figure 3-21: (A) Simplified diagram of the microRNA172 (miR172) target site swap constructs used to transform *ap2-5A* *Triticum turgidum* cv ‘Kronos’ (containing a knockout mutation in *AP2-5A*). Exons are shown using blue boxes, introns are depicted by black lines, and upstream/downstream sequences are shown by grey lines. The miR172 binding sites are shown using pink or green lines, highlighted by an arrow. The miR172 binding site from *APETALA2-2A* (*AP2-2A*) is shown in green, while the miR172 binding site from *APETALA2-5A* (*AP2-5A*) is shown in pink. (B) A K3946 wheat (*Triticum turgidum* cv ‘Kronos’) TILLING line primary spike. The spike was harvested three weeks after heading. The blue arrow highlights a rudimentary basal spikelet. Adapted from Debernardi, et al.<sup>32</sup> under a CC BY 3.0 license (<https://creativecommons.org/licenses/by/3.0/>).

If the *AP2-5A-AP2-2A* construct completely complements the K3946 loss-of-function phenotype, we will be able to conclude that, as suggested by the dual luciferase assay, the *AP2-2* and *AP2-5* target sites are bound equally by miR172 in the context of the *AP2-5A* mRNA sequence. If the construct only partially complements the phenotype, this suggests that miR172 does not target the *AP2-2A* target site as efficiently as the *AP2-5A* target site in the context of the *AP2-5A* mRNA sequence.

Due to wheat’s long generation time, I began developing these constructs before I had carried out the dual luciferase assay described in Section 3.4.7.2. Given these results, I would develop a different set of transgenics to test the effect of *AP2-2* and *AP2-5* mRNA structure on miR172

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

binding. It is possible that *AP2-2* is not targeted by miR172 because its target site is not accessible. Therefore, I would carry out the complementary experiment by transforming *ap2-2* loss-of-function mutants with a wildtype *AP2-2A* construct and an *AP2-2A* construct containing the *AP2-5A* binding site. This would show whether the mRNA context of the *AP2-2* binding site affects miR172 binding to the *AP2-5A* target site.

#### 3.4.7.4 Hypothesis 2 conclusions

The results of the *in silico* TargetFinder prediction pipeline suggest that miR172 may not target *AP2-2*, although this is contradictory to previous studies which have shown that SNPs in the *AP2-2* miR172 binding site have phenotypic effects <sup>42</sup>.

The dual luciferase assay showed *in vivo* that miR172a is capable of repressing both *AP2-2* and *AP2-5* target sites and that the level of repression is not significantly different between the two genes. Therefore, the SNP at the 3' end of the binding site does not have a significant effect on repression by either miR172a or miR172b, providing evidence against Hypothesis 2.

As the dual luciferase assay is used to test miRNA binding sites out of their mRNA context, I cannot rule out that differential mRNA folding of *AP2-2* and *AP2-5* affects miR172 repression. The literature shows that miRNA binding and the mechanism of miRNA-mediated repression is affected by the secondary structure of the target mRNA <sup>87,244</sup>. The transgenic lines described in Section 3.4.7.3 will allow us to begin to understand the effects of *AP2-2* and *AP2-5* mRNA secondary structures on miR172 binding.

## 3.5 Discussion

In this chapter my aim was to gain a more nuanced understanding of the key interaction between miR172 and *AP2*-like genes during wheat spike development. Initial results showed that *AP2-2* and *AP2-5* have opposite expression profiles during wheat spike development, despite both being regulated by the same miRNA, miR172 <sup>32,42</sup>. I narrowed my focus to attempt to explain this observation.

### 3.5.1 The opposite expression patterns of *AP2-2* and *AP2-5* reflects the distinct functions of their orthologues in *A. thaliana*

Although both genes are AP2/ERF transcription factors regulated by miR172, *AP2-2* and *AP2-5* have different functions. Disrupting the function of these genes affects some shared phenotypes yet at the same time there are distinct effects <sup>42</sup>. It is therefore unsurprising that *AP2-2* and *AP2-5* lie in different phylogenetic clades and have different orthologues in

**Chapter 3: How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by *microRNA172*?**

Arabidopsis and rice. *AP2-2* is orthologous to *AP2* in Arabidopsis, an A-class floral development gene<sup>42</sup>. *AP2-5* is orthologous to *AtRAP27*, a floral repressor<sup>42</sup>.

Contradictory to previous studies, our data show that *AP2-5* expression decreases during spike development while *AP2-2* expression increases<sup>42</sup>. Debernardi, *et al.*<sup>32</sup> showed that *AP2-5* expression decreases from the floral primordia stage (W3.5<sup>41</sup>) onwards. *AP2-2* was thought to have a similar expression pattern<sup>42</sup>. It is logically consistent that during spike development floral repressor signals would decrease, while floral signals would increase. We cannot be certain, however, that the broad functions of *AP2*-like genes have been conserved between Arabidopsis and wheat. I will explore the roles that *AP2-2* and *AP2-5* play during development further in Chapter 4.

### 3.5.2 Advances in technology and resources have allowed us to deepen our understanding of miR172 in wheat

We know from Arabidopsis that members of the same miRNA family can have different expression patterns and functions<sup>163,164</sup>. Previous work on miR172 has been based solely on miR172a<sup>32</sup>, however recent advances in genetic resources in wheat have allowed us to develop a more nuanced understanding of this miRNA.

Prior to this work, six *MIR172* loci and one mature miR172 sequence had been identified in *T. aestivum* in the IWGSC v1.0 annotation<sup>15</sup>, while no miR172 entries existed in MIRBASE<sup>94</sup>. A meta-analysis of the literature and publicly available sRNA-Seq datasets has revealed an additional four *MIR172* loci and three distinct miR172 mature sequences.

The sRNA-Seq time course described in Chapter 2 revealed a fourth mature sequence, miR172d (which is identical to the single mature sequence in the IWGSC v1.0 annotation<sup>15</sup>). It's interesting that miR172d has a similar expression pattern to miR172b, not miR172a. This is because *MIR172A-4*, *MIR172A-5* and *MIR172A-6* loci may produce both miR172d and miR172a. This is likely as *MIR172A-4* has been used as the precursor sequence in dual luciferase assays in Debernardi, *et al.*<sup>32</sup> and Section 3.4.7.2 of this thesis. In Section 3.4.7.2 I showed that constitutive expression of *MIR172A-4* significantly represses a target site with perfect complementarity to miR172a, but not a scrambled version of this target site. miR172d is different to miR172a in 18 of 21 positions and bears no sequence complementarity to the miR172a perfect target site. One would expect miR172a and miR172d to have similar expression patterns as they would be produced from the same duplex. It may be possible that there may be a switch in which strand is preferred, from miR172d to miR172a, after the W4.0 (terminal spikelet) stage, a phenomenon which has been observed in rice<sup>79</sup>. It would be interesting to test using a dual luciferase assay whether *MIR172A-5* and *MIR172A-6*

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

significantly repress perfect miR172a target sites, and whether *MIR172A-4/5/6* are able to repress perfect miR172d target sites. This would allow us to establish whether single loci are producing multiple miR172 mature sequences at the same loci, or if there has been some level of sub-functionalisation between the *MIR172A* homoeologues on group 6 chromosomes.

The *MIR172* loci I identified during my analysis form three syntenic homoeologous groups, showing that these are well-conserved loci which were present in the A-, B- and D-genome progenitors of hexaploid bread wheat. This high level of conservation across evolutionary time emphasises their importance to proper wheat spike development.

This comprehensive review provides a much more complete view of miR172 in wheat. I have established a nomenclature for miR172 in wheat which will provide clarity for future work on this key miRNA. This work laid the foundation which allowed me to investigate subtly different miR172 family members independently, which has been shown to be important in *Arabidopsis*<sup>163,164</sup>. To fully understand miRNA function we must characterise each miRNA family member. The nuance of these distinct interactions is important, however this work is time consuming. Consideration needs to be given for whether protocols, such as RT-qPCR, can distinguish between miRNA family members. Sometimes more complex, expensive methods, such as sRNA-Seq, must be used to disentangle closely related miRNAs.

### 3.5.3 Developing hypotheses to explain the opposite expression profiles of *AP2-2* and *AP2-5*

Two of the mature miR172 sequences I identified were highly expressed, which led me to hypothesize that miR172a and miR172b may have different functions (as seen in *Arabidopsis*<sup>163,164</sup>). This functional differentiation could be due to differences in miRNA expression patterns (Hypothesis 1) or sequence-mediated (Hypothesis 2). Their different levels of complementarity to *AP2*-like genes led me to hypothesise that miR172a may preferentially target *AP2-5* while miR172b may preferentially target *AP2-2*, leading to the opposite expression profiles of these closely related genes.

#### 3.5.3.1 *There are few regions where AP2-2 is expressed in the absence of AP2-5*

The MERFISH experiments in Section 3.4.6.1 show that in the vast majority of tissues in the developing wheat spike *AP2-2* and *AP2-5* transcripts overlap. Hypothesis 1 (that the expression domains of *AP2*-like genes and miR172 do not overlap spatially or temporally), relies on *AP2-2* not being expressed in the same cells at the same time as miR172. Across the W5.0 spike *AP2-2* is expressed in the absence of *AP2-5* in just 9.19% of cells. The cell cluster with the highest number of cells with *AP2-2* expression only was cluster 17 (lodicles), where 70.31% of cells where *AP2-2* was expressed, it was expressed on its own. However, the

*Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?*

lodicules are very small organs with this cell cluster containing just 1.85% of cells across the W5.0 spike. Therefore, even if miR172 is not expressed here, allowing *AP2-2* expression to increase, it is unlikely that a phenomenon in such a small region would confer the clear pattern seen in the RNA-Seq data in Figure 3-10 and Figure 3-11.

*3.5.3.2 miR172a and miR172b expression profiles are unique, but both overlap with AP2-like gene expression*

To further test Hypothesis 1, I established whether miR172 is expressed at the same time as *AP2*-like genes. sRNA-Seq showed that miR172 is expressed at the W4.0 and W5.0 stages of wheat spike development which overlaps with the expression of both *AP2-2* and *AP2-5*. Therefore, the increase in *AP2-2* expression is not the result of a decrease in miR172 expression.

By using sRNA-Seq rather than RT-qPCR, I have been able to quantify the abundance of individual miR172 family members. This showed that miR172a and miR172b abundance diverges at W5.0; miR172a abundance is maintained while miR172b abundance decreases. This, along with the differences in complementarity between miR172 siblings and *AP2*-like genes, lends weight to Hypothesis 2, that different mature miR172 family members target different *AP2* genes. miR172a abundance stays high, repressing *AP2-5*, while miR172b abundance decreases, allowing *AP2-2* expression to increase.

*3.5.3.3 miR172a and miR172b repress AP2-2 and AP2-5 target sites equally*

To test Hypothesis 2 empirically, I used a dual luciferase assay to quantify repression of *AP2*-like gene target sites by miR172 family members. This assay showed that miR172 siblings target *AP2-2* and *AP2-5* target sites equally, therefore the SNP at the 3' end of the target site does not make a significant difference to the level of binding and repression, providing evidence against Hypothesis 2. A previous study has used a dual luciferase assay to quantify differences in miRNA-mediated repression between a perfectly complementary target site and a site with a single mismatch at the 5' end; they found that there was a significantly reduced level of repression when this mismatch was present<sup>207</sup>. Therefore, this system should be sensitive enough to detect changes in miRNA-mediated repression resulting from single nucleotide changes.

Interestingly, miR172b did not significantly repress any of the *AP2*-like gene target sites. However, the variation between the ratios of biological replicates was higher when testing miR172b binding (average CV (coefficient of variation) across all treatments = 53.65%) than miR172a binding (average CV = 44.62%). In Figure 3-20, the ratios of *AP2*-like gene target sites appear lower than those for scrambled target site controls, although the difference is not

**Chapter 3: How do *APETALA2-2* and *APETALA2-5* have different expression patterns in developing spikes when they are both regulated by *miR172*?**

statistically significant. I cannot rule out that miR172b does repress *AP2*-like genes but the high level of variability in our assay meant we could not detect it.

A limitation of this assay is that the miRNA binding site is out of context; the binding sites in developing wheat spikes exist in the context of the *AP2* mRNA. It has been shown that mRNA structure and its effect on target site accessibility can affect miRNA binding<sup>244</sup>. Additionally, the structure of the two nucleotides immediately downstream of the miRNA binding site have been shown to specifically affect mRNA cleavage but not miRNA binding, so perhaps affecting the dominant miRNA-mediated repression mechanism<sup>87</sup>. This raises the possibility that *AP2-5* is regulated by miR172 via mRNA cleavage, while *AP2-2* is regulated via translational repression. This would be consistent with the fact that we observe a decrease in *AP2-5* mRNA while miR172 abundance increases, but an increase in *AP2-2* mRNA. I will discuss this possibility further in Chapter 5.

The transgenic lines that are currently being generated will enable us to begin to test whether the mRNA context of the *AP2*-like miR172 binding sites affects miR172-mediated repression of these genes. We now have the resources in wheat to be able to test hypotheses such as these directly *in planta* by complementing a mutant line. This will allow us to test a wide range of hypotheses in more direct ways in the future, significantly improving our understanding of wheat development.

### 3.6 Conclusions and future perspectives

I generated three hypotheses to explain the opposite expression profiles of *AP2-5* and *AP2-2*. sRNA-Seq revealed that miR172 expression overlaps with *AP2-5* and *AP2-2* expression temporally. Spatial transcriptomics show that *AP2-2* and *AP2-5* transcripts broadly overlap and there are only small regions where *AP2-2* is independently expressed. In the future, it would be interesting to reveal the spatial expression pattern of miR172a and miR172b in wheat. LNA (locked nucleic acid) *in situ* hybridisations can be used to spatially resolve miRNA expression and is capable of distinguishing between mature sequences with single SNP differences<sup>183</sup>. This would show whether the miR172 siblings have the same pattern of expression and where in the developing wheat spike miR172 overlaps with *AP2*-like gene transcripts.

I have also shown that miR172a and miR172b repress *AP2-2* and *AP2-5* target sites equally *in vivo*. This suggests that there is no sequence-based functional differentiation between miR172a and miR172b, disproving Hypothesis 2. However, the target sites used in this assay are not in their endogenous mRNA context. Therefore, there may be some additional factors which affect miR172 repression of *AP2*-like genes, such as mRNA structure and miRNA target

**Chapter 3: How do APETALA2-2 and APETALA2-5 have different expression patterns in developing spikes when they are both regulated by microRNA172?**

site accessibility. Additional work to uncover the mRNA structures of *AP2-2* and *AP2-5* and how this might affect the accessibility of the miR172 binding sites would be an interesting future avenue of research.

The work in this chapter has shown that miR172a and miR172b are likely to be functionally redundant, and that the opposite expression profiles of *AP2-2* and *AP2-5* may be the result of different underlying levels of transcription (Hypothesis 3). This could be quantified using smFISH to detect immature (unspliced) and mature (spliced) mRNA transcripts of the *AP2*-like genes <sup>245</sup>. Despite this, I have shown the importance of a more nuanced understanding of miRNA-mRNA interactions, as demonstrated by the divergent expression patterns of miR172a and miR172b. Recent developments in genetic resources in wheat allow us to study these interactions in more detail and reveal the nuance of these critical genetic pathways.

## 4 Does *VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2)* act via *microRNA172* to achieve its phenotypic effect?

### 4.1 Chapter summary

In this chapter, I investigated the downstream genetic pathway of *VRT-A2*, a key transcription factor during the vegetative to floral transition. Modifying the expression of *VRT-A2* or *AP2*-like genes affects several of the same phenotypes, so I focused my work on the interactions between these genes. I found that lower *VRT-A2* expression leads to an increase in *miR172* abundance, a genetic effect which is conserved with *Arabidopsis*. In NILs with increased and extended *VRT-A2* expression, *AP2-2* expression was significantly lower while *AP2-5* expression was significantly higher. The increase in *AP2-5* abundance can be explained by a decrease in *miR172*, however the decrease in *AP2-2* cannot be explained using current data.

### 4.2 Introduction

#### 4.2.1 A natural allele of *VRT-A2* leads to increased and extended expression

##### 4.2.1.1 *Triticum polonicum*

*Triticum polonicum*, also known as Polish wheat, first appears in the record in 1687<sup>246</sup> and was characterised by Carl Linnaeus in 1753<sup>107</sup>. Linnaeus noted the longer spikes and glumes double the length of other species<sup>107</sup>; these are the most prominent phenotypes of *T. polonicum* and the morphological characters which led to its species classification. It is a tetraploid wheat which is sometimes described as a subspecies of durum wheat, *T. turgidum*<sup>56</sup>. *T. polonicum* is not widely grown today due to its susceptibility to lodging<sup>247</sup>, however it has been used to show that the patterns of inheritance for quantitative traits follow similar principles as those shown by Mendel for qualitative traits in peas<sup>248,249</sup>. Rowland Biffen used crosses between *T. polonicum* and other wheat species to show that certain features of wheat, such as the presence of awns and susceptibility to the yellow rust pathogen, are inherited as single dominant/recessive factors. He also illustrated incomplete dominance for other traits, such as glume size and flowering time, where the hybrid individuals display an intermediate phenotype. This phenomenon had not been observed by Mendel, although his principles are

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

capable of explaining it. It was during these studies that Biffen also recorded the long grain, early flowering, and lax spike phenotypes present in *T. polonicum*<sup>248</sup>.

**4.2.1.2 VRT-A2 was identified as the causal gene for the long glume and long grain phenotype in *Triticum polonicum***

Although the inheritance of *T. polonicum*'s distinctive phenotypes was studied extensively in the early 20<sup>th</sup> century, the causal locus was unknown until recently. Through further crossing schemes, Engledow showed in 1920 that several *T. polonicum* phenotypes are inherited together, including glume and grain length, and that they are all controlled by a single "factor"<sup>250</sup>, what we now know to be a genomic locus. This locus (often referred to as the 'P' or 'PI' locus) was mapped to chromosome 7A in 1996<sup>251</sup>. In 2021, Adamski, *et al.*<sup>56</sup> used NILs to fine-map *PI* to a 50.3 kbp region containing the gene *VRT-A2*<sup>56</sup>. This 50.3 kbp region is completely linked to the glume length, plant height, spike length, grain length, and TGW phenotypes seen in *T. polonicum*<sup>56</sup>, supporting the hypothesis that these are pleiotropic effects of *VRT2*. Liu, *et al.*<sup>108</sup> also independently mapped the *PI* locus to *VRT-A2*.

**4.2.1.3 A sequence substitution in intron 1 of VRT-A2 is responsible for the *T. polonicum* phenotype**

*VRT-A2* was identified by Adamski, *et al.*<sup>56</sup> and Liu, *et al.*<sup>108</sup> as the locus responsible for the *T. polonicum* phenotype<sup>56</sup>. Through sequencing of the candidate genes and the surrounding region, a 160 bp sequence substitution was identified in intron 1 of the *T. polonicum VRT-A2* allele<sup>56</sup> (hereafter referred to as the *VRT-A2b* allele). This 160 bp sequence replaces a 563 bp sequence in the wildtype allele (*VRT-A2a*). From further analysis, this 160 bp substitution was shown to contain motifs which flank the 563 bp wildtype sequence<sup>56</sup>, consistent with a sequence re-arrangement occurring during DNA repair. The intron 1 substitution is linked to the *T. polonicum* phenotype, as 36 separate *T. polonicum* accessions from 17 countries were shown to carry the *VRT-A2b* allele, while 297 other accessions with 'normal' length glumes carry the *VRT-A2a* allele<sup>56</sup>. This suggests that the *VRT-A2a* allele is the ancestral form, while the *VRT-A2b* allele is derived<sup>56</sup>.

**4.2.1.4 The sequence substitution in intron 1 of VRT-A2b causes increased and ectopic expression in vegetative and reproductive tissues**

The two *VRT-A2* alleles exhibit different expression patterns; *VRT-A2b* confers ectopic *VRT-A2* expression<sup>56</sup>. Adamski, *et al.*<sup>56</sup> quantified the expression of all three *VRT2* homoeologues (*VRT-A2*, *VRT-B2*, and *VRT-D2*) using RT-qPCR in several tissues and at multiple timepoints<sup>56</sup>. In all tissues, at all timepoints measured, *VRT-B2* and *VRT-A2* expression not significantly different in plants with the *VRT-A2a* or *VRT-A2b* allele<sup>56</sup>.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (*VRT-A2*) act via *microRNA172* to achieve its phenotypic effect?**

*VRT-A2a* is expressed within the shoot apical meristem, especially within the vegetative meristem. Its expression level begins to decrease as the meristem transitions into the reproductive stage and continues to develop and differentiate <sup>56</sup>. While the expression of *VRT-A2b* follows a similar pattern, its level of expression is higher at all stages measured and begins to increase at W4.0 (terminal spikelet) stage <sup>41,56</sup>. Due to the relatively low level of resolution of this expression analysis, it is possible that in tissue types where an increased level of expression is detected, *VRT-A2* is simply being expressed ectopically in a subsection of the tissue sample. On balance, however, the authors concluded that this increased signal is likely due to an increase in expression levels <sup>56</sup>.

While *VRT-A2a* is not expressed in glumes or floral organs (lemma or anthers), *VRT-A2b* is ectopically expressed in these tissues <sup>56</sup>. Expression is particularly high in glumes at the mid-boot stage <sup>56</sup>.

#### 4.2.2 *VRT-A2* is an ortholog of *SVP* in *Arabidopsis*

*VRT2* was originally characterised by Kane *et al.* in 2005 as part of a study of flowering time regulation in wheat <sup>252</sup>. It is an ortholog of *AtSVP* in *Arabidopsis thaliana*, sharing 51% amino acid sequence identity <sup>252</sup>. Like *AtSVP*, *VRT2* has been shown to repress the transition from vegetative to reproductive growth <sup>44,252</sup>. There have been two duplication events within the *SVP* clade in monocots, both of which occurred over 45 million years ago <sup>253,254</sup>. The three clades in wheat are *VRT2*, *SVPI*, and *SVP3*, each of which contains three homoeologues (Figure 4-1).

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

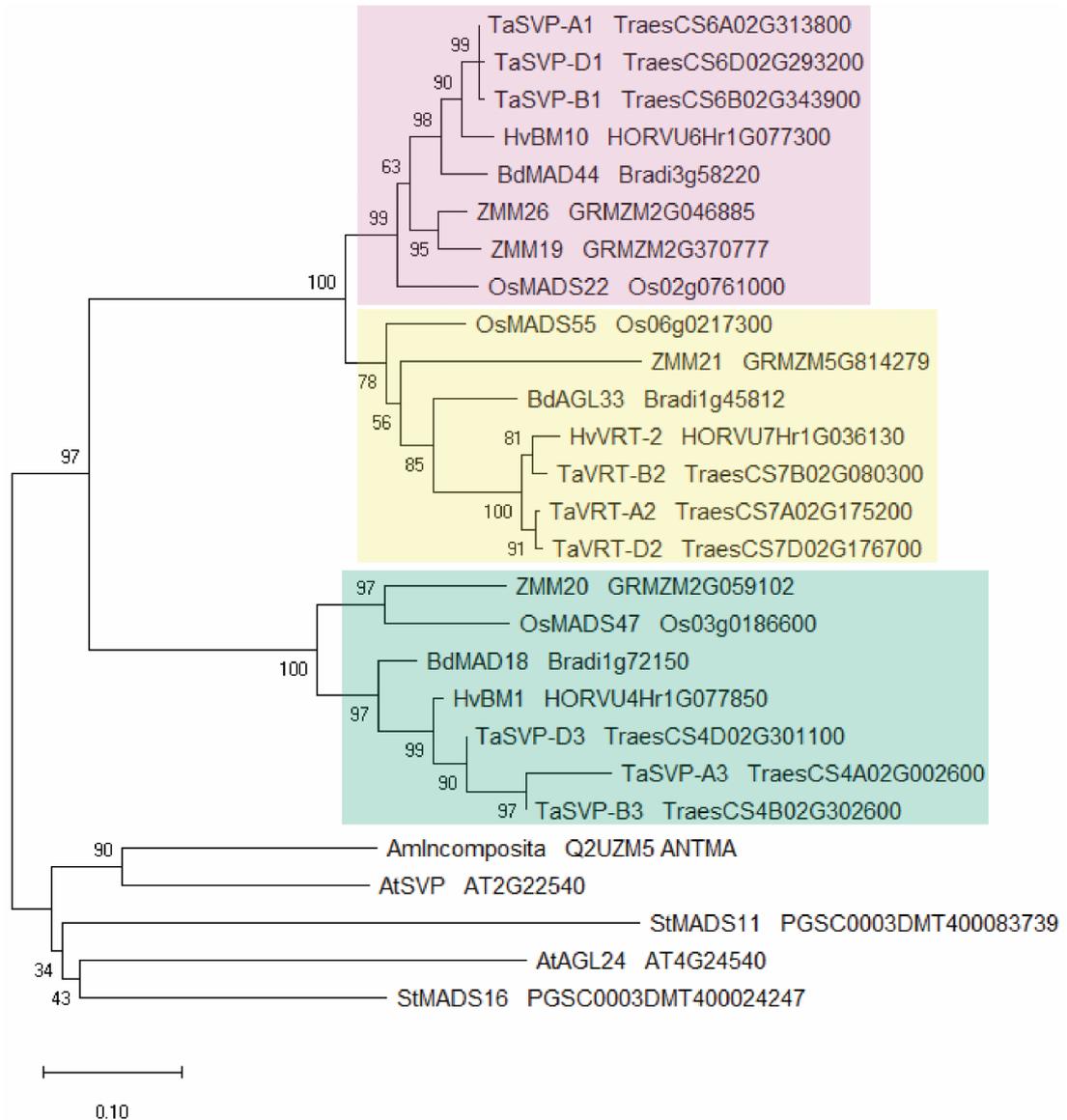


Figure 4-1: Phylogenetic tree of the three SHORT VEGETATIVE PHASE-like (SVP-like) gene clades in wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), Brachypodium distachyon, maize (*Zea mays*), and rice (*Oryza sativa* spp. japonica). The SVP1 clade is highlighted in pink, the VEGETATIVE TO REPRODUCTIVE TRANSITION 2 (VRT2) clade is highlighted in yellow and the SVP3 clade is highlighted in green. Numbers on branch points indicate confidence values based on bootstrap analysis (1000 replicates). The Poisson method was used to calculate branch lengths<sup>255</sup>. Reproduced from Adamski, et al.<sup>56</sup>.

**4.2.2.1 The three SVP genes in wheat have different expression patterns**

The three SVP gene clades in wheat all exhibit different expression patterns. SVP1 is expressed in both vegetative and floral tissues<sup>56</sup>, while SVP3 and VRT2 are only endogenously expressed in vegetative tissues; they are not expressed in floral organs<sup>22,56</sup>. This may be an example of sub-functionalisation, where duplicate gene copies specialise in a particular function<sup>256</sup>.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

Although, this process would not be complete in this case as *SVPI* retains both vegetative and floral tissue expression.

Double *svp1 vrt2* mutants have stronger phenotypes than either single mutant for flowering time, plant height, number of spikelets, and leaf number, suggesting there is some functional redundancy between the two, although their effects are dosage dependent<sup>44,56</sup>. This mutant is a double knockout of the two most closely related genes in the *SVP* clade; a triple *svp1 svp3 vrt2* mutant has not yet been described in the literature.

**4.2.2.2 The AtSVP regulatory network has been well characterised**

Although it is known that the differential expression of *VRT-A2* causes the *T. polonicum* long glume phenotype, the mechanism by which this occurs is unknown given the very different floral morphology of Arabidopsis. Regardless, some hypotheses can be formed on the role of *SVP* genes in floral transition based on knowledge in Arabidopsis.

ChIP-seq analyses have shown that AtSVP proteins bind to many hundreds, if not thousands of targets *in vivo*<sup>3,257</sup>. *AtSVP* is endogenously expressed at two distinct stages of development; during vegetative growth and floral development<sup>258</sup>. *AtSVP* is downregulated during the floral transition itself<sup>257,258</sup>. It has been postulated that *AtSVP* has distinct functions during these two stages, and that its effect is determined by its interactions with different protein partners<sup>258</sup>.

During vegetative growth *AtSVP* acts as a floral repressor, integrating signals from multiple pathways to determine the timing of the floral transition<sup>258,259</sup>. It regulates flowering time by binding directly to other floral integrators such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*AtSOC1*)<sup>259,260</sup>, *AtFT*, and *TWIN SISTER OF FT* (*AtTSF*)<sup>260</sup>. It has also been shown to bind members of the *CLAVATA-WUSCHEL* (*AtCLV-AtWUS*) pathway, which maintains shoot apical meristem size<sup>257,261</sup>.

When *AtSVP* was first characterised, knockout mutants were shown to have wildtype floral architecture; it was therefore posited that either *AtSVP* function is dependent on an interacting factor that is not present during floral development, or that the function of *AtSVP* during floral development is redundant<sup>258</sup>. The latter has been shown to be the case. *AtSVP* and *AGAMOUS-LIKE 24* (*AtAGL24*) have opposite functions during vegetative growth, with *AtAGL24* promoting the floral transition while *AtSVP* represses it<sup>262</sup>. However, after the floral transition these two genes have redundant functions, both having the ability to form a dimer with *APETALA1* (*AtAPI*)<sup>262</sup>, an A-class gene in the ABCDE floral development model<sup>262,263</sup>.

There is strong evidence that miR172a is directly negatively regulated by *AtSVP* in Arabidopsis<sup>3</sup>. As miR172 represses *AtAP2* and *AtRAP27*, this interaction would result in higher *AtAP2* and *AtRAP27* protein levels. There are studies showing that *AtAP2* and *AtRAP27*

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

are also directly bound and upregulated by overexpressed AtSVP protein in nine day old plants in a ChIP-chip experiment <sup>3</sup> and are downregulated in an *svp* loss-of-function mutant in ten day old plants <sup>264</sup>. This was contradictory to other studies where *AtAP2* and *AtRAP27* were not bound by an AtSVP::GFP (green fluorescent protein) protein <sup>257</sup> and were not differentially expressed in an *Atsvp* loss-of-function line <sup>265</sup>, however the samples for these studies were taken at later timepoints (two weeks after bolting and during the transition to flowering, respectively). This suggests that the nature of any interaction may change over time.

### 4.2.3 The regulatory network of VRT-A2 needs to be investigated

The above examples are a selection of key pathways that *AtSVP* has been shown to interact with; others have been identified <sup>3,257</sup>. This diversity of targets show how important *AtSVP* is during various developmental stages, in addition to being an integrator of upstream signals. As an ortholog of *AtSVP*, *VRT-A2* may also have multiple functions important to development. Therefore, elucidating its function may be an essential step in understanding wheat development better at a fundamental level and eventually improving yields. The results of recent studies can provide some leads for this analysis.

#### 4.2.3.1 VRT2 has been shown to interact with SQUAMOSA proteins

The downstream targets of *VRT2* have been studied previously, often in the context of flowering time regulation. For instance, it was shown in 2007 that *VRT2* binds to and represses *VERNALIZATION1 (VRN1)*, a member of the SQUAMOSA gene clade <sup>44,266</sup>. Expression data from Ubi::VRT2 overexpression lines has been analysed for differentially expressed genes that may be regulated by *VRT2* <sup>44</sup>. Several members of the SQUAMOSA and SEPALLATA clades were differentially expressed. The SQUAMOSA clade includes *VRN1*, *FRUITFULL 2 (FUL2)*, and *FRUITFULL 3 (FUL3)* <sup>44</sup>. The E-class SEPALLATA clade gene is divided into the LOFSEP (*SEPI-2*, *SEPI-4*, and *SEPI-6*) and SEP3 clades (*SEP3-1* and *SEP3-2*). SQUAMOSA clade gene *VRN1*, B-class gene *PISTILLATA1 (PII)*, C-class gene *AG1*, two LOFSEP clade genes (*SEPI-2* and *SEPI-4*), and both SEP3 clade genes showed significantly lower levels of expression in Ubi::VRT2 W4.0 (terminal spikelet) spikes, suggesting these genes are downregulated (either directly or indirectly) by *VRT2* <sup>44</sup>. Additionally, a yeast-2-hybrid study found that *VRT2* has the ability to bind to all three SQUAMOSA clade genes <sup>44</sup>. These interactions were additionally verified using bimolecular fluorescence complementation (BiFC) in wheat protoplasts <sup>44</sup>. Li, *et al.* <sup>44</sup> have proposed a working model accounting for the interactions described above. During vegetative growth dependent on its expression pattern, *VRT2* can act as a floral repressor or promoter as both the knockout and overexpression mutants show a delayed heading time phenotype <sup>44</sup>. This flexibility has also been shown in a 2019 study when Ubi::VRT2 caused early flowering in unvernallized winter

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

wheat, but vernalized plants had wildtype flowering times <sup>267</sup>. After this stage, SVP and SQUAMOSA proteins interact to accelerate the transition of the inflorescence meristem into a terminal spikelet meristem <sup>44</sup>. The downregulation of *SVP* genes during the floral transition is hypothesised to release SQUAMOSA proteins to interact with SEPALLATA proteins, necessary for normal floral development <sup>44</sup>.

**4.2.3.2 An interaction orthologous to those between SVP, AP2 and miR172 in Arabidopsis has not been tested in monocots**

With the exception of the network described in Section 4.2.3.1, the downstream targets of *SVP* orthologues have not been investigated in monocots. In this chapter I will focus on testing whether the interactions between *AtSVP* and *AtAP2*, *AtRAP27* and miR172 in Arabidopsis has been conserved in wheat. This will build upon the work I described in Chapter 3 where I characterised the repression of *AP2-2* (the wheat orthologue of *AtAP2*) and *AP2-5* (the wheat orthologue of *AtRAP27*) by miR172. Understanding how *VRT2* and *AP2-5*, in particular, interact is critical to understanding wheat spike development as they are both such key genes during the vegetative to reproductive transition.

**4.2.4 AP2-5A and miR172 mutants have similar phenotypes to VRT-A2 mutants**

A striking piece of evidence from the literature is the similarity between *VRT-A2* mutants, and mutants of *AP2-5A* and miR172 which may imply that these genes function within the same genetic pathway in wheat <sup>32</sup>.

There are several phenotypes affected by *VRT-A2*, *AP2-5A*, and miR172 expression including plant height, heading date, and spikelet density <sup>32,56</sup>. Plants with the *VRT-A2b* allele are taller, which is the same as the phenotype seen in Ubi::miR172 plants <sup>32,56</sup>. Spikelet density in *VRT-A2b* plants is also consistent with Ubi::miR172 plants <sup>32,56</sup>. In contrast, the heading date in *VRT-A2b* plants is later, which is also seen in MIM172 plants (with lower miR172 levels) <sup>32,56</sup>. This conflicting evidence suggests that if there is any interaction between the three genes, the nature of this interaction may differ according to tissue and developmental stage.

**4.2.5 AP2-2 orthologue loss-of-function mutants in barley have long glumes and grains**

The barley orthologue of *AP2-2*, *HvAP2*, has been well-characterised using loss-of-function and gain-of-function mutants <sup>57</sup>. Shoosmith, *et al.* <sup>57</sup> showed that the *gigas1.a HvAP2* loss-of-function mutant line has longer glumes, lemma, palea and grains than the wildtype. *Hvap2* spikes were also more lax (lower spikelet density) <sup>57</sup>, an observation which is consistent with

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

the higher spikelet density seen in both wheat and barley spikes with a miR172-resistant allele of *AP2-2* (so higher *AP2-2* expression) <sup>42,240</sup>. With the exception of the long palea phenotype, the other observations are reminiscent of wheat spikes with the *VRT-A2b* allele <sup>56</sup>. I therefore hypothesised that the *VRT-A2b* phenotype may be mediated by a reduction in *AP2-2* abundance (Figure 4-2).

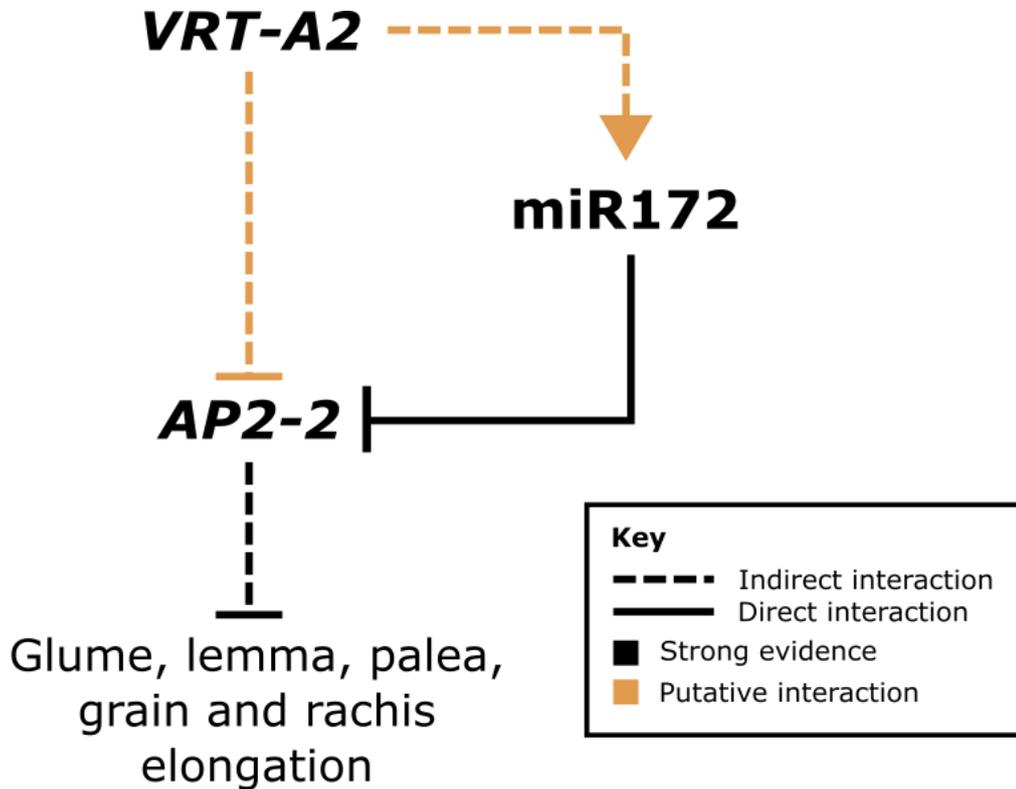


Figure 4-2: A working model describing hypothesised interactions between *APETALA2-2* (*AP2-2*), *VEGETATIVE TO REPRODUCTIVE TRANSITION-A2* (*VRT-A2*), *microRNA172* (*miR172*) and spike phenotypes. The model is based on previous literature describing how increased and extended *VRT-A2* expression and decreased *AP2-2* expression both lead to long floral organs and a lax spike phenotype <sup>56,57</sup>. Interactions with strong evidence are shown using black lines, while interactions which I have hypothesised are shown using orange lines. Interactions which are known to be direct are shown using a solid line, while those where the interaction may be indirect are shown using a dashed line.

#### 4.2.6 Aims and hypotheses

*VRT-A2* has been shown to have key pleiotropic effects during wheat spike development <sup>43,56</sup>. I aimed to understand the mechanism by which *VRT-A2* confers its effects. As a MADS box transcription factor, I aimed to understand its targets and downstream genetic pathway.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

As I have described in Section 4.2.4, modifications to the expression of *AP2-5A* and *VRT-A2* affect several of the same phenotypes such as plant height, heading date, and spikelet density<sup>32,44,56</sup>. Mutations to *AP2-2* and *VRT-A2* also affect similar phenotypes, in this case floral organ length and spikelet density<sup>42,56,57</sup>. Therefore, I hypothesised that these three genes may act within the same genetic pathway. I aimed to test this hypothesis in this chapter. Initial results led me to refine my aim to test whether *VRT-A2* interacts with *AP2-5* and *AP2-2* directly, or via *miR172*.

## 4.3 Methods

### 4.3.1 Paragon RNA-Seq

Details of the Paragon RNA-Seq dataset can be found in 3.3.2.

### 4.3.2 *ap2-5A VRT-A2b* phenotyping

#### 4.3.2.1 Germplasm

I was provided with F<sub>3</sub> seed of crosses between three *ap2-5A* mutants (K2726, K2992, and K3946) and a NIL (*Langdon PI<sup>POL</sup>*) by Dr Nikolai Adamski (Table 4-1).

All *ap2-5A* mutant lines are TILLING lines in a tetraploid *T. turgidum* cv ‘Kronos’ background and they all carry characteristic transitions from guanine (G) to adenine (A) residues within the *AP2-5A* gene. In line K2726, the transition occurs at the splice donor site of the first intron. In line K2992, the transition is located in the first base of the second exon, leading to a premature stop codon at amino acid residue 138 (W138\*). Line K3946 contains a G to A transition in the fifth exon, conferring a premature stop codon at amino acid residue 190 (W190\*). These truncated AP2-5A proteins are predicted to lack both (K2726, K2992) or one (K3946) AP2 domain and can thus be assumed to be non-functional. *Langdon PI<sup>POL</sup>* is a BC<sub>3</sub>F<sub>2</sub> NIL carrying the *VRT-A2b* allele (which confers increased and extended *VRT-A2* expression) in a *T. turgidum* cv ‘Langdon’ background.

#### 4.3.2.2 Plant growth

Seeds were germinated in petri dishes as described previously<sup>268</sup>. Once germinated, they were sown in 96-cell trays containing ‘peat & sand mix’ (85% fine peat, 15% grit, 2.7 kg/m<sup>3</sup> osmocote 3-4 months, wetting agent, 4 kg/m<sup>3</sup> maglime, 1 kg PG mix). After approximately 4 weeks, seedlings were potted up into 1 L pots containing ‘John Innes cereal mix’ (40% medium grade peat, 40% sterilised soil, 20% horticultural grit, 1.3 kg/m<sup>3</sup> PG mix 14-16-18 + Te base fertiliser, 1 kg/m<sup>3</sup> osmocote mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3 kg/m<sup>3</sup> maglime, 300 g/m<sup>3</sup> exemptor). Plants were grown in heated glasshouses with supplementary

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

lighting according to a 16 h 20°C light / 8 h 15°C dark cycle. Primary spikes were harvested after physiological maturity and dried down at 30°C for 3 d.

*Table 4-1: Details of germplasm used in this experiment. The mutation location refers to the location of interest and is relative to the IWGSC v1.0 assembly<sup>15</sup>. The gene IDs provided are according to the IWGSC v1.1 annotation<sup>98</sup>.*

Accession	Mutation location (IWGSC v1.0 assembly)	Mutation of interest IWGSC v1.1 gene model	Genetic background	Reference
K2992	chr5A:650130297	TraesCS5A02G473800	Triticum turgidum cv 'Kronos'	Debernardi, <i>et al.</i> <sup>32</sup>
K3946	chr5A:650129824	TraesCS5A02G473800	Triticum turgidum cv 'Kronos'	Debernardi, <i>et al.</i> <sup>32</sup>
K2726	chr5A:650130395	TraesCS5A02G473800	Triticum turgidum cv 'Kronos'	Debernardi, <i>et al.</i> <sup>32</sup>
<i>Langdon</i> <i>PJ<sup>POL</sup></i>	N/A	TraesCS7A02G175200	Triticum turgidum cv Langdon	Adamski, <i>et al.</i> <sup>56</sup>

#### 4.3.2.3 DNA extraction

DNA was extracted based on the protocol published by Pallotta, *et al.*<sup>269</sup>. A 3 mm tungsten bead was placed in each well of a 1.2 mL collection plate along with approximately 2.5 cm leaf tissue from seedlings less than 4 weeks old. Plates were placed in a freeze-dryer overnight at -20 °C. The samples were homogenised using a Geno/Grinder (SPEX SamplePrep) at 1500 RPM for 90 s and briefly centrifuged. If required, the homogenisation step was repeated, and the plate was briefly centrifuged again. 500 µL of extraction buffer (0.1 M Tris-HCl (pH 7.5), 0.05 M EDTA (pH 8.0), 1.25% SDS) at 65 °C was added to each well. The buffer was mixed with the ground tissue by inverting the plates several times. The plates were incubated at 65 °C for 30-60 min then cooled for 15 min at 4 °C. Next, 250 µL of 6 M ammonium acetate at 4 °C was added to each well and mixed by plate inversion. Plates were incubated at 4 °C for 15 min and centrifuged at 5,000 RPM for 15 min. The supernatant was transferred to another 1.2 mL collection plate containing 360 µL isopropanol per well and mixed by plate inversion. The plates were incubated at 4 °C for 5-90 min and centrifuged at 5,000 RPM for 15 min. The supernatant was discarded, then 500 µL 70% ethanol was added per well and mixed by plate inversion. The plates were centrifuged at 5,000 RPM for 15 min. The supernatant was discarded, and the DNA pellets were dried either by incubating the plates

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (*VRT-A2*) act via *microRNA172* to achieve its phenotypic effect?**

at 65 °C for 30 min, or at room temperature overnight. 200 µL sterile distilled water was added per well and the plates were vortexed. The samples were incubated at 65 °C for 15 min, and vortexed again. Five random samples per plate were quantified using the DeNovix DS-11 spectrophotometer to assess the approximate quality and quantity of DNA extracted.

#### 4.3.2.4 *PACE genotyping*

The 3CR Bioscience PACE system was used to genotype plants according to manufacturer's instructions<sup>270</sup>. The total reaction volume per well was adjusted to 4 µL instead of the recommended 5 µl, as described previously<sup>56</sup>. An initial 30 cycle reaction was used, and five additional cycles were added if necessary after reading the plates using the BMG LabTech PHERAstar. The primers used are listed in Section A.1, Table A-1.

For clarity, in the remainder of this chapter I will refer to wildtype plants as *VRT-A2a/AP2-5A*, *ap2-5A* loss-of-function mutants as *VRT-A2a/ap2-5A*, plants with the *VRT-A2b* allele as *VRT-A2b/AP2-5A*, and plants with the *ap2-5A* loss-of-function mutation and the *VRT-A2b* allele as *VRT-A2b/ap2-5A*.

#### 4.3.2.5 *Phenotyping*

Heading date was defined as the day on which the peduncle of the main spike could be first observed and the spike had fully emerged from the flag leaf sheath. Plants were checked daily, with heading time measured in days from sowing date to heading date. Plant height was measured from the soil to the tip of the tallest spike (excluding awns).

The main spike of each plant was used for phenotyping. Spike length was measured from the peduncle/rachis junction to the end of the terminal spikelet (excluding awns) once the spikes had been dried at 30°C for 3 d. Spikelet number was counted after drying, including sterile basal spikelets. Sham ramification (an extended rachilla including attached florets) was scored after drying as not present (0), moderate (1), or extreme (2). Extreme sham ramification was defined as the rachilla being over double the approximate average wildtype rachilla length, while moderate sham ramification describes a spikelet with a longer than average rachilla less than double the approximate average length. After these measurements were taken, the spike was dissected, and the glumes were arranged on plate films according to spikelet number (including sterile basal spikelets). Lemma, palea, and grains of fertile florets (those which produced a grain) were arranged on plate films according to spikelet number (including sterile basal spikelets) and floret number (including infertile florets). The plate films were scanned (600 dpi, greyscale) and the organs were measured using FIJI's particle measurement feature (including particles 0.05-10 cm<sup>2</sup>) (Schindelin, *et al.*<sup>271</sup>). When necessary for accurate dissection, a Leica MZ16 microscope with a Leica CLS100x white light source and Leica DFC420 colour camera was used.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

**4.3.2.6 Statistical analysis**

I used R v4.0.3 for all data analyses<sup>272,273</sup>. Two-way ANOVAs were used to assess the independent effect of *AP2-5A* and *VRT-A2* alleles as well as any interaction between the two factors on plant height, heading time, number of spikes, number of spikelets per spike, basal sterility, and spike length. The spikelet architecture data (*i.e.*, the number of glumes, bracts, and florets and percentage fertility) did not satisfy the assumptions of a two-way ANOVA. Therefore, I used a non-parametric Kruskal-Wallis test to assess the effect of the *VRT-A2/AP2-5A* genotype on the phenotype. A pairwise Wilcoxon rank sum test was used for pairwise comparisons between genotypes to produce a Benjamini & Hochberg adjusted *p*-value. The effect of *VRT-A2* and *AP2-5A* genotype on glume dimensions was analysed using an LMM (measurement ~ genotype + (1|plant\_id / spikelet\_number)) to include spikelet position as a random effect as glume size varies across the speltoid wheat spike<sup>56</sup>. The effect of *VRT-A2* and *AP2-5A* genotype on floral organ dimensions was analysed using an LMM (measurement ~ genotype + (1| plant\_id / spikelet\_number / floret\_number)) to include spikelet and floret position as nested random effects as floral organ size varies across the speltoid wheat spike and according to floret number<sup>56</sup>.

**4.3.3 *ap2-2 ap2-5A VRT-A2b* phenotyping**

*ap2-2 ap2-5A* seed was kindly provided by the Dubcovsky lab (UC Davis, USA) (Table 4-2). I crossed this line with a NIL carrying the *VRT-A2b* allele in a *T. turgidum* cv ‘Langdon’ background by Dr Nikolai Adamski and allowed the F<sub>1</sub> generation to self to generate an F<sub>2</sub> population segregating for all four loci (*ap2-2A* loss-of-function mutant, *ap2-2B* loss-of-function mutant, *AP2-5A* loss-of-function mutant, and *VRT-A2b* ectopic expression allele) (Table 4-2).

F<sub>1</sub> and F<sub>2</sub> plants were grown as described in Section 4.3.2.2. James Simmonds sampled tissue for DNA extraction, and the JIC Genotyping platform performed DNA extraction and PACE genotyping reactions as described in Sections 4.3.2.3 and 4.3.2.4. Sham ramification, number

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

of sterile bracts, glume width and length, and grain width and length were measured for primary F<sub>2</sub> spikes by me and Pamela Crane as described in Section 4.3.2.5.

Table 4-2: Details of germplasm used in this experiment. The mutation location refers to the location of interest and is relative to the IWGSC v1.0 assembly<sup>15</sup>. The gene IDs provided are according to the IWGSC v1.1 annotation<sup>98</sup>.

Accession	Mutation location (IWGSC v1.0 assembly)	Mutation of interest IWGSC v1.1 gene model	Genetic background	Reference
<i>ap2-2</i> <i>ap2-5A</i>	chr2A:738437789 chr2B:740216839 chr5A:650129824	TraesCS2A02G514200 TraesCS2B02G542400 TraesCS5A02G473800	Triticum turgidum cv ‘Kronos’	Debernardi, <i>et al.</i> <sup>42</sup>
<i>Langdon</i> <i>PI<sup>POL</sup></i>	N/A	TraesCS7A02G175200	Triticum turgidum cv Langdon	Adamski, <i>et al.</i> <sup>56</sup>

#### 4.3.4 sRNA-Seq

Details of sRNA-Seq methods can be found in Sections 2.3.1 and 2.3.2.

#### 4.3.5 miR172 stem-loop RT-qPCR

##### 4.3.5.1 Plant growth

*VRT-A2b* transgenic line seed (zero- and high-copy number) was kindly provided by Dr Nikolai Adamski. These lines are described in Adamski, *et al.*<sup>56</sup> (Table 4-3). In brief, these transgenic lines are in a *T. aestivum* cv ‘Fielder’ background, which contains an endogenous *VRT-A2a* allele. Both the zero-copy and high-copy number lines underwent transformation with a construct containing the genomic *VRT-A2b* extended and increased expression allele from *T. polonicum*. The zero-copy line contains no detectable transgene copies, while the high-copy line contains 9-35 transgene copies.

*vrt2* loss-of-function mutant and *VRT-A2* overexpression line seed was kindly provided by the Dubcovsky lab (UC Davis, USA) and are described in Li, *et al.*<sup>44</sup>. In brief, the *vrt2* loss-of-function mutant is in a *T. turgidum* cv ‘Kronos’ background and contains a premature stop codon in the fourth exon of the A-genome copy and a splice site mutation at the end of the fourth exon of the B-genome copy. The *VRT-A2* overexpression line is also in a *T. turgidum* cv ‘Kronos’ background and contains a transgenic copy of the *VRT-A2* coding region from *Tritium monococcum* driven by a constitutive *UBIQUITIN* promoter from *Z. mays*.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

Table 4-3: Details of germplasm used to quantify *miR172* expression using step-loop RT-qPCR. The mutation location refers to the location of interest and is relative to the IWGSC v1.0 assembly<sup>15</sup>. The gene IDs provided are according to the IWGSC v1.1 annotation<sup>98</sup>.

Accession	Mutation location (IWGSC v1.0 assembly)	Mutation of interest IWGSC v1.1 gene model	Genetic background	Reference
<i>VRT-A2</i> zero copy	N/A	N.A	<i>Triticum aestivum</i> cv 'Fielder'	Adamski, <i>et al.</i> <sup>56</sup>
<i>VRT-A2</i> high copy	N/A	N/A	<i>Triticum aestivum</i> cv 'Fielder'	Adamski, <i>et al.</i> <sup>56</sup>
<i>vrt2</i>	chr7A:128830572 chr7B:90193092	TraesCS7A02G175200 TraesCS7B02G080300	<i>Triticum turgidum</i> cv 'Kronos'	Li, <i>et al.</i> <sup>44</sup>
<i>VRT-A2</i> overexpression	N/A	N/A	<i>Triticum turgidum</i> cv 'Kronos'	Li, <i>et al.</i> <sup>44</sup>

Seeds were germinated in petri dishes as described previously<sup>268</sup>. Once germinated, seeds were sown in 15-cell trays containing 'John Innes Cereal Mix' as described in 2.3.1. All plants were grown in growth chambers at 65% humidity with 16 h photoperiods and 20 °C/15 °C day/night temperatures.

#### 4.3.5.2 Sampling

Three to five whole W5.0 spikes<sup>41</sup> per genotype were dissected as described in Section 2.3.1 into 2 mL microcentrifuge tubes on dry ice then flash-frozen in liquid nitrogen.

#### 4.3.5.3 RNA extraction

I added two flash-frozen 3 mm tungsten carbide beads (washed 1x using RNase Blitz, 1x 70% ethanol and 2x using ddH<sub>2</sub>O) to each sample tube. I ground the tissue in a Geno/Grinder (SPEX SamplePrep) at 1500 RPM for 30 s, then re-froze in liquid nitrogen. This was repeated two times (three total grinds).

I extracted RNA using the Direct-zol miniprep kit (Zymo Research) according to the manufacturer's instructions<sup>274</sup> including the on-column DNase I treatment. I quantified and checked the quality of the total RNA using a Nanodrop 1000 (Thermo).

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

**4.3.5.4 cDNA synthesis**

cDNA synthesis was based on the protocol in Chen, *et al.*<sup>165</sup>. I used SuperScript IV reverse transcriptase (Invitrogen) for cDNA synthesis according to the manufacturer's instructions<sup>275</sup> using 1 µg total RNA per reaction. I used a 0.05 µM final concentration (each) of the SLOmiR172 and snoR101\_R primers from Debernardi, *et al.*<sup>32</sup> (sequences in Section A.1, Table A-1). I included no template controls in the reaction.

**4.3.5.5 Stem-loop RT-qPCR**

The stem-loop RT-qPCR was based on the protocol in Chen, *et al.*<sup>165</sup> and Debernardi, *et al.*<sup>32</sup>. I used the Lightcycler 480 Sybr Green I Master mix (Roche). I set up reactions as follows: 5 µL Lightcycler 480 Sybr Green I Master mix (Roche), 2.5 µL 1:10 cDNA, 1.25 µL 2 µM forward primer, 2 µM reverse primer. I ran the reactions in a LightCycler 480 instrument (Roche) under the following conditions: 5 min at 95 °C; 45 cycles of 10 s at 95 °C, 15 s at 60 °C, 30 s at 72 °C, 1 s at 78 °C (single acquisition); dissociation curve from 60 °C to 95 °C (five acquisitions per °C) to determine primer specificity. *snoR101* was used as the reference gene. I determined primer efficiencies using a two-fold cDNA dilution series; the *snoR101* primers (Debernardi\_snoR101\_F and Debernardi\_snoR101\_R) had an efficiency of 90% and miR172 primers (Debernardi\_Uni\_MIRs and Debernardi\_RTmiR172a) had an efficiency of 99% (primer sequences in Section A.1, Table A-1).

I quantified miR172 expression for three biological replicates per genotype with three technical replicates per biological replicate. I included no template controls for each primer combination. I used *snoR101* as the reference gene. I calculated mean Ct values for each biological replicate and excluded any technical replicates which were >1 cycle different from the other two technical replicates or had two peaks in the melt curve analysis. I calculated relative expression using the  $2^{-\Delta\Delta Ct}$  method<sup>276</sup> from the mean value for each biological replicate.

**4.3.5.6 Statistical analysis**

I carried out statistical analyses using R v4.3.2<sup>123</sup>. I analysed the data in Kronos and Fielder backgrounds separately. I used the raw  $2^{-\Delta\Delta Ct}$  values for the analysis of Fielder background and log-transformed  $2^{-\Delta\Delta Ct}$  values for the analysis of Kronos background data. I used a one-way ANOVA with a Tukey post-hoc test to test for differences in miR172 expression between lines.

### 4.3.6 Generating VRT-A2b::FLAG transgenic lines

#### 4.3.6.1 LI Plasmid construction

To generate a FLAG-tagged *VRT-A2b* construct, I modified the pGGG-AH-VRT-A2 construct described in Adamski, *et al.*<sup>56</sup> which was kindly provided by the authors. The 3X FLAG tag sequence was kindly provided by Dr Estee Tee (JIC, UK). The tag sequence was synthesised by Genewiz in the context of 1826 bp of the *VRT-A2b* genomic sequence and 177 bp downstream sequence with MunI and OsiI restriction sites at the 5' and 3' ends respectively (sequence in Section A.4). This allowed me to clone the construct directly into pGGG-AH-VRT-A2.

I received an agar plate of *E. coli* carrying the pGGG-AH-VRT-A2 plasmid. I picked single colonies and used these to inoculate 30 µg/mL kanamycin LB starter cultures which I then incubated overnight at 37 °C, 200 RPM. I extracted plasmid DNA from the culture using a QIAprep spin miniprep kit (QIAGEN) according to the manufacturer's instructions<sup>208</sup>. I quantified plasmid yield using a Nanodrop 1000 (Thermo). I checked the plasmid was intact using a diagnostic digest. I combined 5 µL pGGG-AH-VRT-A2 plasmid DNA, 5 µL 10x rCutSmart buffer (New England Biolabs), 1 µL HindIII-HF (New England Biolabs) and 39 µL nuclease-free water. I mixed the reaction by pipetting and incubated at 37 °C for 1 h. I ran the digested and undigested plasmid on a 1% agarose gel at 120 V for 45 min. The observed banding pattern was as expected.

I received the synthesised FLAG tag construct in a pUC-GW-Amp vector. I transformed the plasmid into Library Efficiency DH5α *E. coli* (Invitrogen) according to the manufacturer's instructions<sup>214</sup> with the following modifications: I used 50 µL cells with 1 µL plasmid DNA, I used 450 µL SOC, and I incubated the cells for 1.5 h after transformation. I centrifuged the cells at 300 RCF for 3 min. 400 µL supernatant was removed and the cells were resuspended. I inoculated two 100 µg/mL carbenicillin LB agar plates (50 µL per plate). I incubated the plates overnight at 37 °C. I picked single colonies and used these to inoculate 100 µg/mL carbenicillin LB starter cultures which I then incubated overnight at 37 °C, 200 RPM. I made glycerol stocks from these cultures by storing 500 µL culture with 500 µl 40 % glycerol at -80 °C. I extracted plasmid from the remaining culture using a QIAprep spin miniprep kit (QIAGEN) according to the manufacturer's instructions<sup>208</sup>. I quantified plasmid yield using a Nanodrop 1000 (Thermo). The insert was confirmed by Sanger sequencing using M13F(-21) and M13R primers (Section A.1, Table A-1) by GENEWIZ.

I digested 1 µg each of the FLAG tag and pGGG-AH-VRT-A2 plasmids in a 20 µL reaction using 2 µL 10x FastDigest (clear for pGGG-AH-VRT-A2, green for FLAG tag) buffer (Thermo Fisher), 1 µL FastDigest OsiI and 1 µL FastDigest MunI. I digested the plasmids for

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

1 h at 37 °C. I phosphorylated the pGGG-AH-VRT-A2 digest product using rAPid Alkaline Phosphatase (Roche) according to the manufacturer's instructions <sup>216</sup>. I ran the FLAG tag plasmid digestion product on a 1% agarose gel (120 V, 1 h). I excised the band corresponding to the insert and extracted DNA using the Wizard SV Gel and PCR Clean-Up system (Promega). I used the phosphorylated vector and insert DNA in 5:1 and 3:1 (insert:vector) ligation reactions using T4 ligase (New England Biolabs) according to the manufacturer's instructions <sup>217</sup>. I incubated the ligation reaction at 4 °C overnight. I inactivated the ligase at 65 °C for 10 min, then used 10 µL reaction product to transform 5-alpha Competent *E. coli* (High Efficiency) (New England Biolabs) according to the manufacturer's high efficiency protocol <sup>218</sup> with the following modifications: I used 100 µL cells, I used 450 µL SOC medium, I incubated the transformed cells for 1.5 h at 200 RPM. I inoculated 30 µg/mL kanamycin LB agar plates with 100 µL cells and incubated them overnight at 37 °C. I picked single colonies and used these to inoculate 30 µg/mL kanamycin LB starter cultures which I then incubated overnight at 37 °C, 200 RPM. I made glycerol stocks from these cultures by storing 500 µL culture with 500 µl 40 % glycerol at -80 °C. I extracted plasmid from the remaining culture using a QIAprep spin miniprep kit (QIAGEN) according to the manufacturer's instructions <sup>208</sup>. I quantified plasmid yield using a Qubit dsDNA HS assay (Thermo Fisher) according to the manufacturer's instructions <sup>209</sup>. Plasmids were sequenced by Plasmidsaurus. The plasmid sequence was as expected (Figure 4-3) except for two SNPs in highly repetitive regions of the *OsAct1* promoter, which were likely sequencing errors.

**Chapter 4:** Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?

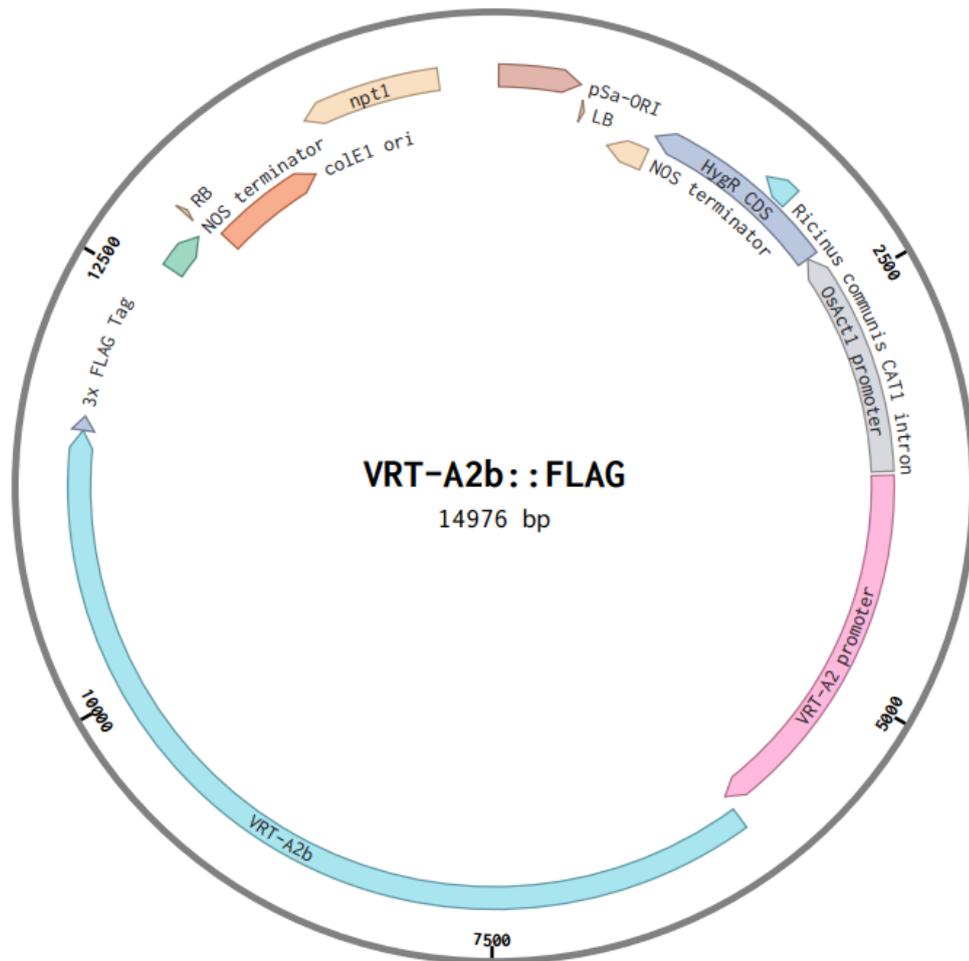


Figure 4-3: Plasmid map of VRT-A2b::FLAG (total size: 14, 976 base pairs (bp)). The plasmid contains the following features which are annotated on the map: RB (right border of T-DNA region); colE1 ori (origin of replication for ColE1 plasmids); nptI (neomycin phosphotransferase I gene from *E. coli* which confers resistance to kanamycin); pSa-ORI (origin of replication from the bacterial pSa plasmid); LB (left border of T-DNA region); NOS terminator (nopaline synthase terminator and nopaline synthase poly(A) signal); HygR (hygromycin resistance gene); *Ricinus communis* CAT1 intron; *OsAct1* promoter (promoter of the rice (*Oryza sativa*) actin1 gene); VRT-A2 promoter (promoter from the VRT-A2 gene from *Triticum polonicum*); VRT-A2b (open reading frame of the VRT-A2b gene from *Triticum polonicum* including introns); 3x FLAG tag.

#### 4.3.6.2 L2 plasmid construction

The VRT-A2b::FLAG construct was cloned into a level 2 pGoldenGreenGate plasmid<sup>230</sup> by Mark Smedley using Golden Gate assembly as described by Weber, *et al.*<sup>229</sup>.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

**4.3.6.3 Transformation**

Transformation of *T. aestivum* cv ‘Fielder’ with the L2 *VRT-A2b::FLAG* construct was performed by Sadiye Hayta. The L2 plasmid was co-electroporated with the helper plasmid pAL155<sup>277</sup> into *A. tumefaciens* AGL1 competent cells<sup>278</sup> as described in Hayta, *et al.*<sup>279</sup>. Single *A. tumefaciens* colonies were used to inoculate 10 mL 50 µg/mL kanamycin, 50 µg/mL rifampicin LB and incubated at 28 °C, 200 RPM for approximately 65 h. *A. tumefaciens* standard inoculums were prepared for transformation as described in Hayta, *et al.*<sup>279</sup>. *T. aestivum* cv ‘Fielder’ was transformed with the *VRT-A2b::FLAG* constructs as described in Hayta, *et al.*<sup>279</sup>.

Copy number analysis was performed using Taqman RT-qPCR as described in Hayta, *et al.*<sup>277</sup>. Transgene copy number was calculated using published methods<sup>276</sup>. T<sub>0</sub> and T<sub>1</sub> plants were grown as described in Section 3.3.6. T<sub>1</sub> seed was germinated as described in Simmonds, *et al.*<sup>106</sup>.

**4.3.6.4 Phenotyping**

I measured the glumes of T<sub>0</sub> primary spikes at physiological maturity as described in Section 4.3.2.5. I calculated the Pearson’s correlation coefficient between *VRT-A2b::FLAG* copy number and glume length using the `cor.test()` function in R v4.3.2<sup>123</sup>.

I measured the spike length and glumes of T<sub>1</sub> primary spikes at 14 dpa (days post anthesis, defined as visible anther extrusion). I measured spikes from the peduncle-rachis junction to the top of the terminal spikelet (excluding awns). I performed a one-way ANOVA and Tukey post-hoc test in R v4.3.2<sup>123</sup> to test for differences in spike length between lines. I measured glumes as described in Section 4.3.2.5. I analysed differences in glume length between using an LMM (`measurement ~ line + (1|spikelet_number)`) to include spikelet position as a random effect as glume size varies across the wheat spike<sup>56</sup>. I used the ‘lme4’ (Bates, *et al.*<sup>205</sup>) and ‘emmeans’<sup>221</sup> packages in R v4.3.2<sup>123</sup>.

**4.3.6.5 RT-qPCR**

I dissected total rachis and central glume (both glumes from three central spikelets) tissue from five biological replicates (T<sub>1</sub> plants) per line. Tissue was flash-frozen in liquid nitrogen and stored at -80 °C before proceeding.

Dr Yunchuan Liu extracted total RNA from all glumes and rachis samples. The tissue was ground using a mortar and pestle with liquid nitrogen. Total RNA was extracted from the ground samples using the Spectrum Plant Total RNA kit (Sigma) according to the manufacturer’s protocol<sup>280</sup> (Protocol A, 750 µL binding solution). Dr Yunchuan Liu DNase treated the samples using RQ1 RNase-free DNase (Promega) according to manufacturer’s

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

instructions<sup>281</sup>. 300  $\mu$ L of RNase-free water were added to each sample and then purified using the Spectrum Plant Total RNA kit (Sigma) according to the manufacturer's protocol<sup>280</sup> starting at step 4 (Protocol A, 750  $\mu$ L binding solution). Dr Yunchuan Liu synthesised cDNA from 1  $\mu$ g total RNA per sample using the SuperScript™ IV Reverse Transcriptase (Invitrogen) according to manufacturer's instructions<sup>275</sup>.

Dr Yunchuan Liu set up RT-qPCR reactions using LightCycler 480 SYBR Green I Master (Roche) according to manufacturer's instructions<sup>282</sup> for all samples and a no template control using *TaVRT-A2* primers with *ACTIN* as a reference gene (Section A.1, Table A-1). All reactions were performed with three technical replicates per sample. Reactions were run in a LightCycler 480 Instrument II (Roche) as described in Section 4.3.5.5 with an annealing temperature of 62 °C. I calculated relative expression values using the  $2^{-\Delta\Delta C_t}$  method<sup>276</sup> as described in Section 4.3.5.5.

I used log-transformed  $2^{-\Delta\Delta C_t}$  values for the statistical analysis of glume and rachis expression data. I used a one-way ANOVA with a Tukey post-hoc test in R v4.3.2<sup>123</sup> to test for differences in *VRT-A2* expression between lines.

#### 4.3.6.6 Western blot

Dr Yunchuan Liu carried out a Western blot on W5.0 spikes from 08.04, R1.01.02 and 01.01 T<sub>1</sub> *VRT-A2b::FLAG* plants. Spikes were dissected as described in section 2.3.1. A 3 mm tungsten carbide beads was placed in each tube. Samples were ground using a Genogrinder 2000 with liquid nitrogen, 28 RPM, 1 min. 300  $\mu$ L extraction buffer (125 mM Tris-HCl pH 6.8, 4% w/v SDS, 20% v/v glycerol, 2% v/v  $\beta$ -mercaptoethanol, 0.001% v/v bromophenol blue) was added to each tube and incubated on ice for 20 min. The samples were centrifuged at 4 °C, 12,000 RPM for 10 min.

An SDS-PAGE gel was prepared first by pouring an 8% separating gel (8% acrylamide/bis-acrylamide, 380 mM Tris-HCl pH 8.8, 0.1% v/v SDS, 0.1% v/v ammonium persulfate, 0.06% v/v TEMED) which was covered in absolute ethanol until solidified. The ethanol was poured off and stacking gel solution was poured over the top (4.98% acrylamide/bis-acrylamide, 126 mM Tris-HCl pH 6.8, 0.1% v/v SDS, 0.1% v/v ammonium persulfate, 0.1% v/v TEMED) and allowed to set.

40  $\mu$ L supernatant was combined with 10  $\mu$ L 5X SDS-loading buffer (0.1% v/v bromophenol blue, 60 mM Tris-HCl pH 6.8, 25% v/v glycerol, 5% v/v  $\beta$ -mercaptoethanol, 2% v/v SDS) and boiled for 10 min. 15  $\mu$ L of each sample was run on an SDS-PAGE gel at 200 V for 1 h. While the gel was running, a nitrocellulose membrane and filter paper was placed in transfer buffer (0.084% w/v NaHCO<sub>3</sub>, 0.032% w/v Na<sub>2</sub>CO<sub>3</sub>, 20% v/v methanol) for 20 min. Protein

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (*VRT-A2*) act via *microRNA172* to achieve its phenotypic effect?**

bands were transferred onto the nitrocellulose membrane at 4 °C or on ice from 2 h (200 mA) or overnight (100 mA). The membrane was blocked using blocking buffer (0.8% w/v NaCl, 20 mM Tris-HCl pH 7.6, 0.1% v/v Tween 20, 5% w/v skim milk powder) for 30 min at RT or 4 °C overnight. Monoclonal actin (mAbGEa) (Invitrogen #MA1-744) and ANTI-FLAG® M2 (Merck #F1804) primary antibodies were diluted to a 1:5,000 in blocking buffer. The membrane was probed with the diluted primary antibodies at 4 °C overnight or at RT for 2 h. The membrane was washed five to six times with TBST buffer (0.8% w/v NaCl, 20 mM Tris-HCl pH 7.6, 0.1% v/v Tween 20) at room temperature for 10 min per wash. Goat anti-mouse IgG H&L (HRP) secondary antibody (abcam # ab6789) was diluted 1:10,000 in blocking buffer. The membrane was incubated with the diluted secondary antibody at 4 °C overnight or at RT for 1 h. The membrane was washed five to six times with TBST buffer at room temperature for 10 min per wash. The Clarity and Clarity Max ECL Western Blotting Substrates kit (Bio-Rad) was used for detection according to the manufacturer's instructions<sup>283</sup>. Blots were imaged using an ImageQuant 8000 (Amersham).

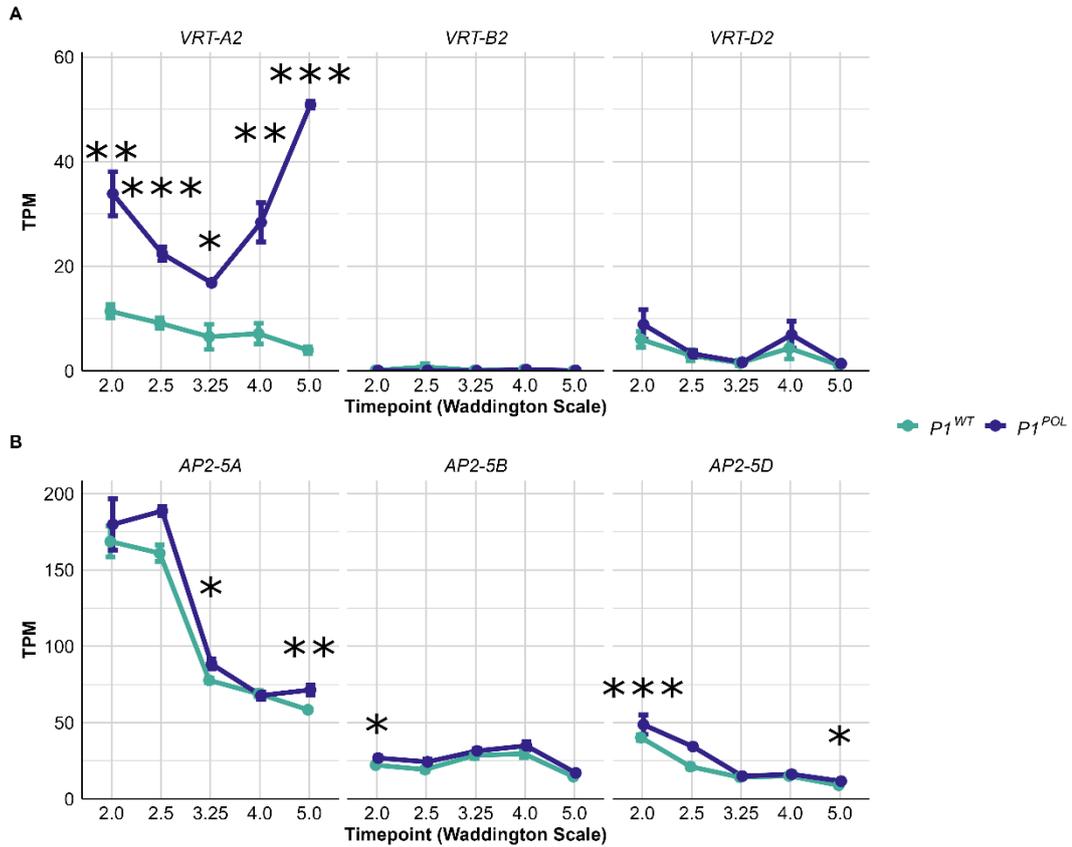
## 4.4 Results

### 4.4.1 *AP2-5* expression is significantly higher in a NIL with ectopic *VRT-A2* expression

As changes to *VRT-A2* and *AP2-5A* expression affect some of the same phenotypes, I used the Paragon RNA-Seq dataset generated by Dr Nikolai Adamski, Dr Anna Backhaus and Max Jones (see Section 3.4.2) to test for differences in *AP2-5A* expression in germplasm with different levels of *VRT-A2* expression.

RNA-Seq was carried out on spikes from *PI<sup>WT</sup>* and *PI<sup>POL</sup>* NILs at timepoints from W2.0 to W5.0 (Figure 4-4).

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**



**Figure 4-4: Expression of A) VEGETATIVE TO REPRODUCTIVE TRANSITION 2 (VRT2) and B) APETALA2-5 (AP2-5) homoeologues in developing spikes of  $P1^{WT}$  and  $P1^{POL}$  near isogenic lines (NILs) (both *T. aestivum* cv ‘Paragon’) carrying the VRT-A2a or VRT-A2b allele, respectively. Expression values were quantified using RNA-Seq and shown here in transcripts per million (TPM). Spike tissue was sampled at W2.0 (early double ridge), W2.5 (late double ridge), W3.25 (lemma primordia), W4.0 (terminal spikelet), and W5.0 (carpel extending round three sides of ovule, approximately 6 d after terminal spikelet) timepoints according to the Waddington scale<sup>41</sup>. One-way ANOVAs were used to assess differences in expression between NILs at each timepoint. Significance levels for the one-way ANOVAs are indicated with asterisks at each timepoint: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Mean expression values are plotted and error bars represent standard errors of the mean.  $n =$  four biological replicates per genotype.**

As shown by Adamski, *et al.*<sup>56</sup>, VRT-A2 expression is significantly higher in NILs containing the VRT-A2b ectopic expression allele ( $P1^{POL}$ ) compared to the ancestral VRT-A2a allele ( $P1^{WT}$ ) at all timepoints measured (one-way ANOVA: W2.0  $F(1,6) = 26.16$ ,  $p = 0.002$ ; W2.5  $F(1,6) = 65.91$ ,  $p < 0.001$ ; W3.25  $F(1,4) = 18.08$ ,  $p = 0.01$ ; W4.0  $F(1,6) = 25.05$ ,  $p = 0.002$ ; W5.0  $F(1,6) = 2451$ ,  $p < 0.001$ ) (Figure 4-4A). VRT-A2a expression decreased throughout development from W2.0 to W5.0. VRT-A2b expression also decreased from W2.0 to W3.25.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

After this, while *VRT-A2a* expression decreased, *VRT-A2b* expression increased once again. The W5.0 stage is later than any stage sampled in Adamski, *et al.*<sup>56</sup>, so provides some insight into how expression continues to increase after W4.0. *VRT-B2* (one-way ANOVA: W2.0  $F(1,6) = 0.72, p = 0.43$ ; W2.5  $F(1,6) = 1.22, p = 0.31$ ; W3.25  $F(1,4) = 4.24, p = 0.11$ ; W4.0  $F(1,6) = 0.002, p = 0.97$ ; W5.0  $F(1,6) = 2.56, p = 0.16$ ) and *VRT-D2* (W2.0  $F(1,6) = 0.78, p = 0.41$ ; W2.5  $F(1,6) = 0.06, p = 0.81$ ; W3.25  $F(1,4) = 0.06, p = 0.82$ ; W4.0  $F(1,6) = 0.60, p = 0.47$ ; W5.0  $F(1,6) = 0.89, p = 0.38$ ) expression was not significantly different at any stage in the two NILs.

*AP2-5A* expression started at a high level of around 169 TPM at W2.0, which then decreased to a still substantial level of expression of approximately 70 TPM from W3.25 onwards. There was a significant difference in expression between the NILs at the W3.25 (one-way ANOVA:  $F(1,4) = 8.95, p = 0.04$ ) and W5.0 (one-way ANOVA:  $F(1,6) = 14.94, p = 0.01$ ) timepoints, with *AP2-5A* expression being higher in the *PI<sup>POL</sup>* NIL. Expression of the pseudogene *AP2-5B* remained consistent throughout time at approximately 25 TPM. *AP2-5B* expression was higher at the W2.0 stage in the *PI<sup>POL</sup>* NIL (one-way ANOVA:  $F(1,6) = 7.45, p = 0.03$ ). *AP2-5D* expression followed a similar pattern to the A-genome homoeologue, decreasing until the W3.25 stage and remaining consistent after this point. *AP2-5D* expression was significantly higher in the *PI<sup>POL</sup>* NIL at the W2.5 (one-way ANOVA:  $F(1,6) = 51.24, p < 0.001$ ) and W5.0 (one-way ANOVA:  $F(1,6) = 11.27, p = 0.015$ ) stages. The general pattern of expression (high expression during double ridge timepoints, decreasing over time) is similar to that of *VRT-A2a*. Liu, *et al.*<sup>108</sup> detected no significant difference in *AP2-5A* or *AP2-5B* expression between *VRT-A2a* and *VRT-A2b* NILs. However, this transcriptome analysis was carried out at a single timepoint significantly later than the scope of this study; the spikes used by Liu, *et al.*<sup>108</sup> were 2.5-3.0 cm long once floral patterning is complete, while the spikes we used are at an earlier stage of development (W5.0 stage, 0.5 cm < spike length < 1.5 cm).

These data show that *AP2-5* is more highly expressed in NILs with increased and extended *VRT-A2* expression. This suggests that *VRT-A2* expression affects *AP2-5* expression, directly or indirectly.

#### 4.4.2 *ap2-5A VRT-A2b* plants are a near phenocopy of *ap2-5A ap2-2* mutants

To further investigate whether *VRT-A2* acts in the same genetic pathways as *AP2-5A*, Dr Nikolai Adamski made a cross between a *VRT-A2b* NIL in a tetraploid background and the *ap2-5A* loss-of-function TILLING lines K2726, K2992, and K3946. I genotyped and phenotyped three F<sub>3</sub> individuals per homozygous genotypic combination from these *VRT-A2b* x *ap2-5A* crosses.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

The phenotypes of the four different mutant combinations were visibly different (Figure 4-5). The wildtype (*VRT-A2a/AP2-5A*), *VRT-A2b/AP2-5A* and *VRT-A2a/ap2-5A* spikes were awned, while the *VRT-A2b/ap2-5A* spikes either had no awns or very short awns. Additionally, we observed sham ramification (where there is an extended rachilla with a large number of often sterile florets) in *VRT-A2a/ap2-5A* and *VRT-A2b/ap2-5A* spikes.

Plants with the *VRT-A2b* allele headed on average 2.12 d later (two-way ANOVA with pairwise comparisons:  $t(76) = -2.14$ ,  $p = 0.04$ , 95% CI [-4.09, -0.14]), their spikes were on average 25.7 mm longer (two-way ANOVA with pairwise comparisons:  $t(9) = -3.30$ ,  $p = 0.001$ , 95% CI [-43.30, -8.05]), and the plants were on average 11.1 cm taller (two-way ANOVA with pairwise comparisons:  $t(49) = -3.21$ ,  $p = 0.002$ , 95% CI [-18.00, -4.13]) than wildtype plants. Plants with the *ap2-5A* knockout allele headed 3.07 d earlier (two-way ANOVA with pairwise comparisons:  $t(76) = 3.10$ ,  $p = 0.002$ , 95% CI [1.1, 5.04]) and were on average 14.4 cm shorter (two-way ANOVA with pairwise comparisons:  $t(49.3) = 3.88$ ,  $p < 0.001$ , 95% CI [6.96, 21.9]) than wildtype plants. The phenotypes of *VRT-A2a/ap2-5A* and *VRT-A2b/AP2-5A* plants were consistent with those previously reported<sup>32,56</sup>.

**4.4.2.1 Some *ap2-5A/VRT-A2b* phenotypes were simply combinations of single mutant phenotypes**

The *VRT-A2b/ap2-5A* phenotype showed some similarities to *VRT-A2a/ap2-5A* and *VRT-A2b/AP2-5A* plants. *ap2-5A/VRT-A2b* showed the sham ramification phenotype, as seen in *VRT-A2a/ap2-5A*. I used two-way ANOVAs to assess whether there was any statistically significant interaction between *VRT-A2* and *AP2-5A* in respect to the quantitative phenotypes I observed. There was no significant interaction between *VRT-A2* and *AP2-5A* for the number of spikes per plant (two-way ANOVA:  $F(1,50) = 0.68$ ,  $p = 0.41$ ), the number of spikelets per spike (two-way ANOVA:  $F(1,22) = 2.04$ ,  $p = 0.17$ ), the number of sterile basal spikelets ( $F(1,8) = 0.00$ ,  $p = 1.00$ ), heading time (two-way ANOVA:  $F(1,75) = 1.51$ ,  $p = 0.22$ ), spike length (two-way ANOVA:  $F(1,8) = 0.71$ ,  $p = 0.42$ ), or plant height (two-way ANOVA:  $F(1,48.01) = 0.04$ ,  $p = 0.85$ ). Therefore, these phenotypes in *VRT-A2b/ap2-5A* plants were additive effects of the *VRT-A2b/AP2-5A* and *VRT-A2a/ap2-5A* single mutant phenotypes.

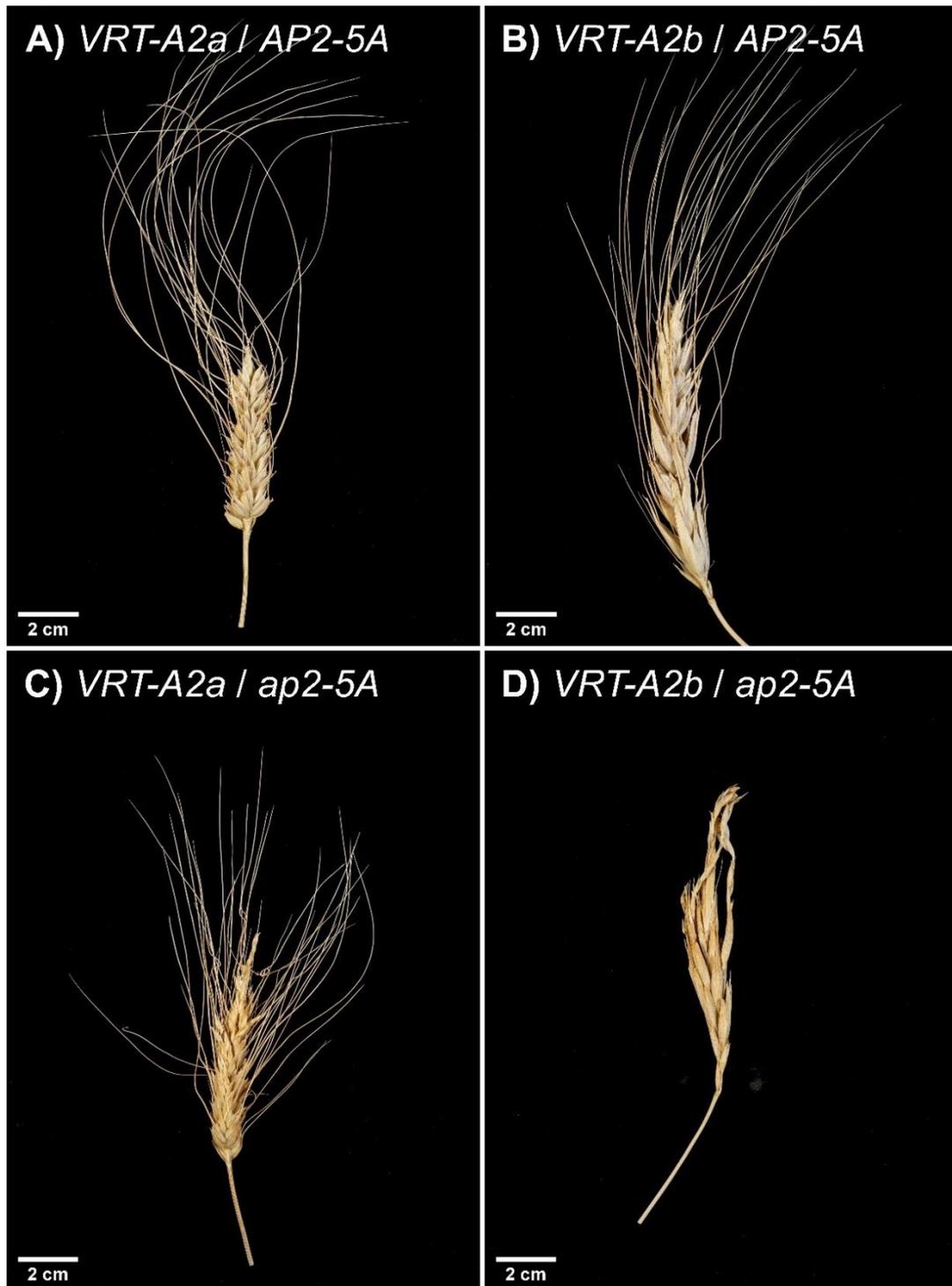


Figure 4-5: Primary  $F_3$  wheat spikes from the VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b x *apetala2-5A* (VRT-A2b x *ap2-5A*) cross at physiological maturity. The primary spike was the first spike which completely emerged from the flag leaf. The VRT-A2 and AP2-5A genotypes are indicated at the top of each image. Each line contains either the ancestral VRT-A2a allele, or the VRT-A2b increased and extended expression allele. The lines contain either the wildtype AP2-5A 'Q' allele, or an *ap2-5A* knockout allele which confers a premature stop codon.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

**4.4.2.2 There is a significant interaction between AP2-5A and VRT-A2 for glume width and length and grain width**

There were some phenotypes for which the *VRT-A2b/ap2-5A* phenotype was not simply an additive effect of the mutant phenotypes, providing evidence for a genetic interaction between *VRT-A2* and *AP2-5A*. The *VRT-A2* allele had a significant effect on glume length (two-way ANOVA:  $F(1, 7.92) = 15.38, p = 0.004$ ), while knocking out *AP2-5A* had no significant effect on glume length compared to the wildtype *AP2-5A* 'Q' allele (two-way ANOVA:  $F(1, 7.92) = 2.49, p = 0.15$ ) (Figure 4-6A). *VRT-A2b/ap2-5A* glumes were closer to the wildtype glume length, so the effect of the *VRT-A2b* allele was greater in an *AP2-5A* background than in an *ap2-5A* background (Figure 4-6C). This was supported by a statistically significant interaction between *VRT-A2* and *AP2-5A* genotypes in respect to glume length (two-way ANOVA:  $F(1, 7.92) = 5.69, p = 0.045$ ). Glumes were significantly narrower in *ap2-5A* spikes (two-way ANOVA:  $F(1, 7.93) = 22.95, p = 0.001$ ), while *VRT-A2b/AP2-5A* glumes were the same width as wildtype glumes (two-way ANOVA:  $F(1, 7.93) = 1.28, p = 0.29$ ) (Figure 4-6B). *VRT-A2b/ap2-5A* glumes were even narrower than *VRT-A2a/ap2-5A* glumes, so the *AP2-5A* genotype had a greater effect on glume width in a *VRT-A2b* background compared to in a *VRT-A2a* background (Figure 4-6D), consistent with the significant interaction between genotypes for glume width (two-way ANOVA:  $F(1, 7.93) = 8.83, p = 0.02$ ).

Both *VRT-A2* (two-way ANOVA:  $F(1, 8.24) = 28.66, p < 0.001$ ) and *AP2-5A* ( $F(1, 8.24) = 30.14, p < 0.001$ ) have a significant effect on grain length (Figure 4-7A). However, there is no significant interaction between *VRT-A2* and *AP2-5A* in respect to grain length (two-way ANOVA:  $F(1, 15.80) = 4.35, p = 0.054$ ), with the marginal *p*-value reflecting the magnitude of the effect rather than directionality. Grains were significantly narrower in spikes with the *ap2-5A* knockout allele compared to the wildtype *AP2-5A* 'Q' allele (two-way ANOVA:  $F(1, 12.74) = 18.25, p < 0.001$ ), and significantly wider in spikes with the *VRT-A2b* allele compared to *VRT-A2a* (two-way ANOVA:  $F(1, 12.74) = 7.48, p = 0.02$ ) (Figure 4-7B). *VRT-A2b/ap2-5A* grains were narrower than *VRT-A2a/ap2-5A* grains so, as with glume width, the *AP2-5A* genotype had a greater effect on glume width in a *VRT-A2b* background compared to in a *VRT-A2a* background (Figure 4-7C). There was a statistically significant interaction between *VRT-A2* and *AP2-5A* genotypes in respect to grain width (two-way ANOVA:  $F(1, 12.74) = 8.27, p = 0.01$ ). However, the grain morphology data should be interpreted with caution, as the three *VRT-A2b/ap2-5A* spikes produced only five grains in total (compared to 77-100 for the other three genotypes). Therefore, a much larger number of spikes would need to be dissected and analysed to draw firm conclusions on the interaction between *VRT-A2* and *AP2-5A* in respect to grain width and length.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

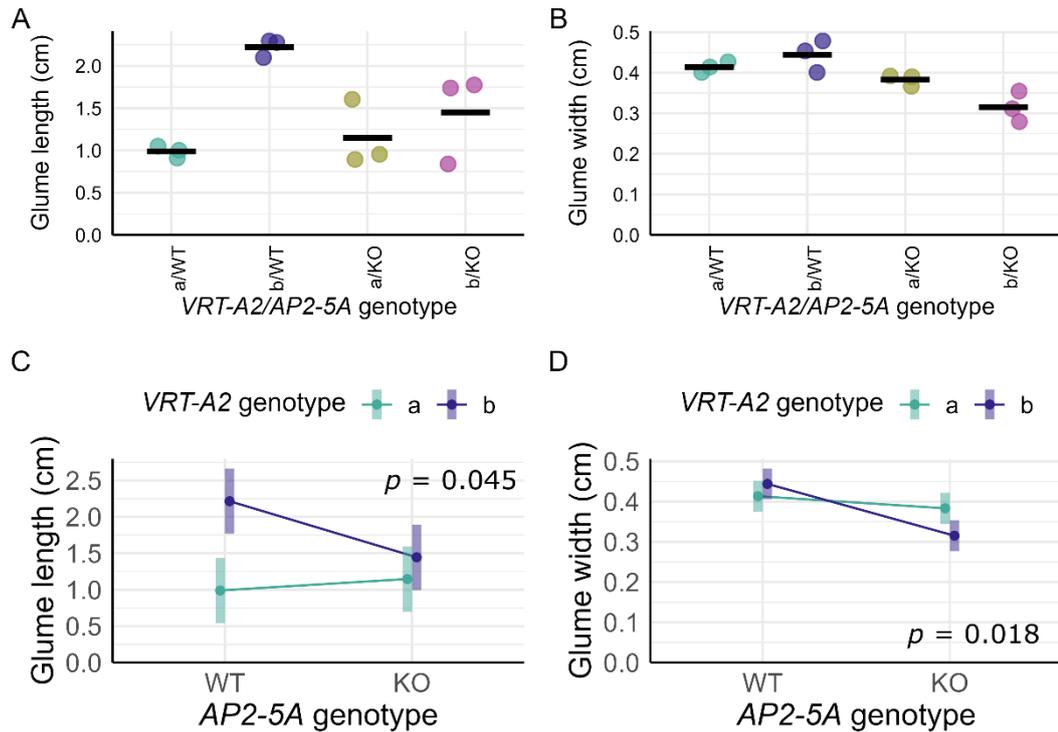


Figure 4-6: Glume size phenotypes of  $F_3$  primary wheat spikes from a VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b x *apetala2-5A* (VRT-A2b x *ap2-5A*) cross. Each line contains either the ancestral VRT-A2a allele, or the VRT-A2b increased and extended expression allele (denoted by an 'a' or 'b' respectively). The lines contain either the wildtype AP2-5A 'Q' allele, or an *ap2-5A* knockout allele which confers a premature stop codon (denoted by a 'WT' or 'KO' respectively).  $n =$  three plants per genotype combination. (A, C) Glume length in cm, (B, D) glume width in cm. In A) and B) values for individual plants are plotted with coloured circles and the mean for  $n=3$  plants is shown by a black horizontal bar. C) and D) show the interaction between VRT-A2 and AP2-5A genotypes for glume length and width, respectively. Shaded areas depict the 95% confidence interval. The  $p$  value for the interaction term in a linear mixed model is shown in the corner of the interaction plots.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

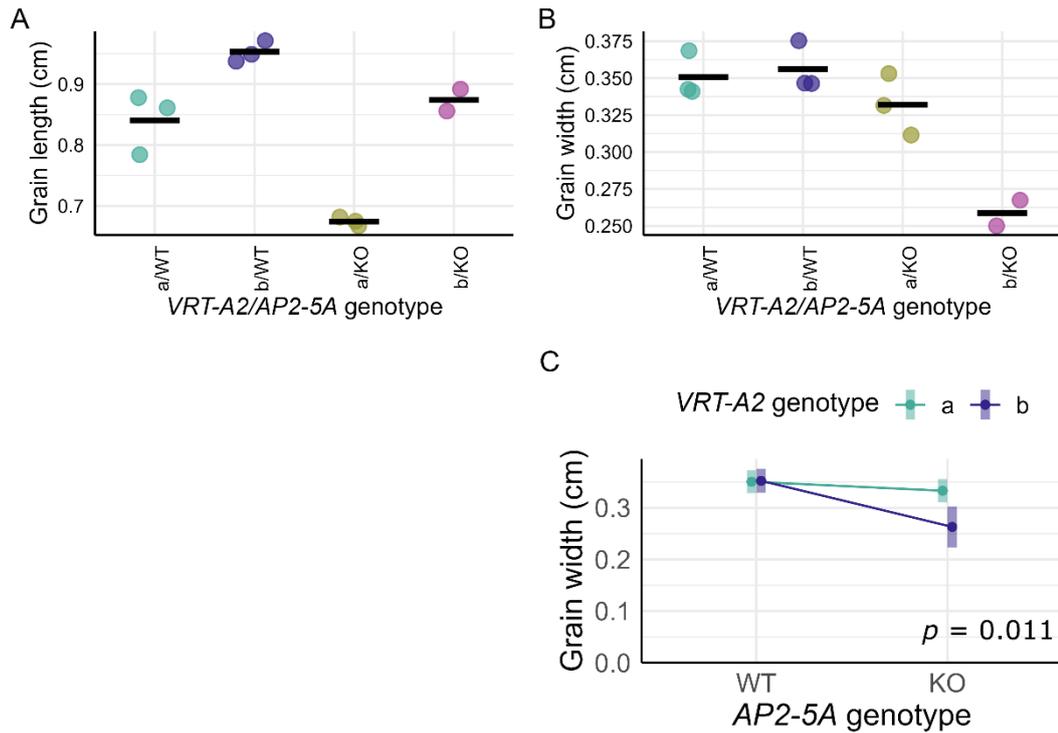


Figure 4-7: Grain size phenotypes of  $F_3$  primary wheat spikes from a VEGETATIVE TO REPRODUCTIVE TRANSITION-2b x *ap2-5A* (*VRT-A2b* x *ap2-5A*) cross. Each line contains either the ancestral *VRT-A2a* allele, or the *VRT-A2b* increased and extended expression allele (denoted by an 'a' or 'b' respectively). The lines contain either the wildtype *AP2-5A* 'Q' allele, or an *ap2-5A* knockout allele which confers a premature stop codon (denoted by a 'WT' or 'KO' respectively).  $n =$  three plants per genotype combination. A) Grain length in cm, B) grain width in cm. In A) and B) values for individual plants are plotted using coloured circles and the mean for  $n = 3$  plants is shown by a black horizontal bar. C) shows the interaction between *VRT-A2* and *AP2-5A* genotypes for grain width. There was no significant interaction between *VRT-A2* and *AP2-5A* for grain length. Shaded areas depict the 95% confidence interval. The  $p$  value for the interaction term in a linear mixed model is shown in the corner of the interaction plot.

#### 4.4.2.3 *VRT-A2b/ap2-5A* plants phenocopy *ap2-2/ap2-5A* mutants

The most striking phenotype in *VRT-A2b/ap2-5A* plants was the high number of sterile bracts I observed in each spikelet (Figure 4-8). *VRT-A2a/ap2-5A* plants have been shown previously to have one or two sterile bracts between the glumes and first lemma of basal spikelets<sup>32</sup>. These bracts have glume-like identities with pronounced keels and short awns<sup>32</sup>. These sterile bracts contain no floral structures (e.g. palea, lodicules, stamens or carpels) between them. *VRT-A2b/AP2-5A* plants have not been reported to have sterile bracts and I did not observe any in the *VRT-A2b/AP2-5A* plants in this study. However, in *VRT-A2b/ap2-5A* plants, the

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

*ap2-5A* single mutant phenotype was exacerbated. There were on average 1.6 bracts per *VRT-A2a/ap2-5A* spikelet whereas I found 6.9 bracts per *VRT-A2b/ap2-5A* spikelet. The majority of *VRT-A2b/ap2-5A* spikelets consisted only of two glumes and numerous bracts (*i.e.*, there was an absence of floral structures) (Figure 4-9). A very small number of florets were observed within *VRT-A2b/ap2-5A* spikelets. While *VRT-A2a/ap2-5A* bracts were almost always found immediately internal to glumes, *VRT-A2b/ap2-5A* bracts could be found after florets on the rachilla. *VRT-A2b/ap2-5A* produced very few grains; there were five grains total in the three *VRT-A2b/ap2-5A* spikes I dissected. In some instances, *VRT-A2b/ap2-5A* florets contained two carpels within the same floret.

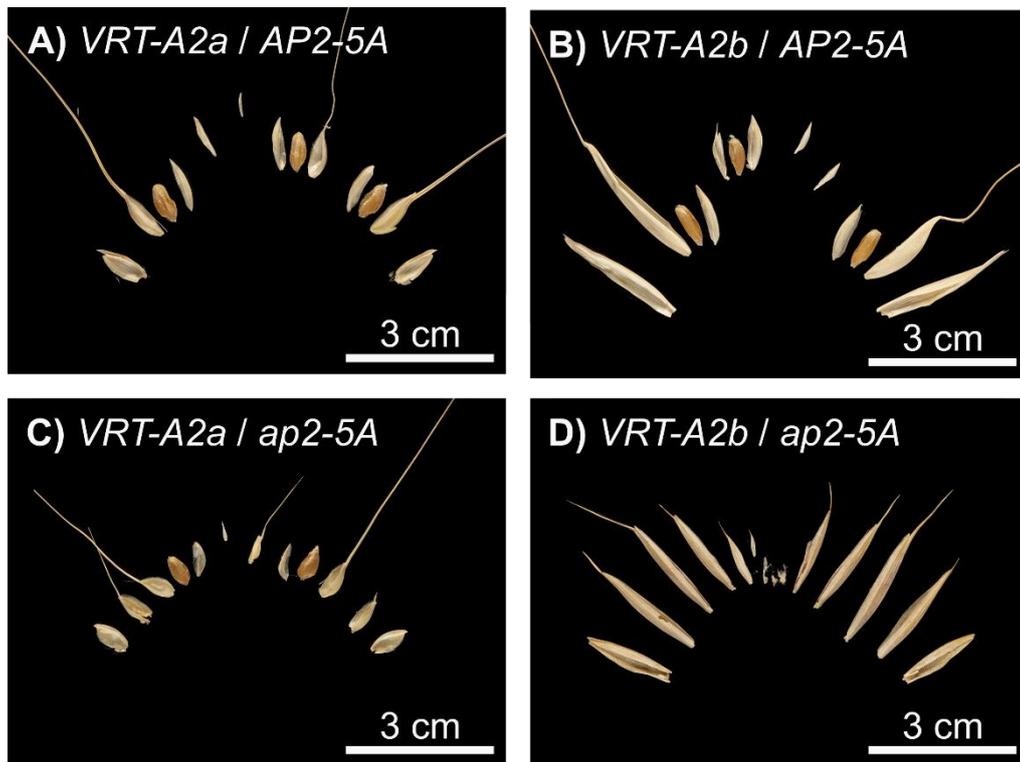


Figure 4-8: Dissected central spikelets from *F3* primary wheat spikes from the VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b x *apetala2-5A* (*VRT-A2b* x *ap2-5A*) cross. The primary spike was the first spike which completely emerged from the flag leaf. The *VRT-A2* and *AP2-5A* genotypes are indicated at the top of each image. Each line contains either the ancestral *VRT-A2a* allele, or the *VRT-A2b* increased and extended expression allele. The lines contain either the wildtype *AP2-5A* 'Q' allele, or an *ap2-5A* knockout allele which confers a premature stop codon.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

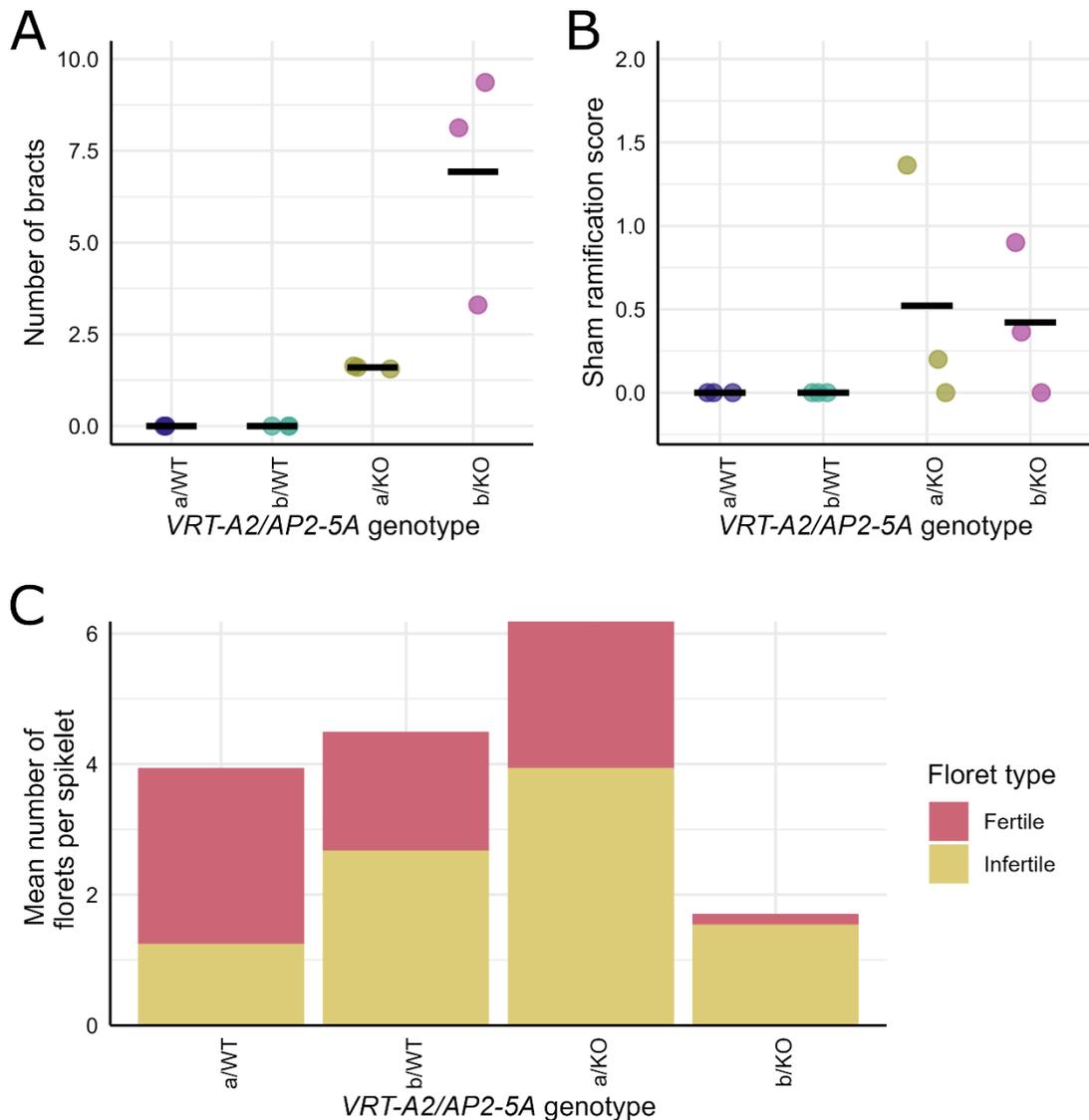


Figure 4-9: Spikelet-level phenotypes of  $F_3$  primary wheat spikes from the VEGETATIVE TO REPRODUCTIVE TRANSITION- $A2b \times ap2-5A$  ( $VRT-A2b \times ap2-5A$ ) cross. A) Mean number of empty bracts (i.e. containing no visible floral organs) per spikelet, B) mean sham ramification score per spikelet (0 = none, 1 = moderate and 2 = extreme) and C) mean number of fertile (defined as grain-bearing) and infertile florets (defined as florets containing floral organs which did not produce a grain) per spikelet.  $n =$  three plants per genotype combination. Each line contains either the ancestral  $VRT-A2a$  allele, or the  $VRT-A2b$  increased and extended expression allele (denoted by an 'a' or 'b' respectively). The lines contain either the wildtype  $AP2-5A$  'Q' allele, or an  $ap2-5A$  knockout allele which confers a premature stop codon (denoted by a 'WT' or 'KO' respectively).

The observations I made of  $VRT-A2b \times ap2-5A$   $F_3$  plants suggest that  $VRT-A2$  genetically interacts with  $AP2-5A$ , either directly or indirectly. The changes seen in  $VRT-A2b/ap2-5A$  spike architecture are more severe than those seen in either  $VRT-A2b$  or  $ap2-5A$  alone; the

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (*VRT-A2*) act via *microRNA172* to achieve its phenotypic effect?**

presence of the *VRT-A2b* allele increases the effect of the *ap2-5A* loss-of-function mutant. This result supports the significant difference in *AP2-5* expression I observed in *PI<sup>WT</sup>* and *PI<sup>POL</sup>* spikes. The RNA-Seq data shows that higher/ectopic *VRT-A2* expression results in higher *AP2-5A* expression.

Additionally, the presence of double carpels and the high number of bracts I observed in *VRT-A2b/ap2-5A* spikes were phenotypes also observed by Debernardi, *et al.*<sup>42</sup> in *ap2-5/ap2-2* spikes (Figure 4-10). This supports my hypothesis that the long glume and long grain phenotype reported in *VRT-A2b* plants by Adamski, *et al.*<sup>56</sup> (and confirmed here) are generated via a direct or indirect interaction with *AP2-2*.

## A *ap2-2 ap2-5A*



## B *VRT-A2b/ap2-5A*



Figure 4-10: (A) *apetala2-2 apetala2-5A* wheat spike and dissected spike, containing knockout alleles of *APETALA2-2A*, *APETALA2-2B*, and *APETALA2-5A*. Reproduced from Debernardi, et al. <sup>42</sup>. (B) *VRT-A2b/ap2-5A* (*VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b / apetala2-5A*) wheat spike and dissected spike, containing the increased and extended expression *VRT-A2b* allele and a knockout allele of *AP2-5A*. Glumes are labelled *G1* and *G2*, and the first two lemma are labelled *L1* and *L2*. Glume-like sterile bracts are highlighted with asterisks.

#### 4.4.3 *ap2-2 ap2-5A VRT-A2b* lines will provide further insight into the interactions between these key genes

To further interrogate this genetic network, I crossed the *ap2-2 ap2-5A* line from Debernardi, *et al.* <sup>42</sup> with a NIL carrying the *VRT-A2b* ectopic expression allele. Using the F<sub>2</sub> segregating population I generated, I aimed to understand the genetic interactions between *AP2-2*, *AP2-5A* and *VRT-A2*.

When generating this population, I aimed to interrogate any possible interaction between *VRT-A2* and *AP2-2* to test the hypothesis I described in Section 4.2.5; that *VRT-A2* negatively regulates *AP2-2* (either directly or indirectly). Unfortunately, no F<sub>2</sub> individuals were recovered with the *VRT-A2b* allele and functional copies of *AP2-2* and *AP2-5A*. Therefore, I could not test for any genetic interaction between *VRT-A2* and *AP2-2*.

I did, however, use this phenotyping data to show that in wheat, an *ap2-2* loss-of-function mutant has longer grains (*t* test:  $t(3) = 1.70, p = 0.19$ ) and narrower glumes (*t* test:  $t(3) = -2.44, p = 0.09$ ) than wildtype plants, although these differences were non-significant. This finding that is consistent with studies in barley <sup>57</sup>. However, knocking out *AP2-2* had no effect on grain width (*t* test:  $t(3) = -0.32, p = 0.77$ ) or glume length (*t* test:  $t(3) = -0.10, p = 0.92$ ), which is inconsistent with the evidence from barley <sup>57</sup>. The sample size for each genotype was small ( $n = 3$  for *ap2-2* and  $n = 2$  for the wildtype). Therefore, this germplasm should be taken forward to the next generation to recover all possible genotype combinations (including *VRT-A2b AP2-5A AP2-2*). This would allow us to test for interactions between all three genes and would increase the statistical power of any phenotypic analyses.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

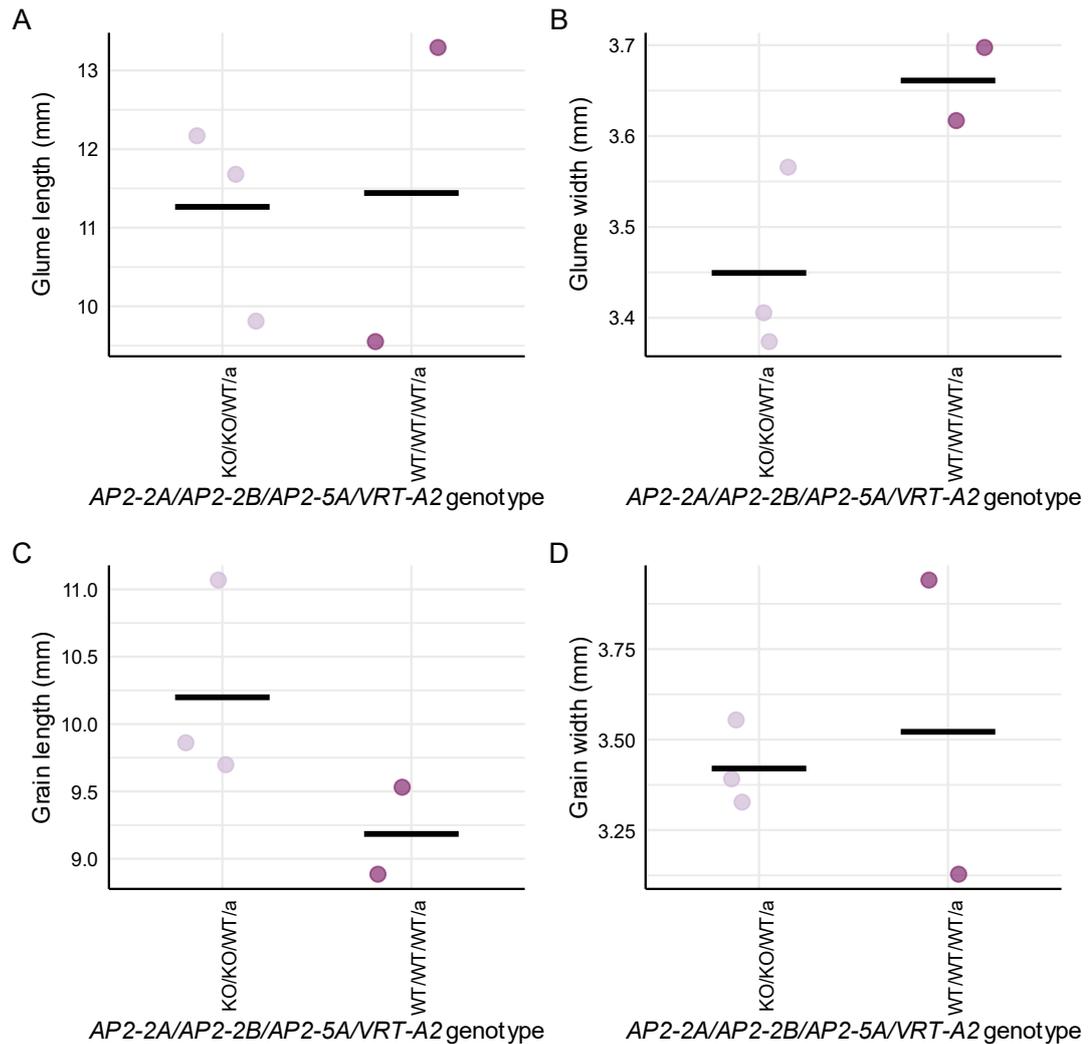
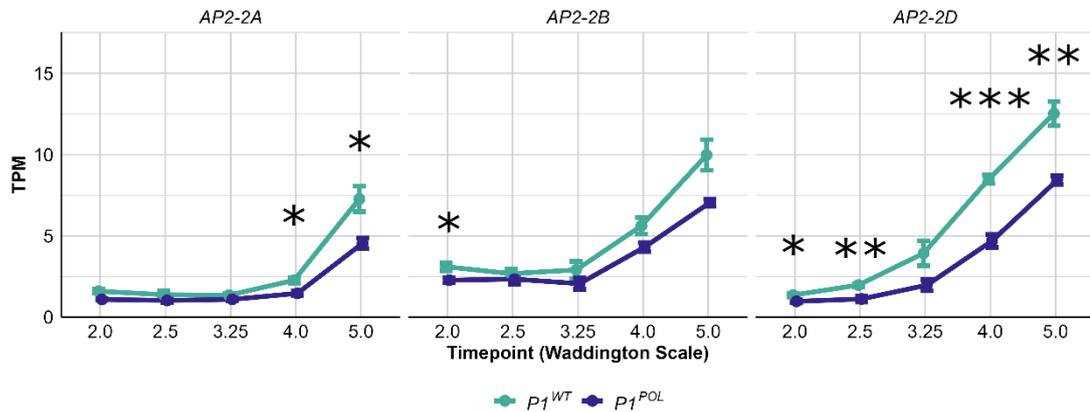


Figure 4-11: Grain size phenotypes of  $F_3$  primary wheat spikes from a VEGETATIVE TO REPRODUCTIVE TRANSITION- $A2b$   $\times$  *apetala2-5A*  $\times$  *apetala2-2* (VRT- $A2b$   $\times$  *ap2-5A*  $\times$  *ap2-2*) cross.  $n =$  three plants for *ap2-2*, two plants for the wildtype. A) Glume length in mm, B) glume width in mm, C) grain length in mm, D) grain width in mm. Mean values for individual plants are plotted and the mean for  $n = 2$  or 3 plants is shown by a black horizontal bar. Each line contains either the ancestral VRT- $A2a$  allele, or the VRT- $A2b$  increased and extended expression allele (denoted by an 'a' or 'b' respectively). The lines contain either the wildtype AP2-5A 'Q' allele, or an *ap2-5A* knockout allele which confers a premature stop codon (denoted by a 'WT' or 'KO' respectively). The lines contain either wildtype AP2-2A and AP2-2B alleles (denoted by a 'WT'), or *ap2-2A* and *ap2-2B* alleles with splice site mutations which introduce a frameshift or affect transcript stability respectively (denoted by a 'KO').

#### 4.4.4 AP2-2 expression is significantly lower in a NIL with ectopic VRT-A2 expression

To further test the hypothesis that *VRT-A2* acts by reducing *AP2-2* expression, I looked at the Paragon RNA-Seq dataset I described in Section 4.4.1. I observed that *AP2-2A* expression is lower in  $P1^{POL}$  spikes at multiple timepoints (Figure 4-12). *AP2-2A* expression is significantly lower in  $P1^{POL}$  spikes at W4.0 (one-way ANOVA:  $F(1, 6) = 13.53, p = 0.01$ ) and W5.0 (one-way ANOVA:  $F(1, 6) = 10.30, p = 0.018$ ), *AP2-2B* expression is significantly lower at W2.0 (one-way ANOVA:  $F(1, 6) = 7.69, p = 0.032$ ), and *AP2-2D* expression is significantly lower at W2.0 (one-way ANOVA:  $F(1, 6) = 7.03, p = 0.038$ ), W2.5 (one-way ANOVA:  $F(1, 6) = 16.53, p = 0.007$ ), W4.0 (one-way ANOVA:  $F(1, 6) = 64.34, p < 0.001$ ) and W5.0 (one-way ANOVA:  $F(1, 6) = 26.90, p = 0.002$ ). This result supports my hypothesis that *VRT-A2* negatively regulates *AP2-2* expression.



*Figure 4-12: Expression of APETALA2-2 (AP2-2) homoeologues in developing wheat spikes of  $P1^{WT}$  and  $P1^{POL}$  near isogenic lines (both *T. aestivum* cv ‘Paragon’, carrying the *VRT-A2a* and *VRT-A2b* alleles, respectively). Tissue was sampled at W2.0 (early double ridge), W2.5 (late double ridge), W3.25 (lemma primordia), W4.0 (terminal spikelet), and W5.0 (carpel extending round three sides of ovule, approximately 6 d after terminal spikelet) timepoints according to the Waddington scale <sup>41</sup>. One-way ANOVAs were used to assess differences in expression between NILs at each timepoint. Significance levels are indicated with asterisks: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Mean expression values are plotted, and error bars represent standard errors of the mean.  $n =$  four biological replicates per genotype.*

The fact that *AP2-2* expression is lower in  $P1^{POL}$  spikes with higher/ectopic *VRT-A2* expression is a surprising result, as expression of the closely related *AP2-5* gene is higher in  $P1^{POL}$  spikes. However, as I showed in Chapter 3, *AP2-5* and *AP2-2* have different expression profiles, although they are both regulated by miR172.

#### 4.4.5 *vrt2* spikes have higher miR172 abundance

As there is phenotypic and genetic evidence that *VRT-A2* expression affects *AP2-2* and *AP2-5* expression, I hypothesized that this interaction may be via miR172, which regulates both *AP2-2* and *AP2-5*. To test this hypothesis, I used the sRNA-Seq dataset described in Chapter 2 to quantify miR172 expression in *P1<sup>WT</sup>* and *P1<sup>POL</sup>* spikes from W2.5 to W5.0 (Figure 4-13). miR172a and miR172b expression was lower in *P1<sup>POL</sup>* spikes, however this difference was not statistically significant (LMM followed by one-way ANOVA: miR172a W2.5  $F(1,1.70) = 7.50$ ,  $p = 0.13$ ; miR172a W4.0  $F(1,4) = 1.34$ ,  $p = 0.31$ ; miR172a W5.0  $F(1,4.03) = 2.18$ ,  $p = 0.21$ ; miR172b W2.5  $F(1,1.46) = 5.03$ ,  $p = 0.20$ ; miR172b W4.0  $F(1,4) = 3.76$ ,  $p = 0.13$ ; miR172b W5.0  $F(1,4.01) = 5.81$ ,  $p = 0.07$ ).

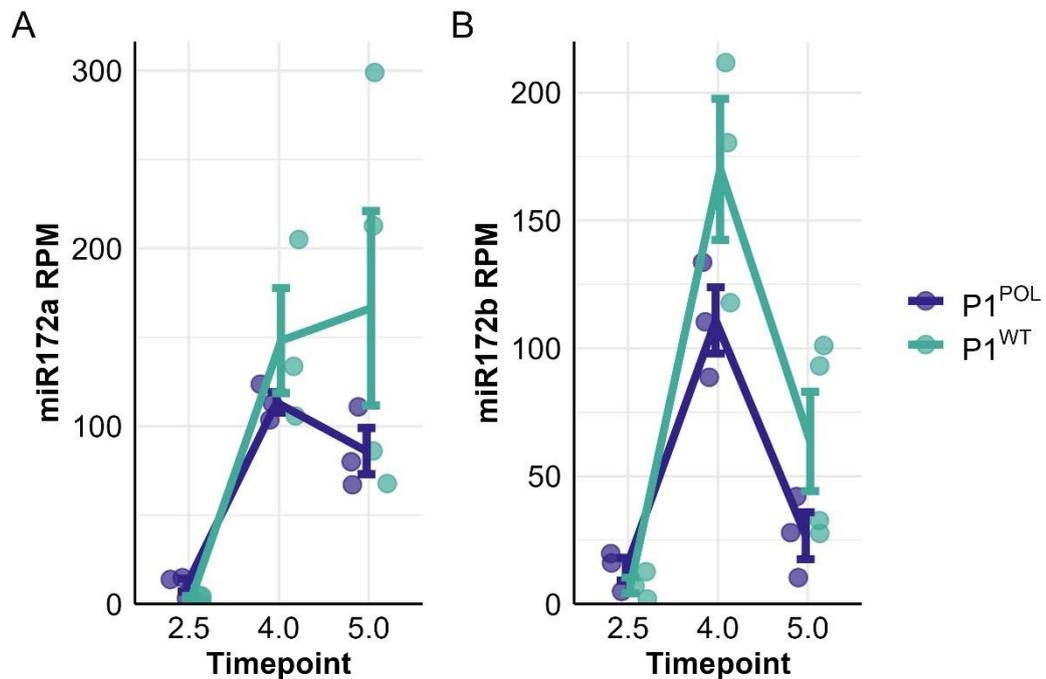


Figure 4-13: A) *microRNA172a* (*miR172a*) and B) *microRNA172b* (*miR172b*) expression in *P1<sup>WT</sup>* and *P1<sup>POL</sup>* near isogenic line (NIL) spikes at W2.5 (double ridge), W4.0 (terminal spikelet) and W5.0 (carpel extending round three sides of ovule, approximately six days after terminal spikelet) according to the Waddington scale<sup>41</sup>. The *P1<sup>WT</sup>* and *P1<sup>POL</sup>* NILs are both *Triticum aestivum* cv ‘Paragon’, carrying the *VRT-A2a* and *VRT-A2b* alleles, respectively.  $n =$  three biological replicates per genotype, except for *P1<sup>WT</sup>* at W5.0 where  $n =$  four biological replicates. Mean expression values are plotted, and error bars represent standard errors of the mean.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via *microRNA172* to achieve its phenotypic effect?**

As *PI<sup>POL</sup>* NILs show a relatively subtle increase in *VRT-A2* expression compared to the wildtype, I quantified miR172 expression in germplasm with greater variation in *VRT-A2* expression using stem-loop RT-qPCR (Figure 4-14). A high copy number (9-35 copies) *VRT-A2b* transgenic line (described in Adamski, *et al.* <sup>56</sup>) did not have a significantly different level of miR172 expression to a zero-copy line or wildtype Fielder (one-way ANOVA:  $F(2,6) = 1.44, p = 0.31$ ) (Figure 4-14A). Weak and intermediate *VRT-A2* overexpression lines (under the control of the maize *UBIQUITIN* promoter, described in Li, *et al.* <sup>44</sup>) had slightly lower miR172 levels than wildtype Kronos, although this difference was not significant (one-way ANOVA followed by Tukey post-hoc test:  $p = 0.50$ , 95% CI [-1.70, 0.63] and  $p = 0.77$ , CI [-1.52, 0.81], respectively) (Figure 4-14B). A *vrt2* loss-of-function mutant (described in Li, *et al.* <sup>44</sup>) had significantly higher miR172 expression than wildtype Kronos (Figure 4-14B) (one-way ANOVA followed by Tukey post-hoc test:  $p = 0.03$ , 95% CI [0.11, 2.44]). This result is consistent with evidence from Arabidopsis that *AtSVP* negatively regulates miR172a and that increasing *AtSVP* expression produces a subtler effect than that observed in an *svp* loss-of-function mutant <sup>284</sup>.

#### 4.4.6 Development of *VRT-A2b::FLAG* transgenics

*VRT-A2* is a MADS box transcription factor and therefore contains a DNA binding domain. To establish whether *VRT-A2* binds directly to *AP2-2*, *AP2-5* or *MIR172*, I developed a *VRT-A2b* construct with a C-terminus FLAG tag. This would allow me to use a technique called CUT&RUN (cleavage under targets & release using nuclease) to pull-down and sequence DNA which the *VRT-A2::FLAG* protein binds to *in planta* at different stages of development.

The details of the *VRT-A2b::FLAG* construct are provided in Section 4.3.6. In brief, the full genomic *VRT-A2b* allele sequence (from the pGGG-AH-*VRT-A2* construct described in Adamski, *et al.* <sup>56</sup>) was fused to a 3X FLAG tag sequence in frame before the stop codon in order to produce a tagged fusion protein. 2,299 bp upstream of the *VRT-A2b* start codon and 1,000 bp downstream of the *VRT-A2b* stop codon was also included to capture any putative *cis*-regulatory elements. By using the *VRT-A2b* allele to transform plants carrying the *VRT-A2a* allele, the phenotype of the transgenic lines would provide a strong indication of whether the fusion protein is functional.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

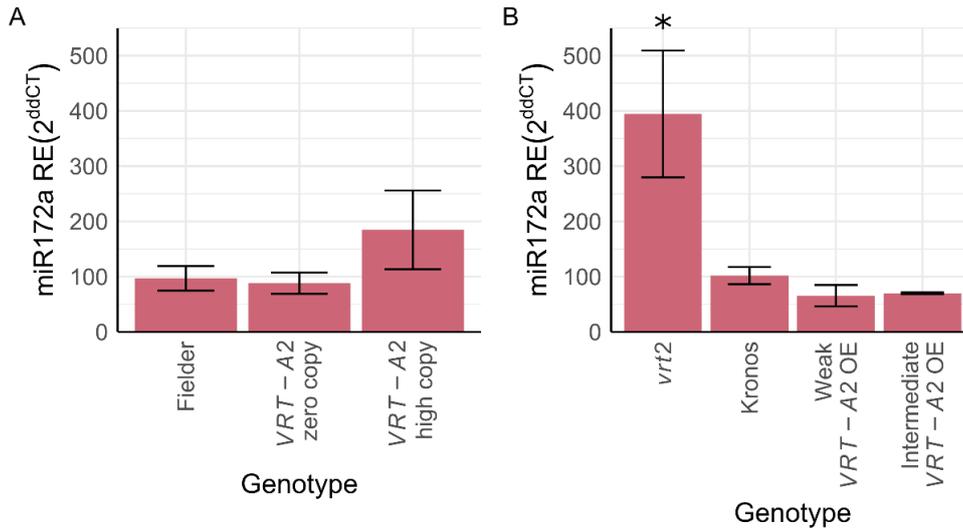


Figure 4-14: microRNA172 (miR172) expression in wheat germplasm with differential VEGETATIVE TO REPRODUCTIVE TRANSITION-A2 (VRT-A2) expression in a (A) *Triticum aestivum* cv 'Fielder' (hexaploid) and (B) *Triticum turgidum* cv 'Kronos' (tetraploid) background. The zero and high copy VRT-A2 lines are transgenic lines containing zero and 9-35 copies of the VRT-A2b allele, respectively. *vrt2* contains loss-of-function mutations in VRT-A2 and VRT-B2. The 'OE' VRT-A2 overexpression transgenic lines are independent lines which contain VRT-A2 under the control of the constitutive *Zea mays* UBIQUITIN promoter.  $n =$  three biological replicates per genotype. Error bars depict the standard error of the mean. The asterisks denote that miR172a relative expression in *vrt2* spikes is significantly higher than for in Kronos wildtype spikes according to a one-way ANOVA ( $p = 0.03$ , 95% CI [0.11, 2.44]).

**4.4.6.1 VRT-A2b::FLAG transgenics have expected phenotypes**

The *VRT-A2b::FLAG* construct was transformed into hexaploid Fielder donor material by Mark Smedley and Sadiye Hayta. We obtained 16 independent T<sub>0</sub> lines with 0-8 copies of the *VRT-A2b::FLAG* construct. To assess the phenotypic effect of differing copy numbers, I measured the glumes of T<sub>0</sub> spikes at maturity (Figure 4-15). The donor material (Fielder) carries the *VRT-A2a* allele, therefore introducing the *VRT-A2b::FLAG* construct should lead to an increase in glume length as observed by Adamski, *et al.*<sup>56</sup> and in Section 4.4.2. Glume length correlated with copy number ( $r(13) = 0.64$ ,  $p = 0.01$ ) in the T<sub>0</sub> plants. This suggests that the *VRT-A2b::FLAG* construct behaves in a similar way to the natural *VRT-A2b* allele described in Adamski, *et al.*<sup>56</sup> and that we can use this construct to draw relevant biological conclusions.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

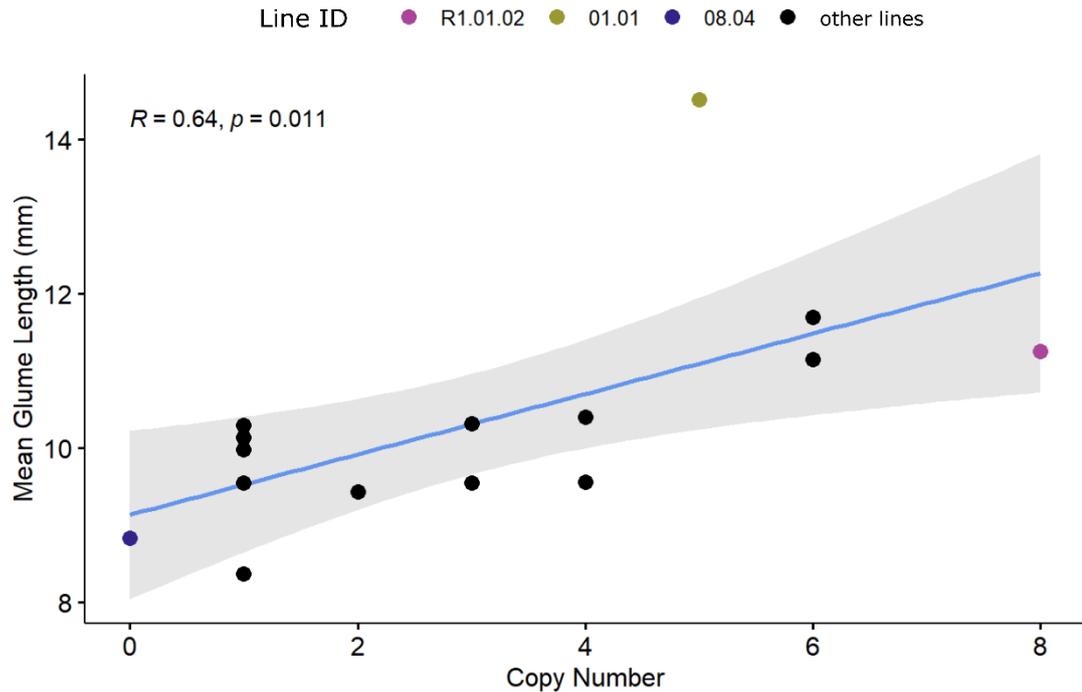


Figure 4-15: Mean glume length from  $T_0$  primary wheat spikes (*Triticum aestivum* cv 'Fielder'), plotted against VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b::FLAG copy number.  $n = 15$   $T_0$  plants. Regression (blue line) and 95% confidence interval (grey shading) are shown.  $T_0$  individuals taken forward to the  $T_1$  generation for further analysis are coloured in dark blue (line 08.04), green (line 01.01) and pink (line R1.01.02), respectively. .

I selected three independent lines with zero (line 08.04), five (line 01.01) and eight (line R1.01.02) copies for further study (Figure 4-16). I would use the 08.04 zero-copy line as a control. The five-copy  $T_0$  plant had much longer glumes than expected for the copy number (Figure 4-15), suggesting a high level of expression from the integrated constructs. The eight-copy  $T_0$  plant had the highest copy number of the independent  $T_0$  lines. The primary spikes of the five- and eight-copy lines had three rudimentary basal spikelets each, in contrast to the zero-copy line which had none; this is consistent with evidence that increased and ectopic expression of VRT-A2 causes an increase in the number of rudimentary basal spikelets<sup>43</sup>.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**



Figure 4-16: Primary spikes of  $T_0$  wheat plants (*Triticum aestivum* cv 'Fielder') containing (A) zero, (B) five and (C) eight copies of the VEGETATIVE TO REPRODUCTIVE TRANSITION- $A2b::FLAG$  construct. The primary spike was the first spike which completely emerged from the flag leaf. White arrows denote rudimentary basal spikelets.

To confirm the phenotypes I observed in the  $T_0$  plants, I phenotyped  $T_1$  plants from the three lines I selected (Figure 4-17). The five- and eight-copy  $VRT-A2b::FLAG$  lines had significantly longer spikes than the zero-copy line (one-way ANOVA with Tukey post-hoc test:  $p = 0.002$ , 95% CI [1.00, 3.92] and  $p = 0.001$ , 95% CI [1.16, 4.08], respectively) (Figure 4-17A) and significantly longer glumes than the zero-copy line (LMM:  $t(435) = -10.36$ ,  $p < 0.001$  and  $t(427) = -8.16$ ,  $p < 0.001$ , respectively) (Figure 4-17B). Dr Yunchuan Liu and I also quantified  $VRT-A2$  expression in glumes and rachis 14 dpa using RT-qPCR. The five- and eight-copy lines had significantly higher  $VRT-A2$  expression in glumes (one-way ANOVA with Tukey post-hoc test:  $p < 0.001$ , 95% CI [5624.69, 17200.35] and  $p = 0.03$ , 95% CI [805.58, 12381.24], respectively) (Figure 4-17D) and rachis (one-way ANOVA with Tukey post-hoc test:  $p < 0.001$ , 95% CI [3.12, 4.74] and  $p < 0.001$ , 95% CI [2.29, 3.82], respectively) (Figure 4-17C) with respect to wildtype.

The phenotypes I observed in the  $VRT-A2b::FLAG$   $T_0$  and  $T_1$  lines are consistent with those observations made in  $VRT-A2b$  transgenic lines in Adamski, *et al.*<sup>56</sup>. Therefore, I can conclude that the c-terminus FLAG tag does not significantly interfere with the function of the VRT-A2 protein, which results in the same phenotypes as previously observed in plants with the  $VRT-A2b$  allele.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

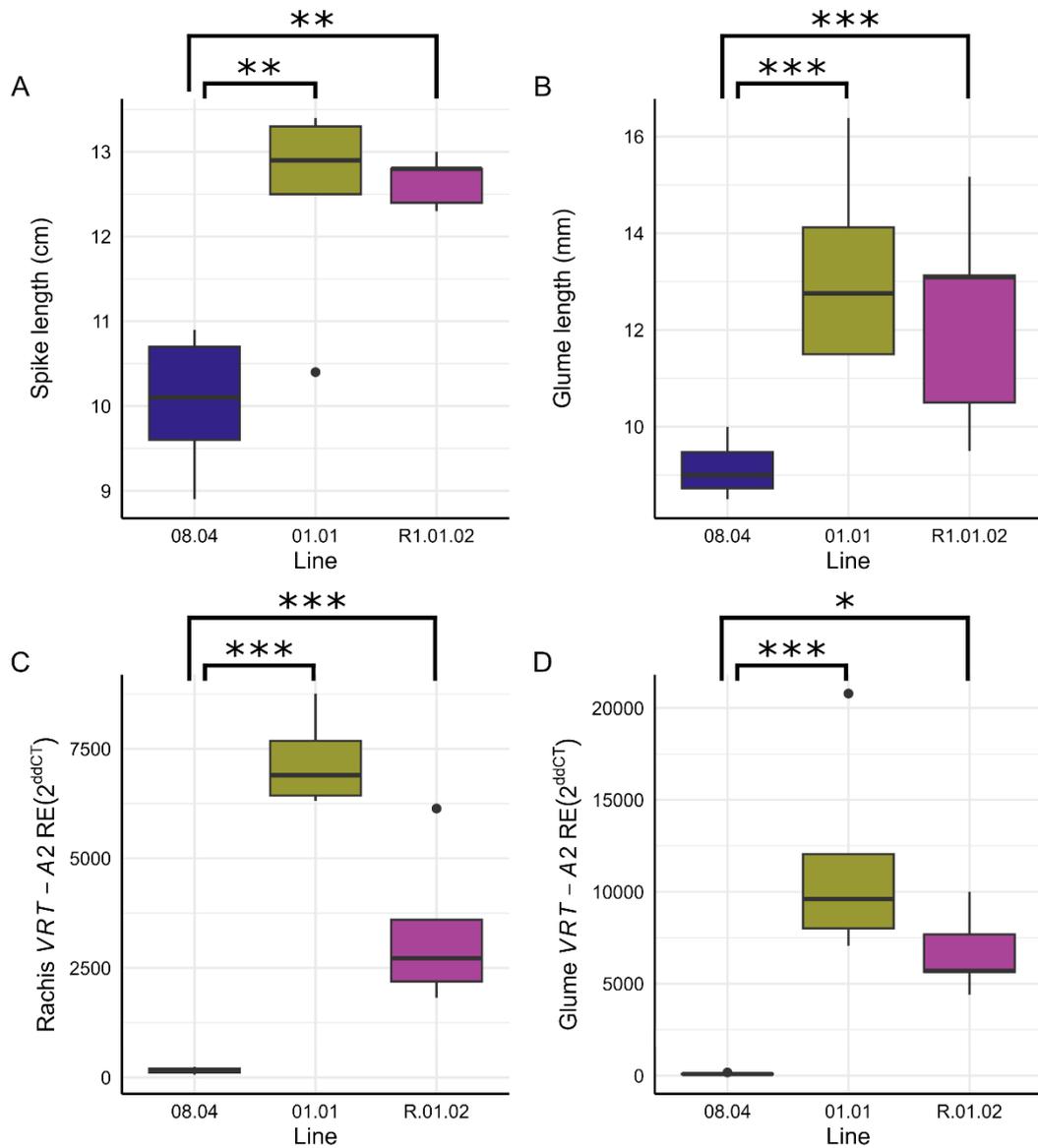


Figure 4-17: Phenotypes of  $T_1$  VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b::FLAG (VRT-A2b::FLAG) wheat plants (*T. aestivum* cv 'Fielder') at 14 days post anthesis (dpa). Lines consist of  $T_1$  plants from independent  $T_0$  individuals with zero (line 08.04), five (line 01.01) and eight (line R1.01.02) copies of the VRT-A2b::FLAG construct ( $n =$  five plants per line). Significance levels are indicated with asterisks: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . (A) Spike length in cm from the base of the first spikelet (including rudimentary basal spikelets) to the apex of the terminal spikelet (excluding awns). (B) Mean glume length in mm for all glumes excluding rudimentary basal spikelets. (C) VRT-A2 relative expression in the rachis and (D) glumes as quantified using qRT-PCR.

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (VRT-A2) act via microRNA172 to achieve its phenotypic effect?**

**4.4.6.2 VRT-A2::FLAG protein can be detected by an anti-FLAG antibody**

To test whether the FLAG tag could be detected by an anti-FLAG antibody, Dr Yunchuan Liu used a Western blot to detect VRT-A2::FLAG in whole protein samples from 08.04, 01.01 and R1.01.02 T<sub>1</sub> spikes (Figure 4-18). There was signal for all samples when a positive control anti-actin antibody was used (Figure 4-18A). There was a strong signal in lanes containing protein samples from either R1.01.02 or 01.01 spikes (Figure 4-18B). There was very faint signal in the lanes containing protein samples from either 08.04 spikes (the zero-copy control line) or wildtype Fielder leaf tissue.

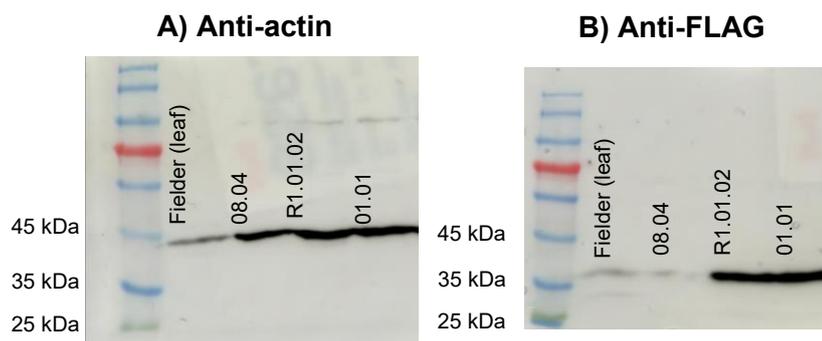


Figure 4-18: Western blot of W4.0 (terminal spikelet according to the Waddington scale<sup>41</sup>) spike samples (unless labelled as leaf tissue) of VEGETATIVE TO REPRODUCTIVE TRANSITION-A2b::FLAG T<sub>1</sub> transgenic wheat (*T. aestivum* cv 'Fielder') lines. A) detection using an anti-actin antibody, B) detection using an anti-FLAG antibody.

The Western blot shows that the VRT-A2::FLAG protein can be detected using an anti-FLAG antibody. The VRT-A2::FLAG protein runs at a slightly higher molecular weight than predicted at approximately 38 kDa; the predicted molecular weight is 28.3 kDa. However this is not an unexpected result as in a previous study, an SVP::HA protein also ran on a protein gel at approximately 38 kDa with a predicted molecular weight of 28.0 kDa<sup>285</sup>. Combined with the fact that the five- and eight-copy *VRT-A2b::FLAG* lines produce this signal when probed with an anti-FLAG antibody, but wildtype Fielder and the zero-copy line do not, I am confident that the *VRT-A2b::FLAG* transgenic lines are producing VRT-A2::FLAG protein which is detectable using an anti-FLAG antibody.

**4.4.6.3 A CUT&RUN assay is being developed to identify DNA that VRT-A2 protein binds to**

To identify DNA sequences that VRT-A2 binds to *in planta*, my aim was to use the *VRT-A2b* transgenic lines I have described in this section as input for a CUT&RUN assay. CUT&RUN is a relatively novel protocol first described in Skene and Henikoff<sup>286</sup>. The principles are similar to the better-known ChIP-Seq (chromatin immunoprecipitation sequencing), however the reaction is carried out in intact nuclei, which means that the input requirement is much

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (*VRT-A2*) act via *microRNA172* to achieve its phenotypic effect?**

lower; ChIP-Seq requires grams of input material per reaction while CUT&RUN requires as little as 10,000 nuclei. This means that the assay can be carried out on nuclei from single young wheat spikes. During the CUT&RUN protocol proteins are crosslinked to DNA they are bound to in the nuclei. By using an anti-FLAG antibody, *VRT-A2::FLAG* protein is pulled-down along with DNA it is bound to. This DNA is then sequenced using NGS.

Unfortunately, this experiment was not able to be completed before the preparation of this thesis. In future, this technique will enable us to identify DNA that the *VRT-A2* protein binds to *in planta*. We can then test whether *VRT-A2* binds to and regulates downstream genes such as *MIR172*, *AP2-2* and *AP2-5* directly.

## 4.5 Discussion

In this chapter I have investigated genetic interactions between *VRT2*, *AP2-2* and *AP2-5* to better understand how *VRT-A2b* confers its phenotypic effects. Using mutant crosses and expression analysis I have tested hypotheses based on known networks in Arabidopsis and begun to shed light on the connections between these key genes.

### 4.5.1 *VRT-A2* downregulates miR172

There is evidence for a direct *SVP*-miR172 interaction in Arabidopsis<sup>3,284</sup>, so I tested whether this interaction is conserved in wheat. First, I used the sRNA-Seq dataset I generated in Chapter 1 to look at differences in miR172 abundance between *VRT-A2a* and *VRT-A2b* NILs. The abundance of both miR172a and miR172b was lower in *VRT-A2b* spikes at two developmental stages, however this difference was not significant. The strength of any effect appeared to be equivalent between miR172a and miR172b. As the increase in *VRT-A2* expression that the *VRT-A2b* allele confers is relatively subtle, I further tested this hypothesis using germplasm with greater variation in *VRT-A2* expression.

I used RT-qPCR to quantify changes in miR172 abundance in spikes from wheat germplasm with different levels of *VRT2* expression and found that *vrt2* loss-of-function spikes had a significantly higher level of miR172 compared to wildtype spikes. There appeared to be a small effect in the opposite direction, *i.e.*, increasing *VRT2* expression led to a decrease in miR172 abundance, however these differences were not statistically significant. This is consistent with evidence from Arabidopsis where Cho, *et al.*<sup>284</sup> showed that decreasing *AtSVP* expression had a greater effect on miR172 abundance than increasing *AtSVP* expression<sup>284</sup>. These results suggest that *VRT-A2* negatively regulates miR172 (either directly or indirectly) in wheat. Unfortunately, the RT-qPCR assay is unlikely to be capable of distinguishing between miR172a and miR172b. Therefore, it would be useful to use sRNA-Seq, as I did to initially test for differences in miR172 abundance in the *PI<sup>WT</sup>* and *PI<sup>POL</sup>* NILs, to test for

**Chapter 4: Does VEGETATIVE TO REPRODUCTIVE TRANSITION-2 (*VRT-A2*) act via *miR172* to achieve its phenotypic effect?**

differences in miR172 abundance between *vrt2* and wildtype spikes. This would allow me to fully test whether there is any difference. Our data and a previous study show that decreasing *VRT-A2* expression has a greater effect on miR172 abundance<sup>284</sup>. Therefore, it would be important to quantify the effect of reducing *VRT-A2* expression on miR172a and miR172b abundance.

I have provided strong evidence that *VRT-A2* directly or indirectly downregulates miR172 in wheat. In Arabidopsis, *AtSVP* directly binds to and downregulates miR172a<sup>284</sup> and in future I would test whether this interaction is also direct in wheat using the *VRT-A2b::FLAG* transgenic lines we have developed and validated.

#### 4.5.2 Increased and extended *VRT-A2* expression leads to higher *AP2-5* expression

A *VRT-A2b* x *ap2-5A* cross revealed a genetic interaction between *VRT-A2* and *AP2-5* in respect to glume width, glume length, grain width and the number of bracts per spikelet. This provides evidence that there is a direct or indirect interaction between *VRT-A2* and *AP2-5*. Furthermore, in young *PI<sup>POL</sup>* NIL spikes with high *VRT-A2* expression, *AP2-5* expression was statistically significantly higher than in *PI<sup>WT</sup>* spikes. This suggests that, consistent with evidence from Arabidopsis<sup>3,264</sup>, increased and extended *VRT-A2* expression leads to higher *AP2-5* expression (either directly or indirectly). This result is somewhat contradictory to evidence from Liu, *et al.*<sup>108</sup>, who identified differentially expressed genes in tetraploid NILs carrying the *VRT-A2a* or *VRT-A2b* allele. They did not identify *AP2-5* as differentially expressed between the NILs, however the samples used were from 2.5-3.0 cm spikes, which are older than any of the timepoints we measured (up to 1.5 cm spikes). Also, the NILs were in a tetraploid background<sup>108</sup>, while our data is from a hexaploid background.

In Arabidopsis this interaction is direct as *AtSVP* protein has been shown to bind to the promoter of *AtRAP27*<sup>3</sup>, the *AP2-5* orthologue. I would in future test whether this interaction is direct or indirect using the *VRT-A2b::FLAG* transgenic lines. Higher *AP2-5* mRNA levels in *PI<sup>POL</sup>* spikes may simply be the result of lower miR172 levels. The fact that *VRT-A2* negatively regulates miR172 abundance and positively regulates *AP2-5* abundance is logically consistent. If *VRT-A2* expression increases, this will lead to a decrease in miR172 abundance which would release some repression of *AP2-5*. Of course, an indirect and direct interaction are not mutually exclusive; *VRT-A2* may regulate *AP2-5* expression directly and indirectly, via miR172.

### 4.5.3 Increased and extended *VRT-A2* expression leads to lower *AP2-2* expression

In *PI<sup>POL</sup>* spikes *AP2-2* expression was significantly lower at multiple timepoints compared to *PI<sup>WT</sup>* spikes suggesting that *VRT-A2* downregulates *AP2-2* (either directly or indirectly). This is consistent with data from Liu, *et al.* <sup>108</sup>, who found that in slightly older spikes from tetraploid NILs carrying the *VRT-A2b* allele, *AP2-2A* was significantly downregulated when compared to the wildtype *VRT-A2a* allele.

This result is consistent with the model I developed in the introduction to this chapter (Figure 4-2), suggesting that the long floral organs in *VRT-A2b* may be the outcome of a decrease in *AP2-2* expression. It is also consistent with the fact that *VRT-A2b ap2-5A* plants are a close phenocopy of *ap2-2 ap2-5A* plants with numerous empty glume-like bracts at each node instead of spikelets. This result suggests that increasing and extending *VRT-A2* expression has a similar effect on wheat spike development to downregulating *AP2-2*.

This result is, however, contradictory to evidence in Arabidopsis. The Arabidopsis *AP2-2* orthologue, *AtAP2*, was found to be bound and upregulated by AtSVP in a ChIP-chip experiment <sup>3</sup> and *AtAP2* expression is lower in *syp* plants <sup>264</sup>. The decrease in *AP2-2* expression in response to increased and extended *VRT-A2* expression observed in both this and previous studies in wheat <sup>108</sup> is difficult to reconcile with the negative regulation of *miR172* by *VRT2*.

In my general discussion (Chapter 5), I will propose two alternative hypotheses which could explain this unexpected observation. I will also recommend experiments which could be used to test these hypotheses.

### 4.5.4 The tagged transgenics will reveal direct targets of the *VRT-A2* protein

To confirm the interactions which have been suggested by the results in this chapter, I developed transgenic lines containing a FLAG-tagged version of *VRT-A2b* and confirmed that these lines have higher *VRT-A2* expression, have *VRT-A2::FLAG* protein detectable with a FLAG antibody, and phenotypes consistent with the *VRT-A2b* allele <sup>56</sup>. In the future, we will use spikes from these lines for CUT&RUN assays to identify direct targets of the *VRT-A2* transcription factor. Potential targets would include cis-regulatory regions of *AP2-2*, *AP2-5*, and *MIR172* genes.

#### 4.5.5 Conclusions and future perspectives

In this chapter I have begun to test whether the downstream network of *AtSVP* in *Arabidopsis* is conserved in wheat to attempt to elucidate the mechanism by which *VRT-A2b* confers its phenotypic effect. In the experiments I have described here I have shown that some interactions in *Arabidopsis* are also present in wheat, for instance increased *VRT-A2* expression results in higher *AP2-5* expression and lower *VRT2* expression leads to a decrease in miR172 abundance. However, the result that increased *VRT-A2* expression results in lower *AP2-2* expression is inconsistent with studies in *Arabidopsis* and with the effect on miR172, although they are consistent with independent RNA-seq data. In Chapter 5, I will bring the data presented here together with data from Chapters 2 and 3 to propose mechanisms which could explain this unexpected observation.

To fully characterise this genetic network, I would also wish to disentangle the effect of *VRT-A2* on individual miR172 family members and use the *VRT-A2b::FLAG* spikes to identify direct DNA targets of the VRT-A2 protein. By doing the above, and expanding our understanding of the *VRT-A2* downstream network, we can implement targeted mutations to modify *VRT2* expression patterns in a way that is most beneficial for plant breeding, as suggested in the paper which first cloned *VRT-A2* <sup>56</sup>.

## 5 General Discussion

### 5.1 Thesis summary

The overall aim of this thesis was to develop our understanding of how miRNAs affect wheat spike development. I aimed to achieve this in a broad sense, by expanding our knowledge of the miRNAs that are expressed, when they are expressed, and the genes they target. I also tackled this aim by interrogating the interactions between miR172 and *AP2-2*, *AP2-5* and *VRT2* which are all key players during early wheat spike development. I aimed to develop a nuanced understanding of these connections. Specifically, I have used spatial and non-spatial transcriptomics, transgenics, phenotypic characterisation, and luciferase-based assays to answer the following questions:

4. What is the role of miRNAs during key developmental transitions in the wheat inflorescence?
5. How do the expression patterns of *AP2-2* and *AP2-5* differ in developing spikes when they are both regulated by miR172?
6. Does *VRT-A2* act via miR172 to achieve its phenotypic effect?

### 5.2 I have developed novel resources for the wheat community

I first sought to develop a coherent and comprehensive list of miRNAs which are present in early developing wheat spikes and the *MIRNA* loci which produce them (Chapter 2). I generated a high-quality, high-depth sRNA-Seq dataset from young wheat spikes and used three independent pipelines to predict candidate miRNAs from the data. I used TargetFinder to identify putative target transcripts of these miRNAs. To gain maximum benefit from the dataset, I looked for variation in miRNA binding sites in the A. E. Watkins landrace collection<sup>34</sup> and EMS-mutagenised TILLING populations<sup>31</sup>. I identified TILLING lines with variation in the miRNA binding sites of genes we know play vital roles in wheat spike development; these lines are currently being grown for further characterisation. I have identified 187 promising SNPs in the Watkins collection which correlate with phenotypic differences between lines.

This work opens new avenues of research and opportunities for introducing novel variation into wheat breeding pipelines. This dataset also provided key insight into the expression of miR172, a miRNA family I focused on in Chapters 3 and 4 of this thesis.

### 5.3 I have developed a more nuanced understanding of miR172-mediated repression of AP2-like genes

To begin to understand in more detail the role individual miRNAs play in developing wheat spike, I took advantage of recently generated genetic resources, including the dataset described in Chapter 2, to unpick a known interaction between miR172 and AP2-like genes<sup>32,42,66</sup>. Until now, studies into miRNA function in wheat have treated miRNA families as single molecules or taken a single family member as a proxy for the entire family<sup>32,95</sup>. I took a more time-consuming but nuanced approach to understand how individual miR172 family members regulate the AP2-like genes AP2-2 and AP2-5 (Chapter 3).

The SNP which differentiates the miR172a and miR172b mature sequences is at the far 5' end of the miRNA. Therefore, using stem-loop RT-qPCR is unlikely to reliably distinguish between miR172a and miR172b<sup>165</sup>. To understand the expression profile of these family members, which are processed from distinct *MIR172* loci with potentially different transcription rates and tissue-specificity, I used sRNA-Seq. By using this approach, I was able to discover that miR172a and miR172b in fact have unique expression profiles during wheat spike development.

I compared the capacity of miR172a and miR172b to reduce the abundance of a luciferase protein containing binding sites from AP2-2 and AP2-5. This assay showed that miR172a and miR172b repress AP2-2 and AP2-5 binding sites equally. This assay is, however, limited by the fact that the binding sites are not within their endogenous mRNA context, which has been shown to be important for miRNA binding and miRNA-mediated cleavage<sup>87,244</sup>. Hence, to test whether miR172 binding is affected by the mRNA context, I have developed constructs containing an AP2-5 gene with either the endogenous AP2-5 miR172 binding site, or the AP2-2 miR172 binding site. I have taken advantage of recent improvements in the efficacy of wheat transformation to start the process of transforming these constructs into *ap2-5A* loss-of-function mutants. Complementing a mutant line with constructs with different miRNA binding sites is an ambitious experimental design; few studies in wheat have used transgenic constructs to complement a mutant, such as in Hyde, *et al.*<sup>287</sup>. The phenotypes of these plants will provide insight into the repression of different AP2-like genes *in planta*.

Finally, I was also able to take advantage of MERFISH technology to compare the spatial expression range of the miR172 targets AP2-2 and AP2-5. I found that AP2-2 and AP2-5 transcripts partially overlap, however there are a significant number of cells in the wheat spike where AP2-2 transcripts are present while AP2-5 transcripts are absent, or vice versa. This

provides the opportunity for miR172 to regulate *AP2-2* and *AP2-5* in a cell-type specific manner due to different amounts of overlap between miR172 and *AP2*-like transcripts.

In Arabidopsis<sup>102,163,164</sup>, rice<sup>185</sup>, and tomato<sup>186</sup>, the functions of different members of the same miRNA family have been studied separately. This is the first study in wheat I am aware of which disentangles the distinct functions of miRNA family members. By doing so, I was able to generate evidence against my hypothesis that based on sequence complementarity, different miR172 family members repress different *AP2*-like genes. Modern resources and protocols allow us to look at miRNA function to a high level of detail and this will be critical to fully understanding the nuance of these key interactions and leveraging them for crop improvement.

## 5.4 *VRT-A2* promotes *AP2-5* abundance and represses miR172

In Chapter 4, I expanded the *AP2-2*, *AP2-5*, miR172 network to include *VRT-A2*. I used mutant crosses to attempt to unravel the role of each of these four players and how they interact with one another.

### 5.4.1 *VRT-A2* represses miR172 abundance

Using the sRNA-Seq data I described in Chapter 2, I found that *PI<sup>POL</sup>* spikes had lower levels of miR172 (although the difference was not significant). The effect appeared to be similar between miR172 family members.

I showed using RT-qPCR that lower *VRT-A2* abundance results in higher miR172 abundance in W5.0 spikes, while higher levels of *VRT-A2* led to a non-significant decrease in miR172 abundance. This result was consistent with evidence from Arabidopsis<sup>284</sup>. While RT-qPCR is not always family member-specific, I was able to use this method given as I had previously established using sRNA-Seq that the effect of increasing and extending *VRT-A2* expression was similar on miR172a and miR172b.

### 5.4.2 *VRT-A2* promotes *AP2-5* abundance

I described the evidence of an interaction between *VRT-A2* and *AP2-5* in Chapter 4. In brief, a cross between a *VRT-A2b* NIL and an *ap2-5A* loss-of-function mutant showed that *VRT-A2* and *AP2-5A* interact (either directly or indirectly). I found that in RNA-Seq data, *VRT-A2b* NILs have higher *AP2-5* expression compared to *VRT-A2a* NILs. This is logically consistent with the fact that lower *VRT-A2* expression leads to an increase in miR172; if *VRT-A2* downregulates miR172, this would lead to an increase in the abundance of miR172 targets such as *AP2-5*.

## 5.5 Patterns of *AP2-2* mRNA abundance cannot be explained using current data

### 5.5.1 *AP2-2* mRNA abundance increases through developmental time

As I discussed in Chapter 3, *AP2-5* and *AP2-2* have opposite expression profiles during early wheat spike development. *AP2-5* expression decreases while *AP2-2* expression increases. The sRNA-Seq dataset I generated showed that miR172 abundance increases, which is a pattern observed across grass species<sup>288,289</sup>. Therefore, it is logically consistent that abundance of the *AP2-5* mRNA targeted by miR172 decreases during the same period. However, *AP2-2* mRNA abundance increases during this time. This expression profile was observed in two independent RNA-Seq datasets (cv Paragon and Chinese Spring). We know from Debernardi, *et al.*<sup>42</sup> and the dual luciferase assay in Chapter 3 of this thesis that the abundance of a protein with an *AP2-2* binding site are lower in the presence of miR172. The dual luciferase assay also showed that *AP2-2* and *AP2-5* binding sites are targeted with the same efficiency by both miR172 family members, so the increase in *AP2-2* mRNA abundance cannot be attributed to the decrease in miR172b abundance seen from W4.0 to W5.0.

It is important to note that this data shows only that *AP2-2* mRNA levels increase from W2.5 to W5.0. We do not currently have data that would enable us to see whether *AP2-2* protein levels also increase during this time. The dual luciferase assay, on the other hand, only showed that a protein with an *AP2-2* binding site decreases in the presence of miR172. Any change in the abundance of the corresponding mRNA was not quantified.

### 5.5.2 Increased and extended *VRT-A2* expression leads to lower *AP2-2* mRNA abundance

Given the evidence that *VRT-A2* represses miR172, we would expect *AP2*-like gene expression to be higher in *VRT-A2b* NILs (which have increased and extended *VRT-A2* expression). We see this pattern in *AP2-5* as described in Section 5.4.2. The fact that *VRT-A2* negatively regulates miR172 abundance and positively regulates *AP2-5* abundance is logically consistent. We would also expect to observe this pattern for *AP2-2*, but yet again *AP2-2* behaves unexpectedly. I showed in Chapter 4 that in *PI<sup>POL</sup>* NIL spikes (increased and extended *VRT-A2* expression), *AP2-2* is higher than in *PI<sup>WT</sup>* spikes. This finding is consistent with a previous study in wheat<sup>108</sup>. As before, this pattern cannot be explained by *AP2-2* not being targeted by miR172 (Section 5.5.1).

## 5.6 Generating hypotheses to explain the unexpected patterns of *AP2-2* mRNA abundance

Above I have laid out ways in which observed patterns of *AP2-2* mRNA abundance are not logically consistent with our gene networks. Therefore, there must be an element missing from the model. Below, I have generated two hypotheses capable of explaining these unexpected observations. I will also describe ways in which these hypotheses could be tested in the future.

### 5.6.1 Variation in the dominant mechanism of miR172-mediated repression (miR172 mechanism hypothesis)

There are two mechanisms by which miRNAs repress genes. The first is DICER-mediated mRNA cleavage, which reduces the abundance of the target mRNA and so the abundance of the target protein. The second is translational inhibition, where the miRNA inhibits translation of the target mRNA into protein. The effects of mRNA cleavage can be observed in RNA-Seq and spatial transcriptomics data, while the effects of translational inhibition cannot as it does not affect mRNA, only protein abundance. My first hypothesis is that there are differences in the mechanisms by which miR172 represses *AP2-2* and *AP2-5*, referred to from here as the miR172 mechanism hypothesis.

We know from previous literature that *AP2-5* is at least partially repressed via mRNA cleavage in wheat<sup>32</sup>. However, we do not know the mechanism by which miR172 represses *AP2-2*. As hypothesised by Houston, *et al.*<sup>240</sup>, the mechanism of miR172-mediated repression may change according to target (Figure 5-1A), tissue (Figure 5-1B), or developmental stage (Figure 5-1C).

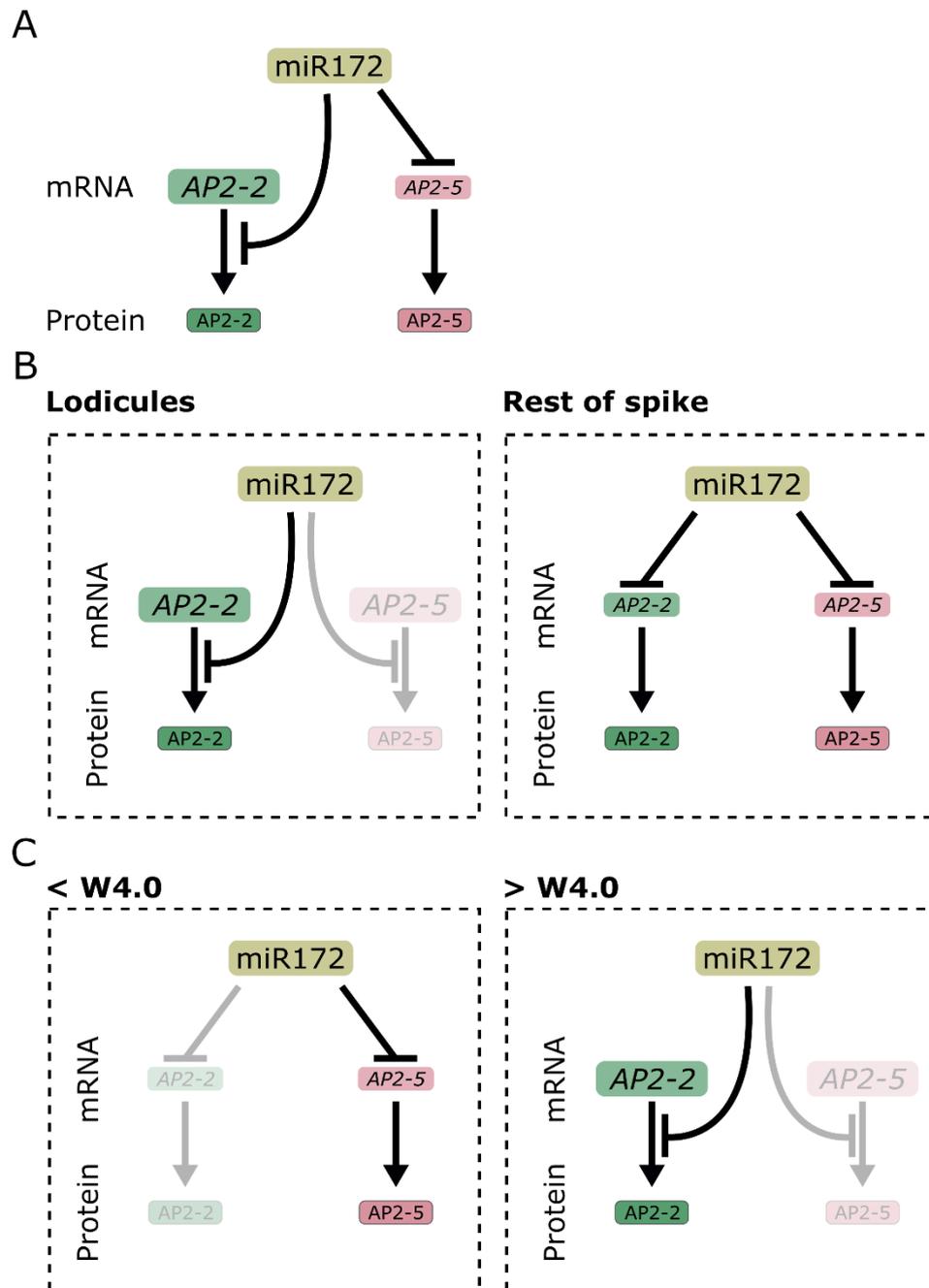


Figure 5-1: Models of the microRNA172 (miR172) mechanism hypothesis to explain unexpected observations. Positive interactions are represented with arrows, while inhibitory interactions are represented with blunt arrows. (A) The mechanism of miR172-mediated repression is target-dependent. miR172 represses APETALA2-5 (AP2-5) via mRNA cleavage and APETALA2-2 (AP2-2) via translational inhibition. (B) The mechanism of miR172-mediated repression is tissue-dependent. miR172 repression is via translational inhibition in lodicules (where AP2-2 expression is dominant), while miR172 repression is via mRNA cleavage in other tissues (where AP2-5 expression is dominant, but AP2-2 is co-expressed). (C) The mechanism of miR172-mediated repression is dependent on the developmental stage. miR172-mediated repression is via mRNA cleavage during very early spike development

before W4.0 (terminal spikelet according to the Waddington scale<sup>41</sup>, when AP2-2 is very lowly expressed), and via translational inhibition during later spike development (after W4.0, when AP2-5 expression is lower).

There is precedent for closely related miR172 targets being repressed by miR172 via different mechanisms in other species. In Arabidopsis, *AtAP2* (the AP2-2 orthologue) and *AtRAP27* (the AP2-5 orthologue) are regulated by miR172-mediated translational inhibition, modifying miR172 abundance results in changes to *AtAP2* and *AtRAP27* protein but not mRNA levels<sup>61,181</sup>. However, this has been shown to not be completely conserved in grasses. For instance, in maize, the two orthologues of *AtRAP27* are regulated by miR172 using different mechanisms<sup>290</sup>. *ZmIDS1* is repressed via translational inhibition<sup>291</sup> while *ZmSID1* is at least partially repressed via mRNA cleavage<sup>290</sup>. This shows that very closely related genes (both *ZmIDS1* and *ZmSID1* are orthologues of AP2-5) can be regulated via different mechanisms<sup>292</sup>. Although there may be subtle differences, broadly *ZmIDS1* and *ZmSID1* are expressed in the same tissues at the same developmental stage (spikelet meristems during early tassel development)<sup>290,291</sup>, showing that the difference in the mechanism of miR172-mediated repression in this case is likely to be target-specific. Our model may be further complicated by the fact that *AtAP2* has been shown to form a negative feedback loop in Arabidopsis; AtAP2 protein represses *AtAP2* transcription<sup>143</sup>.

The miR172 mechanism hypothesis predicts that AP2-5 is predominantly regulated by miR172 via mRNA cleavage, while AP2-2 is mostly regulated by translational repression. This may be due to differences in the mechanism of repression according to target, timepoint, or tissue, however the result would be that AP2-2 is regulated by translational repression. This would mean that the increase in AP2-2 mRNA abundance over developmental time and in *PI<sup>POL</sup>* NILs compared to *PI<sup>WT</sup>* do not account for miR172-mediated repression and are not reflective of protein abundance.

To establish the mechanism of miR172-mediated repression of AP2-2, I would quantify AP2-2 and AP2-5 protein in developing wheat spikes with low (MIM172<sup>32</sup>), wildtype and high (Ubi::miR172<sup>32</sup>) levels of miR172. This would capture the effect of any repression by miR172, irrespective of the mechanism. By taking mRNA levels into account, this experiment may also shed some light on whether the repressive mechanism changes over time. I would use a dual luciferase assay to compare the abundance of AP2-2 mRNA and protein in the presence and absence of miR172. This would indicate whether the mechanism of miR172-mediated repression is target-dependent. This experiment would be in a heterologous system, and we know that the mechanism by which a particular miRNA represses its targets can vary between plant species<sup>61,181,290,291</sup>, therefore any conclusions would be preliminary. One would

need to quantify mRNA and protein levels in the presence and absence of a miRNA *in planta* to conclusively determine the dominant mechanism of miRNA-mediated repression. *AP2-2* and *AP2-5* mRNA levels cannot be reliably quantified using the spatially resolved MERFISH technology described in Chapter 3. Single cell RNA-Seq or RNA-Seq of laser-dissected cells could be used in the future to quantify changes in *AP2-2* and *AP2-5* transcript levels in the presence of low, wildtype, and high levels of miR172. Unfortunately, current technology does not allow spatial protein quantification which would enable us to fully test whether the mechanism of miR172-mediated repression is tissue-dependent. The pace of progress, including the use of ONT for directly sequencing proteins<sup>293</sup>, make this a possibility in the near future.

#### 5.6.1.1 *The miRNA mechanism hypothesis relies on independent changes to the rate of AP2 transcription over time*

As *AP2-2* mRNA abundance increases over time, the rate of *AP2-2* transcription must also increase. I could use smFISH to quantify nascent and mature *AP2-2* mRNA transcripts over developmental time and in *PI<sup>POL</sup>* and *PI<sup>WT</sup>* NIL spikes to quantify the change in transcription rates with spatial resolution. To investigate how this is achieved, *VRT-A2b::FLAG* transgenics could be used to test whether VRT-A2 protein binds directly to *AP2-2*, lending weight to a direct interaction as well as one mediated by miR172. This negative interaction could also be via another pathway or with a co-regulator.

#### 5.6.1.2 *The miR172 mechanism hypothesis relies on changes to the rate of AP2 transcription in the presence of increased and extended VRT-A2 expression*

If the miR172 mechanism hypothesis is correct, the rate of *AP2-2* transcription must also decrease in the presence of increased and extended *VRT-A2* expression. This is supported by my observation that *VRT-A2b/ap2-5A* spikes are a phenocopy of *ap2-2 ap2-5A* spikes with numerous glume-like bracts per spikelet and double carpels. I used this observation to hypothesise that the increased and ectopic *VRT-A2* expression conferred by the *VRT-A2b* allele results in lower *AP2-2* expression.

To uncover this additional element of the genetic network, I would use the spatial transcriptomics data described in Chapter 3 to identify genes which are differentially expressed in *VRT-A2a* and *VRT-A2b* NILs. I would use this data to identify candidate genes which *VRT-A2* may act via to repress *AP2-2*. I would identify or generate (using CRISPR-Cas9) mutants of these candidate genes and quantify expression of *AP2-2* to narrow the list of candidates. I would test any promising candidates using yeast-two-hybrid or yeast-one-hybrid techniques to see if the candidates bind to *AP2-2* DNA or *AP2-2* protein. I would use also make use of existing yeast-two-hybrid data published by Liu, *et al.*<sup>108</sup> to identify any proteins

that VRT-A2 may interact with to downregulate *AP2-2*. I could again use the spatial transcriptomics data to exclude candidates which are expressed in the lodicules.

### 5.6.1.3 Changes to the rate of *AP2* transcription in the presence of increased and extended

*VRT-A2* expression may also be the result of *AP2* protein repressing *AP2-2* transcription

The decrease in *AP2-2* mRNA levels observed in *PI<sup>POL</sup>* NIL spikes may be the result of *AP2* protein negatively regulating *AP2-2* transcription, as in Arabidopsis<sup>143</sup>. In *PI<sup>POL</sup>* spikes, miR172 abundance is lower. This would lead to a higher level of *AP2* protein in the system, which would in turn repress *AP2-2* transcription, leading to a decrease in *AP2-2* mRNA levels which we observed in the RNA-Seq data described in Chapters 3 and 4. This hypothesis is only logically consistent if the main mechanism of miR172-mediated repression of *AP2-2* is translational inhibition.

## 5.6.2 miR172 is absent from lodicules (miR172 absent from lodicules hypothesis)

I showed using MERFISH technology that *AP2-2* and *AP2-5* transcripts overlap spatially in most tissue types in developing wheat spikes, however in cell cluster 17 (lodicules) *AP2-2* was expressed in the absence of *AP2-5* in 32.0% of cells. If miR172 is not expressed in the lodicules, this pattern would allow *AP2-2* expression to increase in the lodicules while *AP2-5* and *AP2-2* transcripts in other tissues would be repressed by miR172, producing these opposite expression profiles. This hypothesis is referred to from here as the miR172 absent from lodicules hypothesis.

To test the miR172 absent from lodicules hypothesis, I would spatially resolve miR172 expression in the wheat spike and specifically look for expression in the lodicules. LNA probes have been used in rice for miR172 *in situ* hybridisation<sup>183</sup>. A more appropriate, although less well-established technique, is sRNA-FISH which uses fluorescent labelling which allows for greater spatial resolution<sup>294</sup>. The LNA probes used in both methods are highly specific and may be capable of distinguishing between miR172a and miR172b.

This putative specificity of miR172 abundance may be the result of differences in the transcriptional control of *MIR172* loci between lodicules and other tissues. *MIR172* transcription may be regulated by a factor which is specifically present or absent from lodicules, or the epigenetic landscape surrounding the *MIR172* loci may differ between lodicules and other tissues. If a low level of miR172 was observed in lodicules compared to other tissues, an interesting future avenue of research would be uncovering the genetic and epigenetic factors which control this high level of spatial specificity of miR172 expression.

If the miR172 absent from lodicules hypothesis is true, there would be no miR172-mediated increase in *AP2-2* levels in P1<sup>POL</sup> lodicules. As lower miR172 abundance in the rest of the spike would result in higher *AP2-2* expression, there must again be some other regulatory interaction present in the wheat spike which we are not currently aware of, as described in Section 5.6.1.2. *VRT-A2* may downregulate *AP2-2* via another pathway. This repression would need to offset the increase in *AP2-2* abundance conferred by an increase in miR172. I believe this hypothesis is unlikely, due to the relatively small area of the lodicules compared to the rest of the spike; a lack of miR172 regulation here would be unlikely to result in the increase we observe in *AP2-2* expression across the spike. Also, plants carrying a miR172-resistant allele of *AP2-2* had smaller lodicules compared to the wildtype<sup>42</sup>. On balance, I believe that the miR172 mechanism hypothesis is likely to better reflect the biological processes behind the unexpected changes observed in *AP2-2* mRNA abundance.

## 5.7 Generating models to summarise how *AP2-2*, *AP2-5*, *VRT-A2* and miR172 work together to control phytomer differentiation

The double mutant phenotypes of *VRT-A2b ap2-5A* and *ap2-2 ap2-5A* are striking, with multiple empty glume-like bracts at almost every node on the rachilla. I have created models describing how *VRT-A2*, miR172, *AP2-2* and *AP2-5* could form a wildtype wheat spike using these mutant lines and our new knowledge of how these genes work together (Chapter 4).

As described in Section 1.3, the wheat spike is a highly branched structure which consists of repeating phytomers. When the spike begins to develop it elongates along the vertical axis and generates alternating phytomers. In brief, each phytomer consists of a lower vegetative meristem, which can differentiate into vegetative structures (such as leaves), and an upper axillary meristem which can form lateral branches known as spikelets. This developmental programme is repeated along the rachillas of each spikelet. The vegetative meristem of each spikelet phytomer can form glumes (where the axillary meristem is repressed) or lemma (where de-repressed axillary meristems which differentiate into florets).

### 5.7.1 Previous studies have hypothesised that *VRT-A2* expression affects the balance of leaf and axillary meristem repression

Previous studies have proposed that the downregulation of *VRT-A2* at W2.5 (double ridge stage) leads to de-repression of the axillary meristem<sup>43</sup>. The proposed model is that, as a vegetative signal, *VRT-A2* represses *SEPALLATA* genes, which act as floral growth signals and promote the development of the axillary meristem into a spikelet<sup>43</sup>. Higher *VRT-A2* levels

in spikes with the *VRT-A2b* allele may lead to continued repression of spikelet growth, leading to the formation of rudimentary basal spikelets<sup>43</sup>.

A study on the orthologue of *AP2-5* in barley showed that a reduction in *HvAP2L-H5* expression produced a terminal spikelet (usually absent in the indeterminate barley spike) and multiple florets per spikelet (where there are none in the wildtype)<sup>40</sup>. The authors hypothesised that *AP2-5* promotes spike indeterminacy and spikelet determinacy<sup>40</sup>. This is consistent with observations made in loss-of-function *ap2-5A* spikes, which display a sham ramification phenotype with a very high number of florets per spikelet (*i.e.*, a lower level of spikelet meristem determinacy)<sup>32</sup>.

### 5.7.2 *AP2*-like genes promote a lemma-like identity, which correlates with differentiation of the axillary meristem into a floret

Knocking out *AP2-2* and *AP2-5* generates a spike with numerous glume-like organs in each spikelet which do not contain a floret, *i.e.*, the axillary meristem is repressed. Increasing *AP2-2* or *AP2-5* expression through mutations in the miR172 binding site generates a spike with lemma-like glumes<sup>42</sup> and increased *AP2-5* expression has been shown to lead to florets forming at the first phytomer of apical spikelets<sup>32</sup>. These results suggest that *AP2*-like genes promote a lemma-like identity, and that a lemma-like identity is correlated with differentiation of the axillary meristem into florets (Figure 5-2).

Combining the *VRT-A2b* allele with a loss-of-function *ap2-5A* mutant also generates this striking phenotype of numerous empty bracts. The bracts, as for *ap2-2 ap2-5A*, appear to have a glume-like identity with very short awns. This phenotype may be conferred by *VRT-A2b* negatively regulating *AP2-2* expression, promoting a glume-like identity. *VRT-A2b* also positively regulates *AP2-5* (likely via miR172), which would promote a lemma-like identity. The absence of full length, functional *AP2-5A* transcripts in *VRT-A2b/ap2-5A* plants means that any promotion of a lemma-like identity via an increase in *AP2-5* expression is absent in this double mutant.

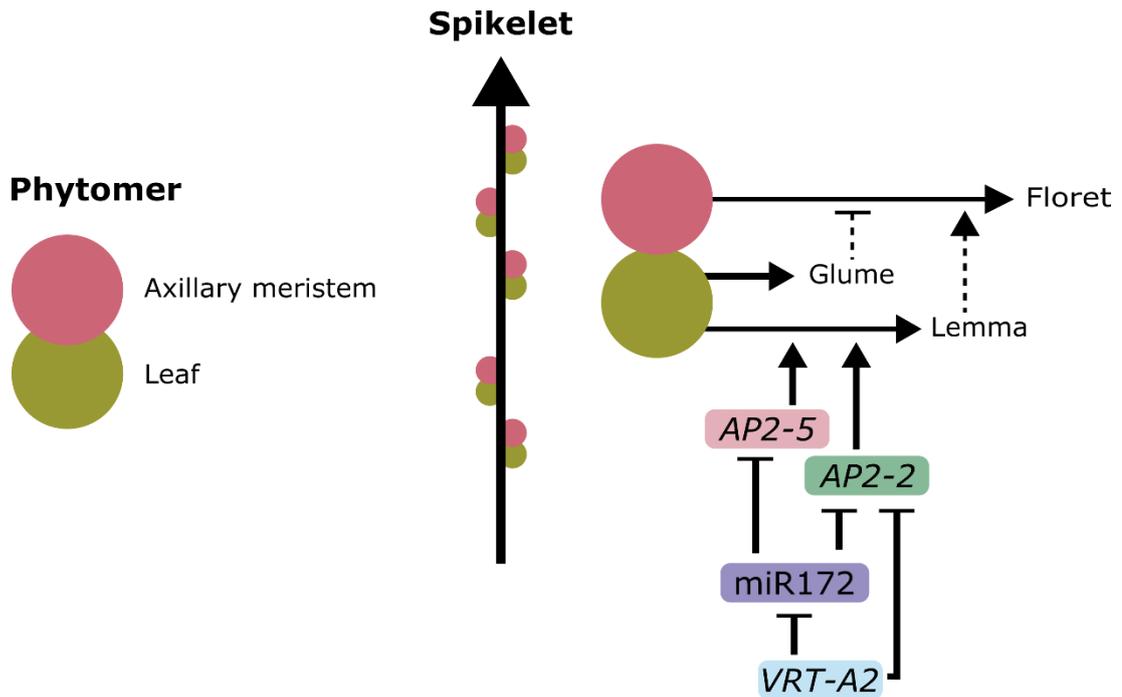


Figure 5-2: A model summarising the putative roles of *APETALA2*-like (*AP2*-like) genes and *VEGETATIVE TO REPRODUCTIVE TRANSITION-A2* (*VRT-A2*) during wheat spikelet phytomer differentiation. Positive interactions are represented with arrows, while inhibitory interactions are represented with blunt arrows. Phytomers are represented using pink and green circles for axillary and leaf meristems respectively. The putative genetic interactions between *VRT-A2*, *AP2-5*, *AP2-2*, and microRNA172 (*miR172*) affecting the development of phytomers within a wheat spikelet are shown.

## 5.8 Concluding statement

In summary, in this thesis I have begun to unravel the key role that miRNAs play during wheat spike development. The miRNA candidates and target transcripts identified here will provide a useful resource for the wheat community, particularly as fundamental researchers and wheat breeders alike can take advantage of the often dominant variation to be found in miRNA binding sites. This step can help us overcome the issue of functional redundancy in this hexaploid crop. By leveraging this dataset, as I have done in Chapters 3 and 4, I show that we can cultivate a highly nuanced understanding of miRNA interactions at the family member level. The pace of change in wheat research means that we have been able to test complex hypotheses directly in wheat by using approaches such as spatial transcriptomics, complementing loss-of-function lines with unique transgenic constructs, and tagging proteins of interest. This presents a step-change for developmental biology in wheat; we can be more ambitious than ever in our quest to understand how a structure as critical to human society as the wheat spike is formed.

# Appendix

## A.1 Primer list

Table A-1: List of primers used in this thesis

Primer name	Primer sequence (5' → 3')	Purpose
AP2-5A CDS L0	TTCCCTACCGCCGCCGCC	Site directed
MIR172 F		mutagenesis
AP2-5A CDS L0	TCGTGATGATGCTGCAGCGTAA	Site directed
MIR172 R	AGC	mutagenesis
M13R	CAGGAAACAGCTATGACCATG	Sequencing
M13F(-21)	TGTAAAACGACGGCCAGT	Sequencing
AP2-5A 3'UTR L0 F	GATGGTCTCACTGAAGCTGGCC	Site directed
	GTTG	mutagenesis
AP2-5A 3'UTR L0 R	CCAATACGCGTCAATC	Site directed
		mutagenesis
K3946_AP2-5A_FAM	GAAGGTGACCAAGTTCATGCT GCACGAACTCCTCCTTGGTT	PACE genotyping
K3946_AP2-5A_HEX	GAAGGTCTGGAGTCAACGGATT GCACGAACTCCTCCTTGGTC	PACE genotyping
K3946_AP2-5A_COM	TTGTCCGATGGTTGATATCTG	PACE genotyping
K2726_AP2-5A_FAM	GAAGGTGACCAAGTTCATGCTT GGGAGTCGCACATCTGA	PACE genotyping
K2726_AP2-5A_HEX	GAAGGTCTGGAGTCAACGGATT TGGGAGTCGCACATCTGG	PACE genotyping
K2726_AP2-5A_COM	GCGATTGATTAGTGATTTGGGT	PACE genotyping
K2992_AP2-5A_FAM	GAAGGTGACCAAGTTCATGCTT AGACCTGCTTCCCGCAATCT	PACE genotyping
K2992_AP2-5A_HEX	GAAGGTCTGGAGTCAACGGATT TAGACCTGCTTCCCGCAATCC	PACE genotyping
K2992_AP2-5A_COM	AATCGCTGGCCTTTGTCCGTCT	PACE genotyping
	G	
VRT-A2B_FAM	GAAGGTGACCAAGTTCATGCTT CCTACTCCAGATCCTACTCCAG	PACE genotyping

VRT-A2B_VIC	GAAGGTCGGAGTCAACGGATT ATGCCTGCGTGCAAAGAGCTC	PACE genotyping
VRT-A2B_COM	TGTAACCCAATAGCAACAACCT AG	PACE genotyping
K2233_AP2-2A_FAM	GAAGGTGACCAAGTTCATGCTC CAGTTCCTCGGCAAGAAG	PACE genotyping
K2233_AP2-2A_HEX	GAAGGTCGGAGTCAACGGATT CCAGTTCCTCGGCAAGAAA	PACE genotyping
K2233_AP2-2A_COM	AGAAGGGAGCTAGGAGGAGAG	PACE genotyping
K3634_AP2-2B_FAM	GAAGGTCGGAGTCAACGGATT AGACCTGTGTGCAACGCAG	PACE genotyping
K3634_AP2-2B_HEX	GAAGGTGACCAAGTTCATGCT AGACCTGTGTGCAACGCAA	PACE genotyping
K3634_AP2-2B_COM	CGAAGCTAAAGCTATGCGGTG	PACE genotyping
VRT-A2_F qPCR	CCGGCAATTCATGCAACAAATT	RT-qPCR
VRT-A2_R qPCR	GAACCGTCATCATTGTCCTGT	RT-qPCR
Actin-Fwd	ACCTTCAGTTGCCAGCAAT	RT-qPCR
Actin-Rev	CAGAGTCGAGCACAATACCAG TTG	RT-qPCR
Debernardi_Uni_MIRs	TGGTGCAGGGTCCGAGGTATT	Stem-loop RT-qPCR
Debernardi_RTmiR172a	GGCGGAGAATCTTGATGATG	Stem-loop RT-qPCR
Debernardi_SLOmiR17 2	GTCTCCTCTGGTGCAGGGTCCG AGGTATTCGCACCAGAGGAGA CATGCAG	Stem-loop RT-qPCR
Debernardi_snoR101_F	GATGTCTTACACTTGATCTCTG AACTT	Stem-loop RT-qPCR
Debernardi_snoR101_R	TGCATCAGGATTGATATAGTGT CC	Stem-loop RT-qPCR

---

## A.2 miR172 precursor sequences used in dual luciferase assay

Mature microRNA sequences highlighted in pink.

### miR172a

>6A dna:chromosome chromosome:IWGSC:6A:607452271:607452782:-1  
CTCGAGTCCCCTTTATTAATTTCTCTCCATCTACCTCTCTTTGTCTTCATCCATTGCATGAG  
AGACTAGTAGTATATTTTTTCTATTTTTCTTCTTCTTCCGCTTCTCTTTGTTCTTCTC  
CACGTACCACAATAATCTCGCCATGGAATACATGGGAGGAATTTTCATGGCCGGAGCAGAGG  
GAAGATATCAGATCACAGTCGGTCTTGCAGGTGCAGCACCACCAAGATTCACATCGGATCC  
GTCGTCGTAAATTAATTTATGCGACGCGCCAGGTGATGAGAATCTTGATGATGCTGCATCA  
GCAGGCACTCGACTCGACTCACATCCTGCCGAAATCACTGTGAAACTGCGAATCAGCTGCT  
GACCGAGGTATGTACTTTGAGCTTTGATCCATCTTAGTATTCTGGCTTTAATTCATATAT  
CTCAAATATTCCGCTGCTGGATTTCAAGTGAAGATCAATTTGACCCAGCCCAAATTGGCAAGT  
GCACCGACCAATGCTTTAATTCGAATTC

### miR172b

>3A dna:chromosome chromosome:IWGSC:3A:741660143:741660535:-1  
CTCGAGATCATCATCATCTTCTAGCTAGATTTGTGAGAGAGTACACATATATGCCACATAGG  
CTAGCGTCGATGTCGAAATCTGCTTGCTGCAGCCAGCCAGCCGGTGATTTCTGGTGTGGCA  
TCATCAAGATTCACACAATGTATGGCATGCATGCGATGCTCTCATCCTTCCCATCCATACA  
TGCATGTGAACTGGAGTTCCGTTTCATGCATGGGAGATGGATAGATTGATGAGCTTTGCTAGC  
TAGCCAGCGTACTATGTGTGGGAATCTTGATGATGCTGCATCGGCGATCAATGGCTAATTAG  
TGATCAATGAACCAACCTACCTTGACGCCGCTCGCTCCTCACAACTCTCTCCCCATTTGT  
ACACTAACTCTCTCAGATCTAGTCCGGGAATTC

## A.3 Synthesised AP2-5A sequences

Golden gate overhangs highlighted in green. Untranslated regions highlighted in blue. Exons highlighted in pink.

### AP2-5A PROM+5' UTR

GGTCTCAGGAGGCAATACTAGAACAGAACCATGTGCTCCAGTACGCAACAAGCGAAGCCATC  
GCGACAAAAACAAAATGAAGAAAAAGAACGGTGCCTCCACCTCCCCCTCCGAGGCTCCGAC  
CCGCCCGCCCGCAGCTCAGCTCAAGGCGCGCACACAGACACAGGCCGGAGGGGGCGTTCCGGG  
CCGGCACGTACCCCCGCCCCACGTCCCGATCACCGGGTGCCTCACCTCACGGGCTCGCTC  
TCAACCGGTCAACGTACTGTACCGCACCGGTGCAGCCATTTACGGCGCTGTGCGGGTGTGCG  
CCGTCTCCCTCCGTCCGTCCATTCATCGGGTCTCCCGTGGCCGTGCCGAGCCCCACGC  
GGTGCCGTGACCGCGACCGTACGCGAGCCCGCCCGCCGGCCGGCCGTGACCACGAGCGTA  
AGGTTACGAGGATGCTACATACATGCGCGTACGGCTTACGTAGACAGAGCGAGTATACATGG

CAGACTTTGGCGGGGTGGGCGTCTCTCGGGTGGTGGGGATGCATTTGTGCGAAAGGAACAA  
CGTTTCCGATGGGGCGAGCGGGGGCGTGACCGGTGACCGGAACGGCACAGTGGCCCCGACA  
GGTACGCCTGTGGCTGCGTACCCGCCCTCGCGTGGGTCTACGTGCAATAATTGCACCCATCC  
CATTACACCCGGGCCCCGGCGAGCAAAACAGTACCCGGACCTAGCCTGCAACCCCCAAGCCC  
CCGACGCGGTACATGCACGCATACACACACGAAAGAGAGCGGTGCACGCAGTACGTACAC  
ACCGGCAGGCGGACGCGGTACGTGCTAGGCTAGGCCAGGCTAGATTGGTCCAGCTGCCGGCT  
CCCCCGTGCTCTCGTCGGCACGGCGGACATGCCGTACGTACCTGCTCCGCCTGTGGCCC  
TTGGCGCTTGCGCCCGGCCGGCCGCGGTGCGGTCAACACACGAGGCCTCCAGATCGGGCGC  
GGCATGCATGCATGCATGCATGCATGTGCCGCCGTACGTATGTACGTATACCGGCGCGGAT  
TAATTTAGAGTTCGATTTGATTAGAGGGAGGGAGGGGGCGTGCGGTGCACTTGGGCAATGT  
AATGCGTCTCGGAGGAGGGATCTCATCTAACCTAGCAGCACAGGGCGTACGGCCGGGGG  
TTATCTTACTCTCGCTAGGTGCCTAAGATAACCAGCATGAGTTGATGGTGCCGGCCGTTAAC  
AATTCCAACAAAAGCTAATCGTCTGGTGCACACGCAATGGTGGACACTATCATAACCATGGA  
TCAGTGGGTGGTTCTTTGTCCATGCCACTGGTTTTTCTCTTTTCGCTAGAGTTAATTAAGC  
GAATGGCTTTTGAATCCGTGTTTGTTCATGTCGGTCAAATCAAATCTCATGAATATGTTGTA  
CCGTTTGTCTAGATGGAGCTGGATTAATAAATTCGGCATTGTGGGTGCCCTCCATACG  
ATGTACTGGAGAGGGTTTTTAAACATTTTGTGCTTAGCGTTTGGCTAGCGATGTAATAACA  
AAATAGTGTGCATGACATCTGCCACAAAGGTGCGTCGACATCCATACTTTACTCTTCGTTGCA  
CTACAAACATATTTTGTGATGTACGCTCCGTGTGATATATACTTGTATCTTTAGAAAGGC  
CTAGAGTCTTAGGTGTAAGACCTTTCTAGCTGCAGGCCTGATTCGCACAAGGTGCTACTTGG  
TTGTTGCATATAAAAAACGGTAATACTATAATACTCTGTCCCTCCGGGACTAAATTACTTAG  
CAAGTTGACATTTGTTTATATTTAGCTTCGGGAACGTCCAACGTCTCCGGTGTAAAGACCTTT  
CTAGTCGCATTCCTGATTCAAACAAGATGCTACTTGGTTGTATATAAAATGATTGGGCTATA  
ATACTCCTCTTCGCAACAATTAATTACTCAGTAGGTGAACATTTTATATATTTCAACTCTAG  
CATATAATACAAATTTGTTGTCGGCATAGAAATGAGTTGCAATCATTTTATAAATAGGAAGA  
ACACATGGCTTATACAAACTAGCAGCCTAAAAAGGTTTTTTTTTTCTTCTTTCTGAGAGGAG  
GCATTTAGCTTGTGGAGCAAAATGTTGAAGCGGCTGGGCGAAAAAACCTCGGCTGATCCGC  
GTGAGGGCACGACACGTGGCGTCCCGTCCACGGGGTGTGTGGCCGTAGCGATTAGCGAGGC  
TCCGCTGCACAAAAATAGTTTACCTCTGATGCCCTTGGCCTCCCCGACGTCCCATCTCGCTT  
TCTCTCTTTCTCTTTCTCCCACTGGCCTGGCCCCCTCTCCTCGTCGTCCTCCAGTCCTCATC  
CCCCGCCCCATGGCCACCACCACCGCCGCTCCCCCTCCCCCACCCTACTTCTAC  
TCCCCCGCCCCGCCCTCGCAGCCCGCGCCACCGCGCTCCCATGCCATAGACGCGACCCCA  
CTCATCGGTCCAGGTCGGTTCGCTCGGAGCCGAGCGGCGGGCGGGCGAGGAGTGCCTTTT  
ATTCGGTCCCGGCGGGCCTCGGATCGGAGAGACC

**AP2-5A CDS**

GGTCTCATCGGAGATGGTGCTGGATCTCAATGTGGAGTCGCCGGCGGACTCGGGCACGTCCA  
GCTCCTCCGTGCTCAACTCCGCGGACGCCGGTGGCGGCGGCTTCCGGTTCGGCCTGCTCGGG  
AGCCCTGATGATGACGACTGCTCCGGCGAGCCGGCGCCGGTTCGGGCCCGGGTTCGTCACGAG  
GCAGCTCTTCCCCGCTCGCCGCCGGGCACGCGGGCGCGCCCGGGGTGACGATGGGGCAGC  
AGGCCCCGGCGCCTGCGCCGATGGCGCCCGTGTGGCAGCCGGCGCGCCGAGGAGCTCCTC  
GTGGCGCAGCGGATGGCGCCCGGAAGAAAACGCGCGGGGCCGAGGTGCGCGAGCTCGCA  
GTACAGGGGCGTCACCTTCTACCGCAGGACCGGCCGGTGGGAGTCGCACATCTGGTACAGCC  
TCCTCTCGTCTCCCTACTCCTCCTCCATGACTATAGTTATTACCCAAATCACTAATCAATCG  
CTGGCCTTTGTCCGTCTGATTCCACCAGGGATTGCGGGAAGCAGGTCTACTTGGGTGAGCTC  
AAACAAATCCCAGCTCGAGCTCCGATCTCCTCGGTGTCTAATTTTCGATTATCTTAGCTGTAT  
GGGCTCGTGATTAACACTGGATAATTTCTTCAGGTGGTTTCGACACTGCGCACGCGGCCGCA  
AGGTGAACTAATTAATTAACCAGGCCTCGTTTGAATTCCTTTTGCCCCGATTTGGCCACGAA  
CTGTGTACTGAGATGAGACGGTGTGGCGTGGCGCAGGGCCTACGATCGCGCGGCGATCAAGT  
TCCGGGGGCTGGAGGCCGACATCAACTTCAATCTGAGCGACTACGAGGAGGATTTGAAGCAG  
GTAATCTTATCCAAGCCTAGTTGATTGCTGTACTACCAAGTAGTGCGCCGATACATATGTA  
TCGGTGGTTTTGTCCGATGGTTGATATCTGGTGGGTGGTGGTGGTGGTTTTTTGCCAGATGAGG  
AACTGGACCAAGGAGGAGTTTCGTGCACATCCTCCGCCCCAGAGCACGGGGTTCGCCAGGGG  
GAGCTCCAAGTACCGCGGCGTACGCTCCACAAGTGCGGCCGCTGGGAGGCAAGGATGGGCC  
AGCTGCTCGGCAAGAAATAAGCAGGCACACACACAGCTCACGCACTAAAAATTAATCACTTC  
GCCACATTATCATAGTAGTAGTTTCTTTTATCAAATGCCATTGACAAGATTCAGTTGAAAT  
GAAATTTACAGACTGCTCATGAACTTGACACTAATTAGTAGTAGATGTGACAGGGCAGCTG  
TTCATGCTGTGGACGTTAATTAGCCTGTGCGAGGTAATCATCTTAGATTACCCCTTTTGAA  
ACATAATCTTAGCTGGTTTTAGGGTAGGGTCATCAAGTTAATCCATGTTGTTAGTTGTTGGCG  
CGTCTGTGTTGGTGTGGTGAGACGTCCACTTCCCCGACACGACACTCGATTGCAGACAT  
CTATTTGGAGCAACTGTTAGGCTCCACATAAGTATATGATCGAGTCGTCCAGACAAAATTAG  
TCTAATCCAATCCGTGCACATTATGGTCCAGACCCAGAATTGTCACCCTACACCTTACACCC  
CCTATCTCCGTTGAGCTGTAGTCTCATCTCATATACTGATACCCACATTATCATCACACGC  
GCAGGTACATATATCTGGGCCTCTTTGACAGCGAAGTTGAAGCTGCAAGTACTTTGATTTG  
CGCTGATCATTAAACATTTGGGCTCACAAAATTCCTTAATCTTTGCTCACTCACTCTAGTGT  
CATGTTGGATTTGGTAGGGCGTACGACAGGGCGGCGATTTCGCTTCAATGGGAGGGAAGCTGT  
GACTAACTTTGAGAGCAGCTCCTACAATGGGGATGCTCCACCCGACGCCGAAAATGAGGGTA  
CTACTACAATCAGTCTCACCTGTGCAATTTCTCCAATCACACCCAAGTCTTATCTCATCGAT  
ATCGTTGTTGTTTTCTTAGCAATTGTTGATGCTGATGCTCTTGACTTGGATCTGCGGATGTC  
GCAACCCACCGCGCACGATCCCAAGAGGGACAACATCATCGCCGGCCTTCAGTTAACTTTTG  
ATTCCCCTGAATCGTCAACCACAATGATCTCTTCTCAGGTAAAGAAACAAAATTATGTTAG  
CACTAGCTAATTCAGTGGTTAGATTTGCTCAAATAGAAAATTTGCTGTTGGCCAGTTGTGT

CGGTGGATGAGATATTATTGATCCAGGTCGTATGGCATATACTCTGTCTAACTTACGATA  
TTTGTGCCTTTTATATCGCTGTAGCCAATGAGCTCATCTTCGTCCCAGTGGCCTGTGCATC  
AACATGGCACGGCAGTAGCACCTCAGCAGCACCAGCGTTTGTACCCATCTGCTTGTGCATGGC  
TTCTACCCGAACGTACAGGTATCATCATCACTACGAGAACGACCACCTCCTCCTCCTCCTCC  
TCCTCTGCTTGGTGCCTGCTCCACCAGCTTACTGAAACTGTTGCTACCCTGAATAATGTT  
TCTGAAGAACTGCTAATAACATTTTCAGTTTTTCGACAAGTTCGTTCCCTGAAATTAAGGTTT  
CCATGCTTCTTTCCTAAAATAATCACGCCTGCTCTCCTAAAACATCTACAGGTGCAAGAAT  
TTTGGGTGTAGTGTGTCTGTTGACTGAGCATCTCGTCTCATGGACAGACAGAGCATGCT  
GTGTGGTAGTACCAGAGTACTTACTAGATGTGGGCACTACTGTTTGCCTGTGAGCTCGCAC  
CACCTTTCGAAAAAACTGCAGTGCGCCCTCCGGTTCTTGCAGTTCATCCGTCCCCATGG  
CACAGCTTTAGATGCAGCAGCAGCTTGCTTAGTTGTAGTACCCTGATCACATGGCGCAGCTT  
TATCTTGGTAGCCACTGTGCATTCACATGAAAGCAAAGCTTGGTGCATGCACGGCCATGA  
CTTGACGCTCTATCTCACTGCTGGTGTGCTGGTGCAGGTGCAGGTGCAGGAGAGGCCCATGGAG  
GCAAGGCCCCCTGAGCAGCCGTCCTTCCCCGGCTGGGGGTGGCAAGCGCAAGCCATGCC  
GCCGGGCTCCTCCCCTCGCCGTTGCTTTACGCTGCAGCATCATCAGGATTTTCTACCGCCG  
CCGCCGGCGCAACCTCGCCCCGCCGCCGCTACCCGGACCACCACCGTTCTACTTCCCC  
CGCCCCGGGACAACCTGAAGAGACC

**AP2-5A 3'UTR**

GGTCTCAGCTTAGCTGGCCGTTGTGACCAGACGGCGGTGGGTGCGCGCGGTGAGGTGTTCC  
CTCCTCGTCGTCGGTAACGCTTGTGTGAACTATAATCGGAGAGAGATGACATTGCCAGGC  
CATGTGTGGTGACACTACTGGCTAGTCTCTCGCCGCTCGCCATGATCGGGATCACGCGGAT  
CATGGCTGTTCAATAGATTCTCATGTATCCAATGTTCAAGTTTCCCAAACGGTTGAAAAAC  
TTTGAAATTTGTGATGGCAAATTCATGCATGGGTGCGCACGGTGCGCCCTTGCAACAAGCGCG  
GGAAGCACTCTGTTGTTTTACAGCAGTCCGCCATTGTTGCGTTAACCGTTGGGCCTTCTGGC  
CGAGGGCGAGCGGCGATTGACAGGCGGGCCCTCCCCGGACGCGACTGGTGGGCGACTAGGTG  
AGGTGAGGTGAGGTGAGTAGGCTAGGCGGTGGGTGAGTATTGTTTCGTTGCCATCCTCCATT  
CTGGGAGGGGAGCAAGCTGTGATTGATGGGCCGTAAACCGCAGCAGGCATCACTTTTTCTC  
CTCGCTTCCCTATTCAATCCCGCACAGTGAAGTGAAGCCTTTGTGGGTGCCCGGCCCTGTG  
GCGCTAGCGATAGCTAATAAGCCAGTACCCATGTTTGCACCCACCCACACACCCATCCA  
TCCATCCACAATAGCAAAGCTACCCGTAGTACTATAACCAGCGCCACTGTGCCATGCCAGTAG  
TACTACGCTGTAGCCTGTAGCTGAGCGTGGGGCTCGCTCTCGCTCTCGCTCTGGCTGGCCTG  
GCCGTGCAGAGTGCCTAGAGCGGGCACTGGAGCGGCACAGTGCGGCCCGGCCGGAGCGGCT  
TCTCGTCCGCTGCGCCCAATGGCGCCGCGCGGTGTGTGCAGCGCCAGAGCATGCATGCGTCC  
CGCCCGCCCTCCCTCCGCGGGCGCGCTGCAGAATCCCCACCCTGTTTTGGCGATTTCGTGCGC  
CTCGGTGCGCCGCAACTGCACTGCAGCTACCTCTGCGGCGCGGACGTAGGTGGCGCCGGCG  
CTGGAGTAGATAGATAGACGGCTGGGAAGGGAAGCCATTGGGGGCCATGGCGTTGGCATGG

CACAGCTGTCCCTTCCGGGTGGCTCCGTGGGGCGCTGTTCCAAGTGCAGTCCCGGGGGACC  
CCCTGCTGCGTTTGACCGGACCTCTTTTTCTTTTTACTACTCCGTCCGCTGCGTGGTGGTTG  
GTCCGTTGGGGCTTCCTCCCTGTTTGTTCCTTGCACGCCCCACTGGGCGGTGGGCTGCC  
ACGCCCACGCACGCACGCTTGCACGGGACCGGGCTCAAATGCTCGCACTTCTTGGTGCCCGA  
CTTTTTCATTTCTAGGTGTTGGTGTGGTGTGGTTTTGAAGTGTGGTGAATGGTCCGGGG  
GCTAGATGATGATGCTGCTGCTGCTGCGTGTGCTCTGCCCAGAACAGCTGCTAGTACACTG  
TGCACCGCTCACTCCTCGCCGTGCATCTCTATCCATTGCCCAAAGTCTAGTTTTCGATAGTTT  
TGGTAAAGTTTAGCCTGT **TAGAAGAGACC**

## A.4 Synthesised 3x FLAG tag sequence

Exons highlighted in pink. 3x FLAG tag highlighted in yellow.

CAATTGTTTCACATGATATATGTGACTGCAG **GACCGGCAATTCATGCAACAAATTAGTGACC**  
**TCCAACACAAG**GTGAGGAATGATCTCTATTGTTTATTTAGGCTTTTGGCAGCAGATTGCGCA  
GTTCCTTTTTTGGATCATGTATGTCAGGCAAATTATCAAAGTACTTACACATGTATATTTCT  
TTACAGAGGGAGTATTAGATATACATGTATGTAATGGAGTCGAATGGGTTTTATGGCAATTA  
CCTGAATTGTTGCAGGCATGTAGGAAATTAATTGTTTGAAGTGGTTCAAAAAAATTGTTTG  
AAAAATGTTGCATGCTAAGAACCCTAATTTATAGTGAATAATATTGTGTATTACATATGTTGC  
TGATTGCTGAAGTCTCAAGTGACCTTAATTTATATGGACTATTGCTAATAGAGCCATTAGT  
AGGAAATTAATGTTAATTCCAATTCTCATAGATTTCATAGGGAATGCCCTTGCAGTATGC  
ACAGAGATGGAGAATGATATGAATATACATGTGGAAGTAGTAGATATGTGCATGTCATGTAC  
AACAGTCAATACCTAAAATTGGCACAGTACATGTTAAACAAGAGGCAAATCGACAAGTGA  
CTCTTTTTTCTTTCCCATATTGATAATGTCAAGAGAGGTTGGGGTAGACCGAACTTGACATG  
GGAGGAGTTCATAAAGAGAGATCTGAAGGACTGGAGTATCACCAAAGAACTAGCCATGGACA  
GGGCATGTGGAAGCTTACTATCCATGTGCCAGAACTATGAGTTGGTTGCGAGATCTTGTGG  
GTTTCACCTCTAGCCTACCCCAACTTGTGGGACTAAAGGCTTTGTTGTTGTTGTTGTTGTTG  
ATATTGAATGAAATGTTATCAGCAATATATCCCTTTATGTGTGGTTTTAAAATTTTATTATT  
TATTTAAGTGAAGAAATTATCTTTTACCACCAACACAGTCAATGGAGTTGCTCCCTGATT  
GCATTATTTTTATTATAAACTATTCAGCAACACAAAATTGACTAGATTAAGTTTATTGGTTT  
TGAGAATATTTTTGGTGTACTTCTACAAGGTGAACATTTGATAATGTAATTTCTCTTTTCT  
TGTAG **GGAACACAGCTGGCAGAGGAAAATATGCGCTTGAAAAACCAAG**TAAATAAGATTTAG  
ATGCATTTCGTAGTCCCTGCACTTAATCAATCTTTATAAGAAGTGCAGCTTGAACAAAAGTCCA  
ATAAAAGTTGAGGTGGCAGAGCACATTTACATCATGTTTCGTGTGCTTGGATGTAG **ATGCA**  
**TGAGGTGCCAACTGCTAGCACCGTGGCCGTGCGGAAGCCGAAAATGTTGTCCCTGAAGATG**  
**CTCATTCATCTGACTCTGTGATGACGGCAGTACATTCGGGAAGCTCACAGGACAATGATGAC**  
**GGTCTGATATATCCCTGAAACTAGC**GTGAGTTTCGTCCAGAAACCCTTGCTCATCATCTTT  
ATCCTGATCATATGCTAGTCTCAAATGCCCTTCATAGAATCTGCGACATGAGAAAAAATATA  
TATAGTAGTTACTCCCTCCGTTCCCAAATTTGTATATTTTGTACTCCCTCTGTACTTATACTT

TTGGGACGGAGGGAGTACTATTATGGGACGGAGGGAGTAGTTAGCTCTGATAACCGGAGATG  
AGTCTGACAGCTCTAATTTCAATGTGTGTTTGTTCGGTTTTTATTTTTGAGGGAACCAATGTG  
TGTTTGTACCTGACTGAGTTACCAACAGTTTCTATCATCAATTCTGAAATGTGGCAGATTC  
TTATTCACTGAACAGGTTACCTTGAAGGATTATAAGGACCATGACGGAGACTATAAGGACC  
ATGACCTCGACGCTGCAGCAGCGGATTATAAGGACGATGACGATAAGTGAAGGACCGTGGGGA  
GGCCACCATGAGTTGCCTTGGTAGAGAGAAGATGTGAAATGCTATGGGGAATCACTCCAAGA  
TAGGCTGGCTGGAATGATCCCATAGTAGAAGCCAGATCAGTTTAACCCGATCGAGTTGTCGT  
TTTATGTCTCCGTTTGCACGGCGCAGTTGTCCACCTTTGTG

## A.5 Candidate miRNAs

Table A-2: Candidate miRNAs generated by miRDeep-P2, miRador and ShortStack from an early wheat spike development sRNA-Seq timecourse referenced in this thesis. The MIRNA loci and strands listed are locations within the IWGSC v1.0 assembly<sup>15</sup>. The known miRNA names listed are predicted by either miRador or ShortStack. The closest miRBase hit is the result with the lowest E-value (below 1.0) from a BLASTn search of the miRBase database. The full dataset can be found in Supplementary Table S1.

<b>miRNA ID</b>	<b>Mature miRNA sequence</b>	<b>Source</b>	<b>Known miRNA name</b>	<b>Closest miRBase hit</b>	<b>MIRNA chromosome</b>	<b>MIRNA strand</b>	<b>MIRNA start</b>	<b>MIRNA end</b>
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr7A	+	8252405	8252522
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr1D	+	418305991	418306183
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr6D	-	460873276	460873362
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr1A	+	515766449	515766583
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr1B	+	566097624	566097802
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr6A	-	607452479	607452567
mirna_candidate-30	AGAAUCUUGAUGAUGCUGCAU	mirdp2		aof-miR172	chr6B	-	702433189	702433275

mirna_candidate-120	CGAAUGUAUUUUUUAUGGCUUG	mirador, mirdp2, shortstack		chr1D	-	325491958	325492077
mirna_candidate-120	CGAAUGUAUUUUUUAUGGCUUG	mirador, mirdp2		chr1A	+	419025397	419025631
mirna_candidate-120	CGAAUGUAUUUUUUAUGGCUUG	mirador, mirdp2, shortstack		chr1B	-	437834545	437834664
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr7D	-	8968232	8968276
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr4D	-	10922263	10922307
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr5D	-	29594820	29594864
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr7D	+	208677267	208677311
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr5D	+	275557436	275557480
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr3D	-	453160517	453160561
mirna_candidate-125	CGCCGGCUGCGCGUUGCCCUCU	mirdp2		chr7D	-	467426664	467426708
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2	tae-miR531	chr2B	+	9690118	9690169
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2	tae-miR531	chr4B	-	14040702	14040753
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2	tae-miR531	chr2B	-	26516118	26516169
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2	tae-miR531	chr7A	-	62272024	62272075
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2	tae-miR531	chr7A	+	86118711	86118763
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2	tae-miR531	chr1A	-	417390284	417390335

mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2		tae-miR531	chr5A	-	437164867	437164918
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2		tae-miR531	chr5B	+	464176870	464176921
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2		tae-miR531	chr5B	+	507633368	507633420
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2		tae-miR531	chr5A	+	616555279	616555330
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2		tae-miR531	chr4B	-	646961073	646961124
mirna_candidate-147	CUCGCCGGAGCAGCGUGCAAC	mirdp2		tae-miR531	chr4B	-	658524819	658524871
mirna_candidate-153	CUCGUCGGAGCAGCGUGCAGC	mirdp2		tae-miR531	chr6B	+	683210485	683210538
mirna_candidate-180	GCAGCACCAACCAAGAUUCACA	mirdp2		ata-miR172b-5p	chr6D	-	460873276	460873362
mirna_candidate-180	GCAGCACCAACCAAGAUUCACA	mirdp2		ata-miR172b-5p	chr6A	-	607452479	607452567
mirna_candidate-180	GCAGCACCAACCAAGAUUCACA	mirdp2		ata-miR172b-5p	chr6B	-	702433189	702433275
mirna_candidate-190	GGAAUCUUGAUGAUGCUGCAU	mirdp2		ata-miR172c-3p	chr3D	-	608542525	608542698
mirna_candidate-190	GGAAUCUUGAUGAUGCUGCAU	mirdp2		ata-miR172c-3p	chr3A	-	741660251	741660421
mirna_candidate-190	GGAAUCUUGAUGAUGCUGCAU	mirdp2		ata-miR172c-3p	chr3B	-	818985888	818986057
mirna_candidate-289	UCGGACCAGGCUUCAUCCCU	mirador, mirdp2, shortstack	miR165	ata-miR5168-3p	chr5D	+	350379778	350379918
mirna_candidate-289	UCGGACCAGGCUUCAUCCCU	mirador, mirdp2, shortstack	miR165	ata-miR5168-3p	chr5B	+	411091726	411091864

mirna_candidate-289	UCGGACCAGGCUUCAUCCCC	mirador, mirdp2, shortstack	miR165	ata-miR5168-3p	chr5A	+	450881167	450881305
mirna_candidate-290	UCGGACCAGGCUUCAUCCCC	mirdp2		aof-miR166d	chr6D	+	418216637	418216753
mirna_candidate-290	UCGGACCAGGCUUCAUCCCC	mirdp2		aof-miR166d	chr6A	+	560691819	560691919
mirna_candidate-290	UCGGACCAGGCUUCAUCCCC	mirdp2		aof-miR166d	chr6B	+	630041200	630041318
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirador, mirdp2, shortstack	miR165	aof-miR166d	chr4A	-	16475386	16475524
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirdp2, shortstack		aof-miR166d	chr1D	+	254662220	254662355
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirdp2, shortstack		aof-miR166d	chr1A	+	326329184	326329319
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirdp2, shortstack		aof-miR166d	chr1B	-	357916215	357916352
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirador, mirdp2, shortstack	miR165	aof-miR166d	chr4D	+	450196289	450196418
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirdp2		aof-miR166d	chr5D	-	458932220	458932295
mirna_candidate-291	UCGGACCAGGCUUCAUCCCC	mirdp2, shortstack		aof-miR166d	chr7D	-	495321574	495321769

mirna_candidate-291	UCGGACCAGGCUUCAUUGCCC	mirdp2, shortstack	aof-miR166d	chr7B	-	523239718	523239915
mirna_candidate-291	UCGGACCAGGCUUCAUUGCCC	mirdp2, shortstack	aof-miR166d	chr7A	-	561152157	561152352
mirna_candidate-291	UCGGACCAGGCUUCAUUGCCC	mirdp2	aof-miR166d	chr5B	-	562984082	562984157
mirna_candidate-291	UCGGACCAGGCUUCAUUGCCC	mirdp2	aof-miR166d	chr5A	-	578021262	578021333
mirna_candidate-292	UCGGACCAGGCUUCAUUGCCU	mirdp2, shortstack	csi-miR166f-3p	chr6D	+	418216615	418216773
mirna_candidate-292	UCGGACCAGGCUUCAUUGCCU	mirdp2, shortstack	csi-miR166f-3p	chr6A	+	560691797	560691939
mirna_candidate-292	UCGGACCAGGCUUCAUUGCCU	mirdp2, shortstack	csi-miR166f-3p	chr6B	+	630041178	630041338
mirna_candidate-293	UCGGACCAGGCUUCAUUGCCU	mirdp2	ata-miR166c-3p	chr6D	+	418216460	418216530
mirna_candidate-293	UCGGACCAGGCUUCAUUGCCU	mirdp2	ata-miR166c-3p	chr6A	+	560691661	560691731
mirna_candidate-293	UCGGACCAGGCUUCAUUGCCU	mirdp2	ata-miR166c-3p	chr6B	+	630041050	630041120
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr3D	-	65610096	65610183
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr3A	+	76564341	76564428
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr6D	+	104978825	104978909
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr3B	-	109363642	109363729
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr6A	+	127182602	127182686
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr6B	+	191531744	191531828

mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr6D	+	287888861	287888949
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr5D	-	339371558	339371671
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr5B	-	398367422	398367533
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr6A	-	437636987	437637075
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr5A	-	440504305	440504418
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr6B	+	448349739	448349827
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr2D	-	456504546	456504635
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr2B	+	537835523	537835611
mirna_candidate-351	UGACAGAAGAGAGUGAGCAC	mirdp2	aof-miR156a	chr2A	-	599784388	599784477
mirna_candidate-367	UGCAGUCCUCGAUGUCGUAG	mirdp2, shortstack	tae-miR9664-3p	chr1D	+	236440928	236441043
mirna_candidate-367	UGCAGUCCUCGAUGUCGUAG	mirdp2, shortstack	tae-miR9664-3p	chr1A	-	309399470	309399588
mirna_candidate-367	UGCAGUCCUCGAUGUCGUAG	mirdp2, shortstack	tae-miR9664-3p	chr1B	+	341933715	341933830
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7B	+	9129757	9129833
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7B	+	9130350	9130428
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7B	+	9131092	9131168
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7D	+	67104716	67104792
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7D	+	67105308	67105385
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7D	+	68403784	68403863

mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7A	-	72697128	72697204
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chr7A	-	72700595	72700671
mirna_candidate-397	UGUAGAUACUCCCUAAGGCUU	mirdp2	ata-miR5200-3p	chrUn	+	84099175	84099254
mirna_candidate-403	UGUGCUCUUUCCUUCUUACCC	mirdp2	aly-miR156h-3p	chr7B	+	41539403	41539474
mirna_candidate-403	UGUGCUCUUUCCUUCUUACCC	mirdp2	aly-miR156h-3p	chr7D	+	90623055	90623153
mirna_candidate-403	UGUGCUCUUUCCUUCUUACCC	mirdp2	aly-miR156h-3p	chr7A	-	91921006	91921103
mirna_candidate-403	UGUGCUCUUUCCUUCUUACCC	mirdp2	aly-miR156h-3p	chr3A	+	275849880	275849947

---

## A.5 Predicted miRNA targets

Table A-3: Predicted targets of candidate miRNAs referenced in this thesis. The TargetFinder score indicates confidence in the prediction. A lower score denotes higher confidence, the cutoff used here was 4. Transcripts and target sequences given are from the IWGSC v1.0 annotation. The full dataset can be found in Supplementary Table S2.

miRNA ID	Transcript target	TargetFinder score	Target site start (transcript-relative)	Target site end (transcript-relative)	Target sequence
mirna_candidate-30	TraesCS5B01G486900.1	2	677	697	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS5D01G486600.1	2	1235	1255	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS5D01G486600.2	2	1241	1261	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS1A01G058400.2	2	1355	1375	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS1B01G076300.3	2	1364	1384	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS1A01G058400.1	2	1795	1815	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS1B01G076300.1	2	1784	1804	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-30	TraesCS7D01G536300.1	3	709	728	AAGCAG-AUCAUCGAGAUUCU
mirna_candidate-30	TraesCS7B01G473200.2	3	709	728	AAGCAG-AUCAUCGAGAUUCU
mirna_candidate-30	TraesCS7B01G473200.1	3	825	844	AAGCAG-AUCAUCGAGAUUCU
mirna_candidate-30	TraesCS7D01G447300.1	3.5	5131	5151	AUGCAGUCUUGUCAAGAUUCU

mirna_candidate-30	TraesCS7B01G359800.1	3.5	5140	5160	AUGCAGUCUUGUCAAGAUUCU
mirna_candidate-30	TraesCS7D01G447300.2	3.5	5137	5157	AUGCAGUCUUGUCAAGAUUCU
mirna_candidate-30	TraesCS7B01G440400.1	4	1121	1141	CAGCAGCAUCAUCAGGAUUCC
mirna_candidate-30	TraesCS7D01G512600.1	4	1249	1269	CAGCAGCAUCAUCAGGAUUCC
mirna_candidate-120	TraesCS7D01G059400.1	4	571	592	CAGGAUCUAAAAAAUACAUUCA
mirna_candidate-147	TraesCS7A01G225000.1	2.5	276	295	GUUGC-UGCUGCUCCGGCGGG
mirna_candidate-147	TraesCS1B01G481000.1	3.5	26	46	GUGGCGCUUUGCUCCGGCGAG
mirna_candidate-147	TraesCS6A01G006000.1	3.5	45	65	UCUCCGCGCUGCUCCGGCGAG
mirna_candidate-147	TraesCS5A01G331200.1	3.5	50	69	GCUGC-UGCUGCUCCGGCGAC
mirna_candidate-147	TraesCSU01G218800.1	4	45	65	UAUCCGUGCUGCUCCGGCGAG
mirna_candidate-147	TraesCS1A01G059600.1	4	34	53	UCUGCACGCU-CUCCGGCGAG
mirna_candidate-147	TraesCS7D01G518600.1	4	503	524	AUGGUAUGGCUGCUCCGGCGAG
mirna_candidate-147	TraesCS7B01G448300.1	4	503	524	AUGGUAUGGCUGCUCCGGCGAG
mirna_candidate-147	TraesCS5B01G487800.1	4	50	71	GUGGCUACCCUGCUCCGGCGGG
mirna_candidate-147	TraesCS5D01G487300.1	4	50	71	GUGGCUACCCUGCUCCGGCGGG
mirna_candidate-147	TraesCS7D01G424700.1	4	438	459	CUUGCAGGACUGCUCCGGCGAU
mirna_candidate-147	TraesCS5D01G216900.1	4	581	600	GCUGCA-GCUGCUGC GGCGAG
mirna_candidate-147	TraesCS7B01G265300.1	4	73	92	GCCGC-CGCUGCUCCGGCGAC
mirna_candidate-153	TraesCS1D01G449800.1	3	112	132	GUCGCAUGCUGCUCCGACGAA
mirna_candidate-153	TraesCS1B01G290200.1	3	525	546	GCUGGACCGCUGCUCCGACGAC
mirna_candidate-153	TraesCS1D01G280300.1	3	513	534	GCUGGACCGCUGCUCCGACGAC

mirna_candidate-153	TraesCS6A01G006000.1	3.5	45	65	UCUCCGCGCUGCUCCGGCGAG
mirna_candidate-153	TraesCS5A01G331200.1	3.5	50	69	GCUGC-UGCUGCUCGGCGAC
mirna_candidate-153	TraesCS4A01G280700.1	3.5	653	672	GCUG-GCGCCGCUCCGACGAG
mirna_candidate-153	TraesCS1D01G252600.1	3.5	1658	1677	UCUGCAU-CUGCUCGGACGAC
mirna_candidate-153	TraesCS1B01G418800.1	4	162	183	GCUGUGCGACUGCUUCGACGAC
mirna_candidate-153	TraesCS1D01G398700.1	4	162	183	GCUGUGCGACUGCUUCGACGAC
mirna_candidate-153	TraesCS7A01G225000.1	4	276	295	GUUGC-UGCUGCUCGGCGGG
mirna_candidate-153	TraesCS1A01G059600.1	4	34	53	UCUGCACGCU-CUCCGGCGAG
mirna_candidate-153	TraesCS3A01G306200.1	4	428	447	GCUGCACGUCGCGC-GACGAG
mirna_candidate-153	TraesCS6B01G098400.1	4	1049	1068	GCAGCACGUCUCUG-CGAG
mirna_candidate-153	TraesCS7A01G047200.1	4	466	485	GCUGCAC-CUGCUUCGCCGAG
mirna_candidate-153	TraesCS5D01G513100.1	4	848	867	UCUGCACGUCUC-GACGAC
mirna_candidate-153	TraesCS5D01G216900.1	4	581	600	GCUGCA-GCUGCUCGGCGAG
mirna_candidate-153	TraesCS1D01G095600.1	4	1292	1313	GCUCUGCGUCUGCUCGGACGAC
mirna_candidate-153	TraesCS7D01G209900.1	4	510	529	GCUGGA-GCUGCUCGAUGGG
mirna_candidate-153	TraesCS1A01G103200.1	4	1008	1029	GCUCUGCGUCUGCUCGGACGAC
mirna_candidate-153	TraesCS7A01G113800.1	4	242	261	CCUGC-CGCUGCUCGGACGGC
mirna_candidate-153	TraesCS7A01G115300.1	4	224	243	CCUGC-CGCUGCUCGGACGGC
mirna_candidate-153	TraesCS7D01G111800.1	4	218	237	CCUGC-CGCUGCUCGGACGGC
mirna_candidate-153	TraesCS7A01G113600.1	4	197	216	CCUGC-CGCUGCUCGGACGGC
mirna_candidate-153	TraesCS7D01G108900.1	4	197	216	CCUGC-CGCUGCUCGGACGGC

mirna_candidate-153	TraesCSU01G095400.1	4	197	216	CCUGC-CGCUGCUCCGACGGC
mirna_candidate-153	TraesCS6A01G033400.1	4	777	798	GCUGCUCGUCUGCUCCGCCGAG
mirna_candidate-153	TraesCS6B01G047900.1	4	774	795	GCUGCUCGUCUGCUCCGCCGAG
mirna_candidate-153	TraesCS6B01G046900.1	4	483	504	GCUGCUCGUCUGCUCCGCCGAG
mirna_candidate-153	TraesCS5A01G339800.1	4	509	528	GCUGC-CGCUGCCCCGACGAC
mirna_candidate-153	TraesCS5B01G338900.1	4	485	504	GCUGC-CGCUGCCCCGACGAC
mirna_candidate-153	TraesCS1D01G381800.1	4	806	825	GCUGC-CUCUGCUCCGACGGC
mirna_candidate-153	TraesCS1B01G394900.1	4	794	813	GCUGC-CUCUGCUCCGACGGC
mirna_candidate-153	TraesCS7B01G265300.1	4	73	92	GCCGC-CGCUGCUCGGGCGAC
mirna_candidate-153	TraesCS7B01G344900.1	4	30	49	GCUGCUC-CUGCUCCGUCGAG
mirna_candidate-153	TraesCS7D01G434000.1	4	30	49	GCUGCUC-CUGCUCCGUCGAG
mirna_candidate-153	TraesCS6B01G046800.1	4	303	324	GCUGCCCUACUGCUCCGACGAC
mirna_candidate-153	TraesCS6A01G033300.1	4	303	324	GCUGCCCUACUGCUCCGACGAC
mirna_candidate-153	TraesCS6D01G401200.1	4	650	669	GCGGCAAGC-GCUCCGACGAG
mirna_candidate-153	TraesCS4A01G299500.1	4	80	101	GCGGCCUCGCUGCUCCGACGAC
mirna_candidate-180	TraesCS6D01G295000.1	3	22	43	UGCUGGGCCUUGGUGGUGCUGC
mirna_candidate-180	TraesCS4A01G104100.1	3	1916	1937	UGCUGUAUGUUGGUGGUGCUGC
mirna_candidate-180	TraesCS4A01G104100.2	3	2471	2492	UGCUGUAUGUUGGUGGUGCUGC
mirna_candidate-180	TraesCS4A01G218900.1	3.5	3075	3095	UAUGGAUCUUGGUGGUGUUGA
mirna_candidate-180	TraesCS6A01G195500.4	3.5	561	581	UGAGAAUUUGGGUGGUGCUGC
mirna_candidate-180	TraesCS6A01G195500.3	3.5	969	989	UGAGAAUUUGGGUGGUGCUGC

mirna_candidate-180	TraesCS6A01G195500.1	3.5	996	1016	UGAGAAUUUGGGUGGUGCUGC
mirna_candidate-180	TraesCS6A01G195500.2	3.5	996	1016	UGAGAAUUUGGGUGGUGCUGC
mirna_candidate-180	TraesCS5D01G090100.4	3.5	1608	1628	UGUGGUUGUUGGUGGUGCUGG
mirna_candidate-180	TraesCS5D01G090100.1	3.5	1674	1694	UGUGGUUGUUGGUGGUGCUGG
mirna_candidate-180	TraesCS5D01G090100.2	3.5	1698	1718	UGUGGUUGUUGGUGGUGCUGG
mirna_candidate-180	TraesCS5D01G090100.3	3.5	1752	1772	UGUGGUUGUUGGUGGUGCUGG
mirna_candidate-180	TraesCS5B01G406800.4	3.5	882	902	UGCUGCUCUUGGUGGUGCUGC
mirna_candidate-180	TraesCS5B01G406800.3	3.5	882	902	UGCUGCUCUUGGUGGUGCUGC
mirna_candidate-180	TraesCS5B01G406800.5	3.5	882	902	UGCUGCUCUUGGUGGUGCUGC
mirna_candidate-180	TraesCS5B01G406800.2	3.5	882	902	UGCUGCUCUUGGUGGUGCUGC
mirna_candidate-180	TraesCS5B01G406800.1	3.5	882	902	UGCUGCUCUUGGUGGUGCUGC
mirna_candidate-180	TraesCS7B01G501900.1	4	411	431	UGUGGAUAUUGGUGCUGCUGU
mirna_candidate-180	TraesCS5D01G030200.1	4	200	220	CGUGGGCGUUGGUGGUGCUGC
mirna_candidate-180	TraesCS5D01G030700.1	4	1544	1564	CGUGGGCGUUGGUGGUGCUGC
mirna_candidate-180	TraesCS6A01G375400.1	4	2433	2453	UGUCGGUAUUGGUGGUGCUGG
mirna_candidate-180	TraesCS6D01G359900.1	4	2439	2459	UGUCGGUAUUGGUGGUGCUGG
mirna_candidate-180	TraesCS7B01G260800.1	4	2344	2364	UUUACUUCUUGGUGGUGCUGC
mirna_candidate-190	TraesCS5B01G486900.1	2.5	677	697	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-190	TraesCS5D01G486600.1	2.5	1235	1255	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-190	TraesCS5D01G486600.2	2.5	1241	1261	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-190	TraesCS1A01G058400.2	2.5	1355	1375	CUGCAGCAUCAUCAGGAUUCU

mirna_candidate-190	TraesCS1B01G076300.3	2.5	1364	1384	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-190	TraesCS1A01G058400.1	2.5	1795	1815	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-190	TraesCS1B01G076300.1	2.5	1784	1804	CUGCAGCAUCAUCAGGAUUCU
mirna_candidate-190	TraesCS7B01G440400.1	3	1121	1141	CAGCAGCAUCAUCAGGAUUC
mirna_candidate-190	TraesCS7D01G512600.1	3	1249	1269	CAGCAGCAUCAUCAGGAUUC
mirna_candidate-190	TraesCS7D01G536300.1	3.5	709	728	AAGCAG-AUCAUCGAGAUUCU
mirna_candidate-190	TraesCS7B01G473200.2	3.5	709	728	AAGCAG-AUCAUCGAGAUUCU
mirna_candidate-190	TraesCS7B01G473200.1	3.5	825	844	AAGCAG-AUCAUCGAGAUUCU
mirna_candidate-190	TraesCS7D01G447300.1	4	5131	5151	AUGCAGUCUUGUCAAGAUUCU
mirna_candidate-190	TraesCS7B01G359800.1	4	5140	5160	AUGCAGUCUUGUCAAGAUUCU
mirna_candidate-190	TraesCS7D01G447300.2	4	5137	5157	AUGCAGUCUUGUCAAGAUUCU
mirna_candidate-190	TraesCS5D01G125500.1	4	476	496	AUGCACCAUCACCAAGAUUUC
mirna_candidate-190	TraesCS4A01G320300.1	4	59	80	AUGAUGGCAUCAUCAAGCUUCC
mirna_candidate-289	TraesCS5D01G052300.1	2.5	677	697	UGGGAAUGAAGCCUGGUCCGG
mirna_candidate-289	TraesCS1D01G155200.1	2.5	573	592	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS1D01G155200.2	2.5	573	592	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS1D01G155200.3	2.5	573	592	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS1B01G173900.1	2.5	579	598	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS1A01G157500.1	2.5	579	598	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS3A01G312800.2	2.5	570	589	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS5B01G378000.1	2.5	624	643	UGGGAU-GAAGCCUGGUCCGG

mirna_candidate-289	TraesCS5D01G385300.1	2.5	624	643	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-289	TraesCS3A01G312800.1	2.5	645	664	UGGGAU-GAAGCCUGGUCCGG
mirna_candidate-290	TraesCS5D01G052300.1	0.5	678	697	GGGAAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS1D01G155200.1	2	573	592	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS1D01G155200.2	2	573	592	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS1D01G155200.3	2	573	592	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS1A01G157500.1	2	579	598	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS1B01G173900.1	2	579	598	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS3A01G312800.2	2	570	589	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS5D01G385300.1	2	624	643	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS5B01G378000.1	2	624	643	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS3A01G312800.1	2	645	664	UGGGAUGAAGCCUGGUCCGG
mirna_candidate-290	TraesCS6D01G237600.1	3	166	185	GUGAAGGAAGCCUGGUCCGU
mirna_candidate-290	TraesCS6A01G256400.1	3	166	185	GUGAAGGAAGCCUGGUCCGU
mirna_candidate-290	TraesCS7D01G501400.1	4	756	775	GGAGAUGAAGCCUGGCCCGG
mirna_candidate-290	TraesCS4A01G158900.1	4	383	402	GUGGAUGAAGCCUGCUCCGG
mirna_candidate-290	TraesCS6A01G078100.1	4	363	381	GGGAAU-AAGCCUGGUCAGC
mirna_candidate-290	TraesCSU01G142400.2	4	906	924	GGGAAU-AAGCCUGGUCAGC
mirna_candidate-290	TraesCSU01G142400.3	4	996	1014	GGGAAU-AAGCCUGGUCAGC
mirna_candidate-290	TraesCS6A01G077700.1	4	973	991	GGGAAU-AAGCCUGGUCAGC
mirna_candidate-290	TraesCSU01G142400.1	4	986	1004	GGGAAU-AAGCCUGGUCAGC

mirna_candidate-291	TraesCS5D01G052300.1	1.5	677	697	UGGGAAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS1D01G155200.1	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS1D01G155200.2	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS1D01G155200.3	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS1A01G157500.1	3	578	598	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS1B01G173900.1	3	578	598	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS3A01G312800.2	3	569	589	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS5B01G378000.1	3	623	643	UUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS5D01G385300.1	3	623	643	UUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS3A01G312800.1	3	644	664	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-291	TraesCS6D01G237600.1	3	165	185	GGUGAAGGAAGCCUGGUCCGU
mirna_candidate-291	TraesCS6A01G256400.1	3	165	185	GGUGAAGGAAGCCUGGUCCGU
mirna_candidate-291	TraesCS4A01G158900.1	4	382	402	GGUGGAUGAAGCCUGCUCCGG
mirna_candidate-292	TraesCS5D01G052300.1	1.5	677	697	UGGGAAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS1D01G155200.1	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS1D01G155200.2	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS1D01G155200.3	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS1B01G173900.1	3	578	598	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS1A01G157500.1	3	578	598	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS3A01G312800.2	3	569	589	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS5B01G378000.1	3	623	643	UUGGGAUGAAGCCUGGUCCGG

mirna_candidate-292	TraesCS5D01G385300.1	3	623	643	UUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS3A01G312800.1	3	644	664	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-292	TraesCS6D01G237600.1	3.5	165	185	GGUGAAGGAAGCCUGGUCCGU
mirna_candidate-292	TraesCS6A01G256400.1	3.5	165	185	GGUGAAGGAAGCCUGGUCCGU
mirna_candidate-292	TraesCS6A01G078100.1	4	362	381	AGGGAAU-AAGCCUGGUCAGC
mirna_candidate-292	TraesCSU01G142400.2	4	905	924	AGGGAAU-AAGCCUGGUCAGC
mirna_candidate-292	TraesCSU01G142400.3	4	995	1014	AGGGAAU-AAGCCUGGUCAGC
mirna_candidate-292	TraesCS6A01G077700.1	4	972	991	AGGGAAU-AAGCCUGGUCAGC
mirna_candidate-292	TraesCSU01G142400.1	4	985	1004	AGGGAAU-AAGCCUGGUCAGC
mirna_candidate-293	TraesCS5D01G052300.1	2	677	697	UGGGAAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS1D01G155200.1	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS1D01G155200.2	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS1D01G155200.3	3	572	592	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS1B01G173900.1	3	578	598	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS1A01G157500.1	3	578	598	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS3A01G312800.2	3	569	589	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS5D01G385300.1	3	623	643	UUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS5B01G378000.1	3	623	643	UUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS3A01G312800.1	3	644	664	CUGGGAUGAAGCCUGGUCCGG
mirna_candidate-293	TraesCS6D01G237600.1	4	165	185	GGUGAAGGAAGCCUGGUCCGU
mirna_candidate-293	TraesCS6A01G256400.1	4	165	185	GGUGAAGGAAGCCUGGUCCGU

mirna_candidate-351	TraesCS7D01G245200.1	1	767	786	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS7D01G245200.2	1	770	789	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS5D01G273900.1	1	824	843	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS7A01G260500.1	1	902	921	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS5D01G273900.2	1	827	846	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS7D01G261500.1	1	923	942	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS5D01G294400.2	1	959	978	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS5D01G294400.1	1	977	996	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS6A01G155300.1	1	802	821	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS7A01G246500.1	1	813	832	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS5B01G286000.1	1	1335	1354	GUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS6A01G110100.2	2	1157	1176	AUGCUCUCUCUCUUCUGUCA
mirna_candidate-351	TraesCS2B01G029900.1	4	703	721	GUGCUCAUUCUCUUUU-UCA
mirna_candidate-351	TraesCS4A01G341400.1	4	958	976	GAGC-CACUCUCUUUUGUCC
mirna_candidate-397	TraesCS5D01G072900.1	3.5	1319	1339	AAGUCUUAGGGAGUAUAUACC
mirna_candidate-397	TraesCS7A01G115400.1	4	240	260	CAACCUUAGGGAGUAUCUCCA
mirna_candidate-397	TraesCS7D01G111600.1	4	240	260	CAACCUUAGGGAGUAUCUCCA

---

## A.6 transcript\_to\_genomic\_loci code

### 01\_Extracting\_exon\_information\_for\_loci\_transcripts.sh

```
#!/bin/bash

data_dir="path/to/working/directory"

### Extracting loci transcript IDs ###
#####

input_file="$data_dir/loci_file"
output_file="$input_file.2"
id_counter=1

# Check if input file exists
if [ ! -f "$input_file.gff3" ]; then
    echo "Input file not found: $input_file.gff3"
    exit 1
fi

# Read the input file line by line, skip lines starting with
"No results for",

# add IDs to the rest, and write to output file

while IFS= read -r line; do
```

```
# Skip lines starting with "No results for". This is a
feature of TargetFinder output.
```

```
if [[ $line == "No results for"* ]]; then
    continue
fi
```

```
# Skip comment lines and empty lines
```

```
if [[ $line == \#* || -z $line ]]; then
```

```
    echo "$line" >> "$output_file.gff3"
```

```
else
```

```
    echo "$id_counter $line" >> "$output_file.gff3"
```

```
    ((id_counter++))
```

```
fi
```

```
done < "$input_file.gff3"
```

```
echo "Modified GFF3 file with IDs written to: $output_file.gff3"
```

```
input_file="$output_file"
```

```
output_file="$input_file.3"
```

```
# Calculating the length of the loci
```

```
awk -F'\t' 'BEGIN {OFS = FS} {difference = $6 - $5 + 1; print
$0, difference}' "$input_file.gff3" > "$output_file.gff3"
```

```
echo "Modified GFF3 file with calculated values written to:
$output_file.gff3"
```

```

input_file="$output_file"

output_file="$input_file.4"

# Extracting transcript IDs from the GFF3 file descriptions
field

awk -F '\t' '{split($2, arr, " "); print $0 "\t" arr[1]}'
$input_file.gff3 > $output_file.gff3

echo "Modified GFF3 file with transcript IDs written to:
$output_file.gff3"

input_file="$output_file"

loci_id_file="$input_file.non-redundant_transcript_ids.txt"

# Extracting transcript IDs into a non-redundant list

awk -F '\t' '{split($2, arr, " "); print arr[1]}'
$input_file.gff3 | sort | uniq > $loci_id_file

echo "Transcript IDs extracted into non-redundant list of
transcripts $loci_id_file"

### Extracting list of exon positions for transcripts ###
#####

input_file="$data_dir/annotation_file"

output_file="$input_file.exons"

# Extracting exons from genome annotation GFF3 file

```

```

awk -F '\t' '$3 == "exon"' $input_file.gff3 > $output_file.gff3

echo "Exons extracted from genome annotation file into
$output_file.gff3"

# Extracting only the relevant exons from the bigger GFF3 file
input_file="$output_file"
output_file="$output_file.targets"

grep -F -f $loci_id_file $input_file.gff3 > $output_file.gff3

echo "Transcript features extracted from genome annotation file
into $output_file.gff3"

# Separating positive and negative strand genes
input_file="$output_file"
output_file_negative="$input_file.negative"
output_file_positive="$input_file.positive"

awk -F '\t' '{ if($7=="-") print $0}' $input_file.gff3 | tr -d
'\r' > $output_file_negative.gff3

echo "Negative strand features extracted"

awk -F '\t' '{ if($7=="+") print $0}' $input_file.gff3 | tr -d
'\r' > $output_file_positive.gff3

```

```

echo "Positive strand features extracted"

# Ordering negative strand exons by exon start location
input_file="$output_file_negative"
output_file="$input_file.2"

sort -k9,9 -k4,4nr "$input_file.gff3" > "$output_file.gff3"
#sorting exons by start location, then transcript ID

# Numbering negative strand exons according to their order in
this sorted file
input_file="$output_file"
output_file="$input_file.3"

awk -F'\t' '{
    split($9, a, ";");
    for (i in a) {
        if (a[i] ~ /^rank=/) {
            gsub(/rank=/, "", a[i]);
            print $0 "\texon_" a[i];
            next;
        }
    }
    print $0 "\t-";
}' "$input_file.gff3" > "$output_file.gff3"

# Calculating negative strand exon lengths

```

```

input_file="$output_file"

output_file="$input_file.4"

awk -F'\t' '{ diff = $5 - $4 + 1; print $0 "\t" diff }'
"$input_file.gff3" > "$output_file.gff3"

# Ordering positive strand exons by exon start location

input_file="$output_file_positive"

output_file="$input_file.2"

sort -k9,9 -k4,4n "$input_file.gff3" > "$output_file.gff3"
#sorting exons by start location, then transcript ID

# Numbering positive strand exons according to their order in
this sorted file

input_file="$output_file"

output_file="$input_file.3"

awk -F'\t' '{
    split($9, a, ";");
    for (i in a) {
        if (a[i] ~ /^rank=/) {
            gsub(/rank=/, "", a[i]);
            print $0 "\texon_" a[i];
        }
    }
}'

```

```

    print $0 "\t-";
}' "$input_file.gff3" > "$output_file.gff3"

# Calculating positive strand exon lengths
input_file="$output_file"
output_file="$input_file.4"

awk -F'\t' '{ diff = $5 - $4 + 1; print $0 "\t" diff }'
"$input_file.gff3" > "$output_file.gff3"

```

## 02\_Translating\_the\_loci.r

```

#####

### Loading necessary libraries ###

#####

library(tidyr)

library(plyr)

library(dplyr)

library(stringr)

#####

### Setting working directory ###

#####

setwd("path/to/working/directory")

#####

```

```

### Loading the output from script 01 ###

#####

positive_exon_info <- read.table(
  "annotation_file.exons.targets.positive.2.3.4.gff3",
  sep="\t", header=FALSE)

negative_exon_info <- read.table(
  "annotation_file.exons.targets.negative.2.3.4.gff3",
  sep="\t", header=FALSE)

target_loci<-
read.table("loci_file.2.3.4.gff3",sep="\t",header=FALSE)

#####

### Translating loci on the positive strand ###

#####

# Tidying up the data #

#####

colnames(positive_exon_info)[colnames(positive_exon_info) ==
"V1"] <- "chr"

positive_exon_info <- subset(positive_exon_info, select = -
c(V2,V3,V6,V7,V8) )

colnames(positive_exon_info)[colnames(positive_exon_info) ==
"V4"] <- "start"

colnames(positive_exon_info)[colnames(positive_exon_info) ==
"V5"] <- "end"

```

```

colnames(positive_exon_info)[colnames(positive_exon_info) ==
"V9"] <- "transcript_id"

colnames(positive_exon_info)[colnames(positive_exon_info) ==
"V10"] <- "id"

colnames(positive_exon_info)[colnames(positive_exon_info) ==
"V11"] <- "length"

# The code below is specific to TargetFinder output. Please
edit this code so that transcript_loci is a dataframe with 5
columns:

# (1) TS, (2) TE, (3) locus_id, (4) Tlength, and (5)
transcript_id. These correspond to

# (1) Transcript-relative locus start site, (2) transcript-
relative locus end site, (3) locus ID, (4) locus length, and
(5) transcript ID.

transcript_loci <- subset(transcript_loci, select = -
c(V1,V2,V3,V4,V7,V8,V9) )

colnames(transcript_loci)[colnames(transcript_loci) == "V12"]
<- "transcript_id"

transcript_loci <-
separate(transcript_loci,col=V10,into=c('col1','locus_id','col
2','col3','col4'),sep='=')

transcript_loci <- subset(transcript_loci, select = -
c(col1,col2,col3,col4) )

transcript_loci <-
separate(transcript_loci,col=locus_id,into=c('locus_id','col1'
),sep=';')

transcript_loci <- subset(transcript_loci, select = -c(col1) )

colnames(transcript_loci)[colnames(transcript_loci) == "V5"]
<- "TS"

```

```

colnames(transcript_loci)[colnames(transcript_loci) == "V6"]
<- "TE"

colnames(transcript_loci)[colnames(transcript_loci) == "V11"]
<- "Tlength"

# Converting the exon information into wide format #
#####

positive_exon_info$transcript_id <-
str_extract(positive_exon_info$transcript_id,
"TraesCS\\w+\\.\\.\\d+") # This line extracts the Traes ID from
the longer notes field. This will need to be edited depending
on the format of the notes field.

positive_exon_info <- positive_exon_info %>%
  mutate(chr_transcript_id = paste0(chr, ",", transcript_id))
positive_exon_info_wide <- positive_exon_info %>%
  pivot_wider(
    id_cols = chr_transcript_id,
    names_from = id,
    values_from = c(start, end, length)
  )

positive_exon_info_wide <-
separate(positive_exon_info_wide,col=chr_transcript_id,into=c(
'chr','transcript_id'),sep=',')

# Merging transcript and target info #
#####

```

```

transcript_loci <- unique(transcript_loci)

positive_targetexon_info <-
merge(transcript_loci,positive_exon_info_wide,by="transcript_i
d", all.x = FALSE, all.y = FALSE)

# Translating the loci #
#####

positive_translatingloci <- function(row) {

  # Splitting the dataframe into rows, so each row is worked on
independently

  row <- as.data.frame(row)

  # Setting up the counter for unassigned target nucleotides.
The number refers to the number of unassigned nucleotides
remaining for the row.

  unassigned_nt <- row$Tlength

  # Generating the numeric variables needed for the for the
function

  exon_no_counter <- 1 # Variable for the exon number we are
currently looking at

  next_exon_counter <- 2 # Variable for the next exon number
from the one we are currently looking at

  region_no_counter <- 1 # Variable for the target region we
are currently assigning

  current_transcript_loc_counter <- row$TS - row$length_exon_1
# Numeric variable for each TS-E1-E2 etc. value

```

```
previous_transcript_loc_counter <- row$TS # Numeric variable
for each TS-E1-E2 etc. value, but from the previous exon. I.e
if we are on exon 3, this will be TS-E1-E2.
```

```
while (unassigned_nt != 0) { # The script will run until the
number of unassigned nucleotides is zero.
```

```
# Generating character variables I need for for loop 1
```

```
region_start <- paste0("R", region_no_counter, "_start") #
Generating the text to label the region start variable.
```

```
region_end <- paste0("R", region_no_counter, "_end") #
Generating the text to label the region end variable.
```

```
exon_start_var <- paste0("start_exon_", exon_no_counter) #
Generating the text to label the exon start variable.
```

```
exon_end_var <- paste0("end_exon_", exon_no_counter) #
Generating the text to label the exon end variable.
```

```
exon_length_var <- paste0("length_exon_",exon_no_counter)
# Generating the text to label the exon length variable.
```

```
next_exon_start_var <- paste0("start_exon_",
next_exon_counter) # Generating the text to label the exon start
variable.
```

```
next_exon_length_var <-
paste0("length_exon_",next_exon_counter) # Generating the text
to label the next exon length variable.
```

```
if (current_transcript_loc_counter <0){ # Does the target
start in this transcript? True if it does. Ifelse statement 1.
```

```
##OUTPUT## Calculating target start site in this exon.
```

```
row[,ncol(row)+1] <- 0 # Make a new column
```

```

names(row)[ncol(row)] <- region_start # Rename this new
column

row[,ncol(row)] <- ((row[[exon_start_var]] +
previous_transcript_loc_counter) - 1) # Insert the value for
this new column

if (current_transcript_loc_counter > ((row$Tlength -1)*-
1)){ # Is TS - E(n) > the negative miRNA length? This is true
if the target site extends beyond this exon. Ifelse statement
2.

##OUTPUT## Calculating the end of this region (the end
of the exon)

row[,ncol(row)+1] <- 0 # Make a new column

names(row)[ncol(row)] <- region_end # Rename this new
column

row[,ncol(row)] <- row[[exon_end_var]] # Insert the
value for this new column (just the end of the exon).

unassigned_nt <- (unassigned_nt - ( row[[region_end]]
- row[[region_start]] +1)) # Updating the counter to show how
many unassigned nucleotides are left.

while (unassigned_nt != 0) { # While loop 2

# Updating numeric variables for this for loop

exon_no_counter <- exon_no_counter + 1 # Increasing
the counter by one so the next loop will look at the next exon.

```

```

        region_no_counter  <-  region_no_counter  +  1  #
Increasing the counter by one so the next loop will label the
region correctly.

        # Update character variables for this for loop

        region_start  <-  paste0("R",  region_no_counter,
"_start") # Generating the text to label the region start
variable.

        region_end <- paste0("R", region_no_counter, "_end")
# Generating the text to label the region end variable.

        exon_start_var      <-      paste0("start_exon_",
exon_no_counter) # Generating the text to label the exon start
variable.

        exon_end_var <- paste0("end_exon_", exon_no_counter)
# Generating the text to label the exon end variable.

        exon_length_var      <-      <-
paste0("length_exon_",exon_no_counter) # Generating the text to
label the exon length variable.

        ##OUTPUT## Calculating the start of this region (the
start of the exon).

        row[,ncol(row)+1] <- 0 # Make a new column.

        names(row)[ncol(row)] <- region_start # Rename this
new column.

        row[,ncol(row)] <- row[[exon_start_var]] # Insert the
value for this new column (just the start of the exon).

        if (unassigned_nt > row[[exon_length_var]]) { # Are
the remaining nucleotides to be assigned more than the length
of this exon? This is true if the target site extends beyond
this exon. Ifelse statement 3.

```

```

        ##OUTPUT## Calculating the end of this region (the
end of the exon).

        row[,ncol(row)+1] <- 0 # Make a new column

        names(row)[ncol(row)] <- region_end # Rename this
new column

        row[,ncol(row)] <- row[[exon_end_var]] # Insert the
value for this new column (just the end of the exon).

        unassigned_nt      <-      (unassigned_nt      -      (
row[[region_end]] - row[[region_start]] +1)) # Updating the
counter to show how many unassigned nucleotides are left.

    } else { # Else condition for ifelse statement 3 -
for if the region ends in this exon.

        ##OUTPUT## Finding the target end site if it's
within this exon.

        row[,ncol(row)+1] <- 0 # Make a new column

        names(row)[ncol(row)] <- region_end # Rename this
new column

        row[,ncol(row)]      <-      ((row[[exon_start_var]]      +
unassigned_nt) -1) # Insert the value for this new column

        unassigned_nt      <-      (unassigned_nt      -      (
row[[region_end]] - row[[region_start]] +1)) # Updating the
counter to show how many unassigned nucleotides are left.

    } # Closing bracket for ifelse statement 3

```

```

    } # Closing bracket for for while loop 2

} else { # Else condition for ifelse statement 2

    ##OUTPUT## Finding the target end site if it's within
this exon.

    row[,ncol(row)+1] <- 0 # Make a new column

    names(row)[ncol(row)] <- region_end # Rename this new
column

    row[,ncol(row)] <- ((row[[region_start]] + row$Tlength)
-1) # Insert the value for this new column

    unassigned_nt <- (unassigned_nt - ( row[[region_end]]
- row[[region_start]] +1)) # Updating the counter to show how
many unassigned nucleotides are left.

} # Closing bracket for ifelse statement 2

} else { # Else condition for ifelse statement 1

    exon_no_counter <- exon_no_counter + 1 # Variable for the
exon number we are currently looking at

    next_exon_counter <- next_exon_counter + 1 # Variable for
the next exon number from the one we are currently looking at

    exon_length_var <-
paste0("length_exon_",exon_no_counter) # Generating the text to
label the exon length variable.

```

```

        previous_transcript_loc_counter      <-
current_transcript_loc_counter

        current_transcript_loc_counter      <-
current_transcript_loc_counter - row[[exon_length_var]] #
Numeric variable for each TS-E1-E2 etc. value

    } # Closing bracket for ifelse statement 1

} # Closing bracket for while statement

return(row)

} # Closing bracket for function

positive_list_total <- list()

for (i in 1:nrow(positive_targetexon_info)){
    result
    positive_translatingloci(positive_targetexon_info[i, ])
    positive_list_total <- append(positive_list_total,
list(result))
}

positive_loci <- bind_rows(positive_list_total, .id = "source")
positive_loci$strand <- "+"

#####

```

```

### Translating loci on the negative strand ###
#####

# Tidying up the data #
#####

colnames(negative_exon_info)[colnames(negative_exon_info) ==
"V1"] <- "chr"

negative_exon_info <- subset(negative_exon_info, select = -
c(V2,V3,V6,V7,V8) )

colnames(negative_exon_info)[colnames(negative_exon_info) ==
"V4"] <- "start"

colnames(negative_exon_info)[colnames(negative_exon_info) ==
"V5"] <- "end"

colnames(negative_exon_info)[colnames(negative_exon_info) ==
"V9"] <- "transcript_id"

colnames(negative_exon_info)[colnames(negative_exon_info) ==
"V10"] <- "id"

colnames(negative_exon_info)[colnames(negative_exon_info) ==
"V11"] <- "length"

# Converting the exon information into wide format #
#####

negative_exon_info$transcript_id <-
str_extract(negative_exon_info$transcript_id,
"TraesCS\\w+\\.\\.\\d+") # This line extracts the Traes ID from
the longer notes field. This will need to be edited depending
on the format of the notes field.

```

```

negative_exon_info <- negative_exon_info %>%
  mutate(chr_transcript_id = paste0(chr, ",", transcript_id))
negative_exon_info_wide <- negative_exon_info %>%
  pivot_wider(
    id_cols = chr_transcript_id,
    names_from = id,
    values_from = c(start, end, length)
  )

negative_exon_info_wide <-
  separate(negative_exon_info_wide, col=chr_transcript_id, into=c(
    'chr', 'transcript_id'), sep=',')

# Merging transcript and target info #
#####

negative_targetexon_info <-
  merge(transcript_loci, negative_exon_info_wide, by="transcript_id",
    all.x = FALSE, all.y = FALSE)

# Translating the loci #
#####

negative_translatingloci <- function(row) {

  # Splitting the dataframe into rows, so each row is worked on
  independently

  row <- as.data.frame(row)

```

```
# Setting up the counter for unassigned target nucleotides.
The number refers to the number of unassigned nucleotides
remaining for the row.
```

```
unassigned_nt <- row$Tlength
```

```
# Generating numeric variables I need for the for the function
```

```
exon_no_counter <- 1 # Variable for the exon number we are
currently looking at
```

```
next_exon_counter <- 2 # Variable for the next exon number
from the one we are currently looking at
```

```
region_no_counter <- 1 # Variable for the target region we
are currently assigning
```

```
current_transcript_loc_counter <- row$TS - row$length_exon_1
# Numeric variable for each TS-E1-E2 etc. value
```

```
previous_transcript_loc_counter <- row$TS # Numeric variable
for each TS-E1-E2 etc. value, but from the previous exon. I.e
if we are on exon 3, this will be TS-E1-E2.
```

```
while (unassigned_nt != 0) { # The script will run until the
number of unassigned nucleotides is zero.
```

```
# Generating character variables I need for for loop 1
```

```
region_start <- paste0("R", region_no_counter, "_start") #
Generating the text to label the region start variable.
```

```
region_end <- paste0("R", region_no_counter, "_end") #
Generating the text to label the region end variable.
```

```
exon_start_var <- paste0("start_exon_", exon_no_counter) #
Generating the text to label the exon start variable.
```

```
    exon_end_var <- paste0("end_exon_", exon_no_counter) #  
Generating the text to label the exon end variable.
```

```
    exon_length_var <- paste0("length_exon_",exon_no_counter)  
# Generating the text to label the exon length variable.
```

```
    next_exon_start_var <- paste0("start_exon_",  
next_exon_counter) # Generating the text to label the exon start  
variable.
```

```
    next_exon_length_var <-  
paste0("length_exon_",next_exon_counter) # Generating the text  
to label the next exon length variable.
```

```
    if (current_transcript_loc_counter <0){ # Does the target  
start in this transcript? True if it does. Ifelse statement 1.
```

```
    ##OUTPUT## Calculating target start site in this exon.
```

```
    row[,ncol(row)+1] <- 0 # Make a new column
```

```
    names(row)[ncol(row)] <- region_end # Rename this new  
column
```

```
    row[,ncol(row)] <- ((row[[exon_end_var]] -  
previous_transcript_loc_counter) + 1) # Insert the value for  
this new column
```

```
    if (current_transcript_loc_counter > ((row$Tlength -1)*-  
1)){ # Is TS - E(n) > the negative miRNA length? This is true  
if the target site extends beyond this exon. Ifelse statement  
2.
```

```
    ##OUTPUT## Calculating the end of this region (the end  
of the exon)
```

```
    row[,ncol(row)+1] <- 0 # Make a new column
```

```

names(row)[ncol(row)] <- region_start # Rename this new
column

row[,ncol(row)] <- row[[exon_start_var]] # Insert the
value for this new column (just the end of the exon).

unassigned_nt <- (unassigned_nt - ( row[[region_end]]
- row[[region_start]] +1)) # Updating the counter to show how
many unassigned nucleotides are left.

while (unassigned_nt != 0) { # While loop 2

# Updating numeric variables for this for loop

exon_no_counter <- exon_no_counter + 1 # Increasing
the counter by one so the next loop will look at the next exon.

region_no_counter <- region_no_counter + 1 #
Increasing the counter by one so the next loop will label the
region correctly.

# Update character variables for this for loop

region_start <- paste0("R", region_no_counter,
"_start") # Generating the text to label the region start
variable.

region_end <- paste0("R", region_no_counter, "_end")
# Generating the text to label the region end variable.

exon_start_var <- paste0("start_exon_",
exon_no_counter) # Generating the text to label the exon start
variable.

exon_end_var <- paste0("end_exon_", exon_no_counter)
# Generating the text to label the exon end variable.

```

```

        exon_length_var                                     <-
paste0("length_exon_",exon_no_counter) # Generating the text to
label the exon length variable.

        ##OUTPUT## Calculating the start of this region (the
start of the exon).

        row[,ncol(row)+1] <- 0 # Make a new column.

        names(row)[ncol(row)] <- region_end # Rename this new
column.

        row[,ncol(row)] <- row[[exon_end_var]] # Insert the
value for this new column (just the start of the exon).

        if (unassigned_nt > row[[exon_length_var]]) { # Are
the remaining nucleotides to be assigned more than the length
of this exon? This is true if the target site extends beyond
this exon. Ifelse statement 3.

        ##OUTPUT## Calculating the end of this region (the
end of the exon).

        row[,ncol(row)+1] <- 0 # Make a new column

        names(row)[ncol(row)] <- region_start # Rename this
new column

        row[,ncol(row)] <- row[[exon_start_var]] # Insert
the value for this new column (just the end of the exon).

        unassigned_nt <- (unassigned_nt - (
row[[region_end]] - row[[region_start]] +1)) # Updating the
counter to show how many unassigned nucleotides are left.

```

```

    } else { # Else condition for ifelse statement 3 -
for if the region ends in this exon.

    ##OUTPUT## Finding the target end site if it's
within this exon.

    row[,ncol(row)+1] <- 0 # Make a new column

    names(row)[ncol(row)] <- region_start # Rename this
new column

    row[,ncol(row)] <- ((row[[exon_end_var]] -
unassigned_nt) +1) # Insert the value for this new column

    unassigned_nt <- (unassigned_nt - (
row[[region_end]] - row[[region_start]] +1)) # Updating the
counter to show how many unassigned nucleotides are left.

    } # Closing bracket for ifelse statement 3

} # Closing bracket for for while loop 2

} else { # Else condition for ifelse statement 2

    ##OUTPUT## Finding the target end site if it's within
this exon.

    row[,ncol(row)+1] <- 0 # Make a new column

    names(row)[ncol(row)] <- region_start # Rename this new
column

    row[,ncol(row)] <- ((row[[region_end]] - row$Tlength)
+1) # Insert the value for this new column

```

```

        unassigned_nt <- (unassigned_nt - ( row[[region_end]]
- row[[region_start]] +1)) # Updating the counter to show how
many unassigned nucleotides are left.

    } # Closing bracket for ifelse statement 2

} else { # Else condition for ifelse statement 1

    exon_no_counter <- exon_no_counter + 1 # Variable for the
exon number we are currently looking at

    next_exon_counter <- next_exon_counter + 1 # Variable for
the next exon number from the one we are currently looking at

    exon_length_var <-
paste0("length_exon_",exon_no_counter) # Generating the text to
label the exon length variable.

    previous_transcript_loc_counter <-
current_transcript_loc_counter

    current_transcript_loc_counter <-
current_transcript_loc_counter - row[[exon_length_var]] #
Numeric variable for each TS-E1-E2 etc. value

} # Closing bracket for ifelse statement 1

} # Closing bracket for while statement

return(row)

} # Closing bracket for function

```

```

negative_list_total <- list()

for (i in 1:nrow(negative_targetexon_info)){

  result <-
negative_translatingloci(negative_targetexon_info[i, ])

  negative_list_total <- append(negative_list_total,
list(result))

}

negative_loci <- bind_rows(negative_list_total, .id = "source")
negative_loci$strand <- "-"

#####

### Combining the loci from positive and negative strands ###

#####

genomic_loci <- rbind.fill(negative_loci, positive_loci)

df_for_export <- genomic_loci[, c("locus_id", "chr", "strand",
grep("^R", names(genomic_loci), value = TRUE))]

write.table(df_for_export, file = "genomic_loci.txt", row.names
= FALSE, col.names = TRUE, quote = FALSE)

```

## Glossary

EMS	Ethyl methanesulfonate
TILLING	Targeting induced local lesions in genomes
SNP	Single nucleotide polymorphism
Indels	Insertion-deletions
CNV	Copy number variant
SAM	Shoot apical meristem
miRNA	microRNA
Pol II	RNA Polymerase II
Pri-miRNA	Primary miRNA
Pre-miRNA	Precursor miRNA
PARE-Seq	Parallel analysis of RNA ends sequencing
ELISA	Enzyme linked immunosorbent assay
sRNA-Seq	smallRNA-Seq
NGS	Next generation sequencing
RRSV	Rice ragged stunt virus
CRISPR	Clustered regularly interspaced short palindromic repeats
IWGSC	International wheat genome sequencing consortium
RPM	Reads per million
tRNA	Transfer RNA
rRNA	Ribosomal RNA
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
PCA	Principal component analysis
TGW	Thousand grain weight
JIC	John Innes Centre
BLUE	Best linear unbiased estimate
ANOVA	Analysis of variance
BLAST	Basic Local Alignment Search Tool
E-value	Expectation value
UTR	Untranslated region
MITE	miniature inverted-repeat transposable element
sPARTA	small RNA-PARE target analyser
GM	Genetically modified
GMO	Genetically modified organism

PAM	Protospacer adjacent motif
ORF	Open reading frame
MERFISH	Multiplexed error-robust fluorescence <i>in situ</i> hybridization
JTT	Jones-Taylor-Thornton
TPM	Transcripts per million
PFA	Paraformaldehyde
PBS	Phosphate-buffered saline
OCT	Optimal cutting temperature compound
LB	Lennox broth
SOC	Super optimal broth with catabolite repression medium
ONT	Oxford nanopore technology
LMM	Linear mixed model
CLD	Compact letter display
CDS	Coding sequence
dsDNA	Double-stranded DNA
UMAP	Uniform manifold approximation and projection
RT-qPCR	Real-Time Quantitative Reverse Transcription PCR
CV	Coefficient of variation
LNA	Locked nucleic acid
smFISH	Single molecule fluorescence <i>in situ</i> hybridization
NIL	Near isogenic line
BiFC	Bimolecular fluorescence complementation
DPA	Days post anthesis
CUT&RUN	Cleavage under targets & release using nuclease
ChIP-Seq	Chromatin immunoprecipitation sequencing

## References

- 1 Shewry, P. R. Wheat. *Journal of Experimental Botany* **60**, 1537-1553 (2009).  
<https://doi.org/10.1093/jxb/erp058>
- 2 Food and Agriculture Organisation of the United Nations. *FAOSTAT*,  
<http://www.fao.org/faostat/en/#home> (2024).
- 3 Tao, Z. *et al.* Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. *The Plant Journal* **70**, 549-561 (2012).  
<https://doi.org/10.1111/j.1365-313x.2012.04919.x>
- 4 United Nations. *World Population Prospects 2019: Highlights*. (2019).
- 5 Hawkesford, M. J. *et al.* Prospects of doubling global wheat yields. *Food and Energy Security* **2**, 34-48 (2013). <https://doi.org/10.1002/fes3.15>
- 6 FAO. *Crop Prospects and Food Situation #1, March 2021: Quarterly Global Report*. 1 edn, (FAO, 2021).
- 7 FAO, IFAD, UNICEF, WFP & WHO. *The State of Food Security and Nutrition in the World 2024: Financing to end hunger, food insecurity and malnutrition in all its forms*. (FAO, IFAD, UNICEF, WFP, WHO, 2024).
- 8 Bentley, A. Broken bread — avert global wheat crisis caused by invasion of Ukraine. *Nature* **603**, 551 (2022). <https://doi.org/10.1038/d41586-022-00789-x>
- 9 Bentley, A. R. *et al.* Near- to long-term measures to stabilize global wheat supplies and food security. *Nature Food* **3**, 483-486 (2022). <https://doi.org/10.1038/s43016-022-00559-y>
- 10 European Council. *How the Russian invasion of Ukraine has further aggravated the global food crisis*, <https://www.consilium.europa.eu/en/infographics/how-the-russian-invasion-of-ukraine-has-further-aggravated-the-global-food-crisis/#0> (2025).
- 11 European Council. *Food security and affordability*,  
<https://www.consilium.europa.eu/en/policies/food-security-and-affordability/>  
(2025).
- 12 Shiferaw, B. *et al.* Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Security* **5**, 291-317 (2013).  
<https://doi.org/10.1007/s12571-013-0263-y>
- 13 Department for Environment, Food and Rural Affairs, Department of Agriculture, Environment and Rural Affairs (Northern Ireland), Welsh Government, Knowledge and Analytical Services & The Scottish Government, Rural and Environment Science and Analytical Service. *Agriculture in the United Kingdom 2023*. (2024).
- 14 Walkowiak, S. *et al.* Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**, 277-283 (2020). <https://doi.org/10.1038/s41586-020-2961-x>

- 15 The International Wheat Genome Sequencing Consortium (IWGSC) *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018). <https://doi.org/10.1126/science.aar7191>
- 16 Adamski, N. *et al.* *Wheat Development*, [https://www.wheat-training.com/wp-content/uploads/Wheat\\_growth/pdfs/Wheat\\_development\\_pdf.pdf](https://www.wheat-training.com/wp-content/uploads/Wheat_growth/pdfs/Wheat_development_pdf.pdf)
- 17 International Rice Genome Sequencing Project & Sasaki, T. The map-based sequence of the rice genome. *Nature* **436**, 793-800 (2005). <https://doi.org/10.1038/nature03895>
- 18 Schnable, P. S. *et al.* The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* **326**, 1112-1115 (2009). <https://doi.org/10.1126/science.1178534>
- 19 The International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, 1251788 (2014). <https://doi.org/10.1126/science.1251788>
- 20 Marcussen, T. *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, 1250092 (2014). <https://doi.org/10.1126/science.1250092>
- 21 Dubcovsky, J. & Dvorak, J. Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. *Science* **316**, 1862-1866 (2007). <https://doi.org/10.1126/science.1143986>
- 22 Ramírez-González, R. H. *et al.* The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018). <https://doi.org/10.1126/science.aar6089>
- 23 Borrill, P., Adamski, N. & Uauy, C. Genomics as the key to unlocking the polyploid potential of wheat. *New Phytologist* **208**, 1008-1022 (2015). <https://doi.org/10.1111/nph.13533>
- 24 Borrill, P., Harrington, S. A. & Uauy, C. Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. *The Plant Journal* **97**, 56-72 (2019). <https://doi.org/10.1111/tpj.14150>
- 25 Slade, A. J. *et al.* Development of high amylose wheat through TILLING. *BMC Plant Biology* **12**, 69 (2012). <https://doi.org/10.1186/1471-2229-12-69>
- 26 Acevedo-Garcia, J. *et al.* *mlo*-based powdery mildew resistance in hexaploid bread wheat generated by a non-transgenic TILLING approach. *Plant Biotechnology Journal* **15**, 367-378 (2017). <https://doi.org/10.1111/pbi.12631>
- 27 Rosyara, U. *et al.* Genetic Contribution of Synthetic Hexaploid Wheat to CIMMYT's Spring Bread Wheat Breeding Germplasm. *Scientific Reports* **9**, 12355 (2019). <https://doi.org/10.1038/s41598-019-47936-5>
- 28 Watson, A. *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**, 23-29 (2018). <https://doi.org/10.1038/s41477-017-0083-8>

- 29 Borrill, P., Ramirez-Gonzalez, R. & Uauy, C. expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform *Plant Physiology* **170**, 2172-2186 (2016). <https://doi.org/10.1104/pp.15.01667>
- 30 Adamski, N. M. *et al.* A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. *eLife* **9**, e55646 (2020). <https://doi.org/10.7554/elife.55646>
- 31 Krasileva, K. V. *et al.* Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences* **114**, E913-E921 (2017). <https://doi.org/10.1073/pnas.1619268114>
- 32 Debernardi, J. M., Lin, H., Chuck, G., Faris, J. D. & Dubcovsky, J. microRNA172 plays a crucial role in wheat spike morphogenesis and grain threshability. *Development* **144**, 1966-1975 (2017). <https://doi.org/10.1242/dev.146399>
- 33 Simmonds, J. *et al.* A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theoretical and Applied Genetics* **129**, 1099-1112 (2016). <https://doi.org/10.1007/s00122-016-2686-2>
- 34 Cheng, S. *et al.* Harnessing landrace diversity empowers wheat breeding. *Nature* **632**, 823-831 (2024). <https://doi.org/10.1038/s41586-024-07682-9>
- 35 Debernardi, J. M. *et al.* A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nature Biotechnology* **38**, 1274–1279 (2020). <https://doi.org/10.1038/s41587-020-0703-0>
- 36 Borrill, P. Blurring the boundaries between cereal crops and model plants. *New Phytologist* **228**, 1721-1727 (2020). <https://doi.org/10.1111/nph.16229>
- 37 McSteen, P. & Leyser, O. Shoot Branching. *Annu. Rev. Plant Biol.* **56**, 353-374 (2005). <https://doi.org/10.1146/annurev.arplant.56.032604.144122>
- 38 Pautler, M., Tanaka, W., Hirano, H.-Y. & Jackson, D. Grass Meristems I: Shoot Apical Meristem Maintenance, Axillary Meristem Determinacy and the Floral Transition. *Plant and Cell Physiology* **54**, 302-312 (2013). <https://doi.org/10.1093/pcp/pct025>
- 39 Brinton, J. & Uauy, C. A reductionist approach to dissecting grain weight and yield in wheat. *J Integr Plant Biol* **61**, 337-358 (2019). <https://doi.org/10.1111/jipb.12741>
- 40 Zhong, J. *et al.* *INTERMEDIUM-M* encodes an *HvAP2L-H5* ortholog and is required for inflorescence indeterminacy and spikelet determinacy in barley. *Proceedings of the National Academy of Sciences* **118**, e2011779118 (2021). <https://doi.org/10.1073/pnas.2011779118>

- 41 Waddington, S. R., Cartwright, P. M. & Wall, P. C. A Quantitative Scale of Spike Initial and Pistil Development in Barley and Wheat. *Annals of Botany* **51**, 119-130 (1983). <https://doi.org/10.1093/oxfordjournals.aob.a086434>
- 42 Debernardi, J. M., Greenwood, J. R., Finnegan, E. J., Jernstedt, J. & Dubcovsky, J. *APETALA 2*-like genes *AP2L2* and *Q* specify lemma identity and axillary floral meristem development in wheat. *The Plant Journal* **101**, 171-187 (2020). <https://doi.org/10.1111/tpj.14528>
- 43 Backhaus, A. E. *et al.* High expression of the MADS-box gene *VRT2* increases the number of rudimentary basal spikelets in wheat. *Plant Physiology* **189**, 1536–1552 (2022). <https://doi.org/10.1093/plphys/kiac156>
- 44 Li, K. *et al.* Interactions between *SQUAMOSA* and *SHORT VEGETATIVE PHASE* MADS-box proteins regulate meristem transitions during wheat spike development. *The Plant Cell* **33**, 3621-3644 (2021). <https://doi.org/10.1093/plcell/koab243>
- 45 Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31-37 (1991). <https://doi.org/10.1038/353031a0>
- 46 Theißen, G. Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology* **4**, 75-85 (2001). [https://doi.org/10.1016/s1369-5266\(00\)00139-4](https://doi.org/10.1016/s1369-5266(00)00139-4)
- 47 Angenent, G. C. & Colombo, L. Molecular control of ovule development. *Trends in Plant Science* **1**, 228-232 (1996). [https://doi.org/10.1016/1360-1385\(96\)86900-7](https://doi.org/10.1016/1360-1385(96)86900-7)
- 48 Irish, V. The ABC model of floral development. *Current Biology* **27**, R887-R890 (2017). <https://doi.org/10.1016/j.cub.2017.03.045>
- 49 Murai, K. Homeotic Genes and the ABCDE Model for Floral Organ Formation in Wheat. *Plants (Basel)* **2**, 379-395 (2013). <https://doi.org/10.3390/plants2030379>
- 50 Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genes directing flower development in *Arabidopsis*. *The Plant Cell* **1**, 37-52 (1989). <https://doi.org/10.1105/tpc.1.1.37>
- 51 Lombardo, F. & Yoshida, H. Interpreting lemma and palea homologies: a point of view from rice floral mutants. *Frontiers in Plant Science* **6** (2015). <https://doi.org/10.3389/fpls.2015.00061>
- 52 Callens, C., Tucker, M. R., Zhang, D. & Wilson, Z. A. Dissecting the role of MADS-box genes in monocot floral development and diversity. *Journal of Experimental Botany* **69**, 2435-2459 (2018). <https://doi.org/10.1093/jxb/ery086>
- 53 Ali, Z. *et al.* Genetic and Molecular Control of Floral Organ Identity in Cereals. *Int J Mol Sci* **20**, 2743 (2019). <https://doi.org/10.3390/ijms20112743>
- 54 Kellogg, E. A. *Flowering Plants. Monocots: Poaceae*. (Springer Cham, 2015).

- 55 Yoshida, H. Is the lodicule a petal: Molecular evidence? *Plant Science* **184**, 121-128 (2012). <https://doi.org/10.1016/j.plantsci.2011.12.016>
- 56 Adamski, N. M. *et al.* Ectopic expression of *Triticum polonicum VRT-A2* underlies elongated glumes and grains in hexaploid wheat in a dosage-dependent manner. *The Plant Cell* **33**, 2296–2319 (2021). <https://doi.org/10.1093/plcell/koab119>
- 57 Shoemith, J. R. *et al.* APETALA2 functions as a temporal factor together with BLADE-ON-PETIOLE2 and MADS29 to control flower and grain development in barley. *Development* **148**, dev194894 (2021). <https://doi.org/10.1242/dev.194894>
- 58 Luo, Y., Guo, Z. & Li, L. Evolutionary conservation of microRNA regulatory programs in plant flower development. *Developmental Biology* **380**, 133-144 (2013). <https://doi.org/10.1016/j.ydbio.2013.05.009>
- 59 D’Ario, M., Griffiths-Jones, S. & Kim, M. Small RNAs: Big Impact on Plant Development. *Trends in Plant Science* **22**, 1056-1068 (2017). <https://doi.org/10.1016/j.tplants.2017.09.009>
- 60 Yamaguchi, A. *et al.* The MicroRNA-Regulated SBP-Box Transcription Factor SPL3 Is a Direct Upstream Activator of *LEAFY*, *FRUITFULL*, and *APETALA1*. *Developmental Cell* **17**, 268-278 (2009). <https://doi.org/10.1016/j.devcel.2009.06.007>
- 61 Chen, X. A MicroRNA as a Translational Repressor of *APETALA2* in *Arabidopsis* Flower Development. *Science* **303**, 2022-2025 (2004). <https://doi.org/10.1126/science.1088060>
- 62 Yant, L. *et al.* Orchestration of the Floral Transition and Floral Development in *Arabidopsis* by the Bifunctional Transcription Factor APETALA2. *The Plant Cell* **22**, 2156-2170 (2010). <https://doi.org/10.1105/tpc.110.075606>
- 63 Aida, M. & Tasaka, M. Morphogenesis and Patterning at the Organ Boundaries in the Higher Plant Shoot Apex. *Plant Molecular Biology* **60**, 915-928 (2006). <https://doi.org/10.1007/s11103-005-2760-7>
- 64 Alonso-Peral, M. M. *et al.* The MicroRNA159-Regulated *GAMYB-like* Genes Inhibit Growth and Promote Programmed Cell Death in *Arabidopsis* *Plant Physiology* **154**, 757-771 (2010). <https://doi.org/10.1104/pp.110.160630>
- 65 Wang, Y. *et al.* *TamiR159* directed wheat *TaGAMYB* cleavage and its involvement in anther development and heat response. *PLoS One* **7**, e48445 (2012). <https://doi.org/10.1371/journal.pone.0048445>
- 66 Greenwood, J. R., Finnegan, E. J., Watanabe, N., Trevaskis, B. & Swain, S. M. New alleles of the wheat domestication gene *Q* reveal multiple roles in growth and reproductive development. *Development* **144**, 1959-1965 (2017). <https://doi.org/10.1242/dev.146407>

- 67 Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes & Development* **16**, 1616-1626 (2002).  
<https://doi.org/10.1101/gad.1004402>
- 68 Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854 (1993). [https://doi.org/10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y)
- 69 Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950-952 (1999).  
<https://doi.org/10.1126/science.286.5441.950>
- 70 Rogers, K. & Chen, X. Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. *The Plant Cell* **25**, 2383-2399 (2013).  
<https://doi.org/10.1105/tpc.113.113159>
- 71 Wang, J., Mei, J. & Ren, G. Plant microRNAs: biogenesis, homeostasis, and degradation. *Frontiers in Plant Science* **10** (2019).  
<https://doi.org/10.3389/fpls.2019.00360>
- 72 Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal* **23**, 4051-4060 (2004). <https://doi.org/10.1038/sj.emboj.7600385>
- 73 Kurihara, Y. & Watanabe, Y. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences* **101**, 12753-12758 (2004). <https://doi.org/10.1073/pnas.0403115101>
- 74 Axtell, M. J. & Meyers, B. C. Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. *The Plant Cell* **30**, 272-284 (2018).  
<https://doi.org/10.1105/tpc.17.00851>
- 75 Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. Methylation Protects miRNAs and siRNAs from a 3'-End Uridylation Activity in *Arabidopsis*. *Current Biology* **15**, 1501-1507 (2005). <https://doi.org/10.1016/j.cub.2005.07.029>
- 76 Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-935 (2005). <https://doi.org/10.1126/science.1107130>
- 77 Moran, Y., Agron, M., Praher, D. & Technau, U. The evolutionary origin of plant and animal microRNAs. *Nature Ecology & Evolution* **1**, 0027 (2017).  
<https://doi.org/10.1038/s41559-016-0027>
- 78 Zhang, X. *et al.* *Arabidopsis* Argonaute 2 Regulates Innate Immunity via miRNA393\*-Mediated Silencing of a Golgi-Localized SNARE Gene, *MEMB12*. *Molecular Cell* **42**, 356-366 (2011). <https://doi.org/10.1016/j.molcel.2011.04.010>
- 79 Hu, W., Wang, T., Yue, E., Zheng, S. & Xu, J.-H. Flexible microRNA arm selection in rice. *Biochemical and Biophysical Research Communications*. **447**, 526-530 (2014). <https://doi.org/10.1016/j.bbrc.2014.04.036>

- 80 Budak, H., Bulut, R., Kantar, M. & Alptekin, B. MicroRNA nomenclature and the need for a revised naming prescription. *Briefings in Functional Genomics* **15**, 65-71 (2016). <https://doi.org/10.1093/bfgp/elv026>
- 81 Meyers, B. C. *et al.* Criteria for Annotation of Plant MicroRNAs. *The Plant Cell* **20**, 3186-3190 (2008). <https://doi.org/10.1105/tpc.108.064311>
- 82 Ó'Maoiléidigh, D. S. *et al.* Systematic analyses of the *MIR172* family members of *Arabidopsis* define their distinct roles in regulation of *APETALA2* during floral transition. *PLOS Biology* **19**, e3001043 (2021). <https://doi.org/10.1371/journal.pbio.3001043>
- 83 Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics* **12**, 99-110 (2011). <https://doi.org/10.1038/nrg2936>
- 84 Yu, Y., Jia, T. & Chen, X. The 'how' and 'where' of plant microRNAs. *New Phytologist* **216**, 1002-1017 (2017). <https://doi.org/10.1111/nph.14834>
- 85 Hutvagner, G. & Zamore, P. D. A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science* **297**, 2056-2060 (2002). <https://doi.org/10.1126/science.1073827>
- 86 Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their Regulatory Roles in Plants. *Annual Review of Plant Biology* **57**, 19-53 (2006). <https://doi.org/10.1146/annurev.arplant.57.032905.105218>
- 87 Yang, M. *et al.* Intact RNA structure reveals mRNA structure-mediated regulation of miRNA cleavage *in vivo*. *Nucleic Acids Research* **48**, 8767-8781 (2020). <https://doi.org/10.1093/nar/gkaa577>
- 88 Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science* **305**, 1434-1437 (2004). <https://doi.org/10.1126/science.1102514>
- 89 Souret, F. F., Kastenmayer, J. P. & Green, P. J. AtXRN4 Degrades mRNA in *Arabidopsis* and Its Substrates Include Selected miRNA Targets. *Molecular Cell* **15**, 173-183 (2004). <https://doi.org/10.1016/j.molcel.2004.06.006>
- 90 Zhang, Z. *et al.* RISC-interacting clearing 3'-5' exoribonucleases (RICEs) degrade uridylated cleavage fragments to maintain functional RISC in *Arabidopsis thaliana*. *eLife* **6**, e24466 (2017). <https://doi.org/10.7554/eLife.24466>
- 91 Iwakawa, H.-o. & Tomari, Y. Molecular insights into microRNA-mediated translational repression in plants. *Molecular Cell* **52**, 591-601 (2013). <https://doi.org/10.1016/j.molcel.2013.10.033>
- 92 CIMMYT. *Increasing Yield Potential in Wheat: Breaking the Barriers* (eds M. P. Reynolds, S. Rajaram, & A. McNab) (CIMMYT, 1996).

- 93 Calderini, D. F. *et al.* Overcoming the trade-off between grain weight and number in wheat by the ectopic expression of expansin in developing seeds leads to increased yield potential. *New Phytologist* **230**, 629-640 (2021).  
<https://doi.org/10.1111/nph.17048>
- 94 Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Research* **47**, D155-D162 (2019).  
<https://doi.org/10.1093/nar/gky1141>
- 95 Dixon, L. E. *et al.* MicroRNA-resistant alleles of *HOMEODOMAIN-2* modify inflorescence branching and increase grain protein content of wheat. *Science Advances* **8**, eabn5907 (2022). <https://doi.org/doi:10.1126/sciadv.abn5907>
- 96 Chen, C. *et al.* sRNAanno—a database repository of uniformly annotated small RNAs in plants. *Horticulture Research* **8**, 45 (2021). <https://doi.org/10.1038/s41438-021-00480-8>
- 97 Zhu, T. *et al.* Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal* **107**, 303-314 (2021).  
<https://doi.org/10.1111/tpj.15289>
- 98 The International Wheat Genome Sequencing Consortium. *IWGSC RefSeq v1.1 annotation is publicly available at URGI*,  
<https://www.wheatgenome.org/resources/annotations/iwgsc-refseq-v1.1-annotation2> (2021).
- 99 Feng, N. *et al.* Transcriptome Profiling of Wheat Inflorescence Development from Spikelet Initiation to Floral Patterning Identified Stage-Specific Regulatory Genes *Plant Physiology* **174**, 1779-1794 (2017).  
<https://doi.org/10.1104/pp.17.00310>
- 100 Li, J. *et al.* Gene Expression Profiles and microRNA Regulation Networks in Tiller Primordia, Stem Tips, and Young Spikes of Wheat Guomai 301. *Genes* **10** (2019).  
<https://doi.org/10.3390/genes10090686>
- 101 Zhang, C. *et al.* Suppression of Jasmonic Acid-Mediated Defense by Viral-Inducible MicroRNA319 Facilitates Virus Infection in Rice. *Molecular Plant* **9**, 1302-1314 (2016). <https://doi.org/10.1016/j.molp.2016.06.014>
- 102 Gupta, A., Hua, L., Zhang, Z., Yang, B. & Li, W. CRISPR-induced miRNA156-recognition element mutations in *TaSPL13* improve multiple agronomic traits in wheat. *Plant Biotechnology Journal* **21**, 536-548 (2023).  
<https://doi.org/10.1111/pbi.13969>
- 103 Tang, J. & Chu, C. MicroRNAs in crop improvement: fine-tuners for complex traits. *Nature Plants* **3**, 17077 (2017). <https://doi.org/10.1038/nplants.2017.77>

- 104 Djami-Tchatchou, A. T., Sanan-Mishra, N., Ntushelo, K. & Dubery, I. A. Functional Roles of microRNAs in Agronomically Important Plants—Potential as Targets for Crop Improvement and Protection. *Frontiers in Plant Science* **8** (2017). <https://doi.org/10.3389/fpls.2017.00378>
- 105 Wu, J. *et al.* Overexpression of *zmm28* increases maize grain yield in the field. *Proceedings of the National Academy of Sciences* **116**, 23850-23858 (2019). <https://doi.org/10.1073/pnas.1902593116>
- 106 Simmonds, J., Faci, I., Buffagni, V. & Uauy, C. *How to grow wheat in controlled and semi-controlled environment*, [dx.doi.org/10.17504/protocols.io.6qpvr37jbvmk/v1](https://doi.org/10.17504/protocols.io.6qpvr37jbvmk/v1) (2024).
- 107 Linnaeus, C. *Species plantarum: exhibentes plantas rite cognitatas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas.* (Laurentii Salvii, 1753).
- 108 Liu, J. *et al.* Ectopic expression of *VRT-A2* underlies the origin of *Triticum polonicum* and *Triticum petropavlovskyi* with long outer glumes and grains. *Molecular Plant* **14**, 1472-1488 (2021). <https://doi.org/10.1016/j.molp.2021.05.021>
- 109 ZYMO RESEARCH. *Direct-zol™ RNA Microprep*, [https://files.zymoresearch.com/protocols/\\_r2060\\_r2061\\_r2062\\_r2063\\_direct-zol\\_rna\\_microprep.pdf](https://files.zymoresearch.com/protocols/_r2060_r2061_r2062_r2063_direct-zol_rna_microprep.pdf) (2023).
- 110 Andrews, S. *FastQC*, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
- 111 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014). <https://doi.org/10.1093/bioinformatics/btu170>
- 112 Garg, V. & Varshney, R. K. in *Plant Bioinformatics: Methods and Protocols* (ed David Edwards) 497-509 (Humana, 2022).
- 113 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10-12 (2011). <https://doi.org/10.14806/ej.17.1.200>
- 114 Alaux, M. *et al.* Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology* **19**, 111 (2018). <https://doi.org/10.1186/s13059-018-1491-4>
- 115 Hammond, R. K., Gupta, P., Patel, P. & Meyers, B. C. miRador: a fast and precise tool for the prediction of plant miRNAs. *Plant Physiol* **191**, 894-903 (2023). <https://doi.org/10.1093/plphys/kiac538>
- 116 Axtell, M. J. ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA* **19**, 740-751 (2013). <https://doi.org/10.1261/rna.035279.112>

- 117 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009). <https://doi.org/10.1186/gb-2009-10-3-r25>
- 118 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). <https://doi.org/10.1093/bioinformatics/btp352>
- 119 Kuang, Z., Wang, Y., Li, L. & Yang, X. miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. *Bioinformatics* **35**, 2521-2522 (2019). <https://doi.org/10.1093/bioinformatics/bty972>
- 120 Hannon Laboratory. *FASTX-Toolkit*, [http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html) (2010).
- 121 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**, 26 (2011). <https://doi.org/10.1186/1748-7188-6-26>
- 122 Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* **49**, D192-D200 (2021). <https://doi.org/10.1093/nar/gkaa1047>
- 123 R Core Team. *R: A Language and Environment for Statistical Computing*, <https://www.R-project.org/> (2023).
- 124 Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-Directed Phasing during *Trans*-Acting siRNA Biogenesis in Plants. *Cell* **121**, 207-221 (2005). <https://doi.org/10.1016/j.cell.2005.04.004>
- 125 Fahlgren, N. *et al.* High-Throughput Sequencing of *Arabidopsis* microRNAs: Evidence for Frequent Birth and Death of *MIRNA* Genes. *PLOS ONE* **2**, e219 (2007). <https://doi.org/10.1371/journal.pone.0000219>
- 126 Fahlgren, N. & Carrington, J. C. in *Plant MicroRNAs: Methods and Protocols* Vol. 592 *Methods in Molecular Biology* (eds Blake Meyers & Pamela Green) 51-57 (Humana Press, 2010).
- 127 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021). <https://doi.org/10.1093/gigascience/giab008>
- 128 Harrison, P. W. *et al.* Ensembl 2024. *Nucleic Acids Research* **52**, D891-D899 (2024). <https://doi.org/10.1093/nar/gkad1049>
- 129 van Rossum, B.-J. *statgenSTA: Single Trial Analysis (STA) of Field Trials*, <https://biometris.github.io/statgenSTA/index.html> (2024).
- 130 van Rossum, B.-J. *Modeling field trials using statgenSTA*, <https://biometris.github.io/statgenSTA/articles/statgenSTA.html> (2024).
- 131 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society:*

- Series B (Methodological)* **57**, 289-300 (1995). <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- 132 Li, T. *et al.* Small RNA and degradome sequencing reveal complex roles of miRNAs and their targets in developing wheat grains. *PLOS ONE* **10**, e0139658 (2015). <https://doi.org/10.1371/journal.pone.0139658>
- 133 Tang, Z. *et al.* Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing *Plant Physiology* **159**, 721-738 (2012). <https://doi.org/10.1104/pp.112.196048>
- 134 Han, R. *et al.* Identification and characterization of microRNAs in the flag leaf and developing seed of wheat (*Triticum aestivum* L.). *BMC Genomics* **15**, 289 (2014). <https://doi.org/10.1186/1471-2164-15-289>
- 135 Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biology* **17**, 13 (2016). <https://doi.org/10.1186/s13059-016-0881-8>
- 136 Altschul, S. F., W. G., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990). [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 137 Liu, J., Cheng, X., Liu, P. & Sun, J. miR156-Targeted SBP-Box Transcription Factors Interact with DWARF53 to Regulate *TEOSINTE BRANCHED1* and *BARREN STALK1* Expression in Bread Wheat. *Plant Physiology* **174**, 1931-1948 (2017). <https://doi.org/10.1104/pp.17.00445>
- 138 Gupta, O. P., Permar, V., Koundal, V., Singh, U. D. & Praveen, S. MicroRNA regulated defense responses in *Triticum aestivum* L. during *Puccinia graminis* f.sp. *tritici* infection. *Molecular Biology Reports* **39**, 817-824 (2012). <https://doi.org/10.1007/s11033-011-0803-5>
- 139 Yadav, A. *et al.* microRNA 166: an evolutionarily conserved stress biomarker in land plants targeting HD-ZIP family. *Physiology and Molecular Biology of Plants* **27**, 2471-2485 (2021). <https://doi.org/10.1007/s12298-021-01096-x>
- 140 Shaltouki-Rizi, M., Smith, N. E., Brown-Guedira, G. & Mohammadi, M. Shared quantitative trait loci underlying root biomass and phenology in wheat (*Triticum aestivum* L.). *Journal of Agronomy and Crop Science* **210**, e12700 (2024). <https://doi.org/10.1111/jac.12700>
- 141 Mallory, A. C. *et al.* MicroRNA control of *PHABULOSA* in leaf development: importance of pairing to the microRNA 5' region. *The EMBO Journal* **23**, 3356-3364 (2004). <https://doi.org/10.1038/sj.emboj.7600340>
- 142 Lewis, B. P., Shih, I.-h., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of Mammalian MicroRNA Targets. *Cell* **115**, 787-798 (2003). [https://doi.org/10.1016/S0092-8674\(03\)01018-3](https://doi.org/10.1016/S0092-8674(03)01018-3)

- 143 Schwab, R. *et al.* Specific Effects of MicroRNAs on the Plant Transcriptome. *Developmental Cell* **8**, 517-527 (2005). <https://doi.org/10.1016/j.devcel.2005.01.018>
- 144 Zeng, X. *et al.* *OsSPL11* positively regulates grain size by activating the expression of *GW5L* in rice. *Plant Cell Reports* **43**, 228 (2024). <https://doi.org/10.1007/s00299-024-03315-7>
- 145 Liu, Y. *et al.* A high-resolution genotype–phenotype map identifies the *TaSPL17* controlling grain number and size in wheat. *Genome Biology* **24**, 196 (2023). <https://doi.org/10.1186/s13059-023-03044-2>
- 146 Lv, B. *et al.* Characterization of *FLOWERING LOCUS T1 (FT1)* gene in *Brachypodium* and wheat. *PLOS ONE* **9**, e94171 (2014). <https://doi.org/10.1371/journal.pone.0094171>
- 147 Wu, L. *et al.* Regulation of *FLOWERING LOCUS T* by a MicroRNA in *Brachypodium distachyon*. *The Plant Cell* **25**, 4363-4377 (2013). <https://doi.org/10.1105/tpc.113.118620>
- 148 Gauley, A. *Connecting the dots for flowering time genes in wheat*, University of East Anglia, (2020).
- 149 Pearce, S., Vanzetti, L. S. & Dubcovsky, J. Exogenous Gibberellins Induce Wheat Spike Development under Short Days Only in the Presence of *VERNALIZATION1*. *Plant Physiology* **163**, 1433-1445 (2013). <https://doi.org/10.1104/pp.113.225854>
- 150 Jian, C. *et al.* Virus-Based MicroRNA Silencing and Overexpressing in Common Wheat (*Triticum aestivum* L.). *Front Plant Sci* **8** (2017). <https://doi.org/10.3389/fpls.2017.00500>
- 151 Nicotra, A. B. *et al.* The evolution and functional significance of leaf shape in the angiosperms. *Functional Plant Biology* **38**, 535-552 (2011). <https://doi.org/10.1071/FP11057>
- 152 Huang, X. *et al.* Natural variation at the *DEP1* locus enhances grain yield in rice. *Nature Genetics* **41**, 494-497 (2009). <https://doi.org/10.1038/ng.352>
- 153 Hou, G. *et al.* Identification of microRNAs in developing wheat grain that are potentially involved in regulating grain characteristics and the response to nitrogen levels. *BMC Plant Biology* **20**, 87 (2020). <https://doi.org/10.1186/s12870-020-2296-7>
- 154 Wang, J. *et al.* A transposon in the vacuolar sorting receptor gene *TaVSR1-B* promoter region is associated with wheat root depth at booting stage. *Plant Biotechnology Journal* **19**, 1456-1467 (2021). <https://doi.org/10.1111/pbi.13564>

- 155 Winter, D. *et al.* An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLOS ONE* **2**, e718 (2007).  
<https://doi.org/10.1371/journal.pone.0000718>
- 156 Shimada, T. *et al.* Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **100**, 16095-16100 (2003). <https://doi.org/10.1073/pnas.2530568100>
- 157 Liu, Y. *et al.* Comparative Small RNA Profiling and Functional Exploration on Wheat With High- and Low-Cadmium Accumulation. *Frontiers in Genetics* **12**, 635599 (2021). <https://doi.org/10.3389/fgene.2021.635599>
- 158 Wei, L. Q., Yan, L. F. & Wang, T. Deep sequencing on genome-wide scale reveals the unique composition and expression patterns of microRNAs in developing pollen of *Oryza sativa*. *Genome Biology* **12**, R53 (2011). <https://doi.org/10.1186/gb-2011-12-6-r53>
- 159 Han, H. *et al.* Small RNA and degradome sequencing used to elucidate the basis of tolerance to salinity and alkalinity in wheat. *BMC Plant Biology* **18**, 195 (2018).  
<https://doi.org/10.1186/s12870-018-1415-1>
- 160 Belli Kullan, J. *et al.* miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. *BMC Genomics* **16**, 393 (2015).  
<https://doi.org/10.1186/s12864-015-1610-5>
- 161 Meyers, B. C. & Axtell, M. J. MicroRNAs in Plants: Key Findings from the Early Years. *Plant Cell* **31**, 1206-1207 (2019). <https://doi.org/10.1105/tpc.19.00310>
- 162 Kakrana, A., Hammond, R., Patel, P., Nakano, M. & Meyers, B. C. *sPARTA*: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. *Nucleic Acids Research* **42**, e139 (2014). <https://doi.org/10.1093/nar/gku693>
- 163 Baker, C. C., Sieber, P., Wellmer, F. & Meyerowitz, E. M. The *early extra petals I* Mutant Uncovers a Role for MicroRNA *miR164c* in Regulating Petal Number in *Arabidopsis*. *Current Biology* **15**, 303-315 (2005).  
<https://doi.org/10.1016/j.cub.2005.02.017>
- 164 Lian, H. *et al.* Redundant and specific roles of individual *MIR172* genes in plant development. *PLOS Biology* **19**, e3001044 (2021).  
<https://doi.org/10.1371/journal.pbio.3001044>
- 165 Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research* **33**, e179 (2005). <https://doi.org/10.1093/nar/gni178>
- 166 Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics* **39**, 1033-1037 (2007).  
<https://doi.org/10.1038/ng2079>

- 167 Jones-Rhoades, M. W. Conservation and divergence in plant microRNAs. *Plant Molecular Biology* **80**, 3-16 (2012). <https://doi.org/10.1007/s11103-011-9829-2>
- 168 Brinton, J. *et al.* A haplotype-led approach to increase the precision of wheat breeding. *Communications Biology* **3**, 712 (2020). <https://doi.org/10.1038/s42003-020-01413-2>
- 169 Genetically Modified Organisms (Deliberate Release) Regulations *Parliament of the United Kingdom* (2002).
- 170 Genetic Technology (Precision Breeding) Bill *Parliament of the United Kingdom* (2022).
- 171 Kloosterman, W. P. & Plasterk, R. H. A. The Diverse Functions of MicroRNAs in Animal Development and Disease. *Developmental Cell* **11**, 441-450 (2006). <https://doi.org/10.1016/j.devcel.2006.09.009>
- 172 Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M. A. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nature Structural & Molecular Biology* **16**, 144-150 (2009). <https://doi.org/10.1038/nsmb.1552>
- 173 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- 174 Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nature Reviews Genetics* **24**, 161-177 (2023). <https://doi.org/10.1038/s41576-022-00541-1>
- 175 Rani, A. *et al.* MicroRNAs for understanding and improving agronomic traits in oilseed Brassicas. *Plant Gene* **34**, 100422 (2023). <https://doi.org/10.1016/j.plgene.2023.100422>
- 176 MacKey, J. *Neutron and X-ray experiments in wheat and a revision of the speltoid problem*, (1954).
- 177 Simons, K. J. *et al.* Molecular characterization of the major wheat domestication gene *Q*. *Genetics* **172**, 547-555 (2006). <https://doi.org/10.1534/genetics.105.044727>
- 178 Sears, E. R. The aneuploids of common wheat. (University of Missouri, College of Agriculture, Agricultural Experiment Station, Colombia, Missouri, 1954).
- 179 Varani, G. & McClain, W. H. The G·U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Reports* **1**, 18-23 (2000). <https://doi.org/10.1093/embo-reports/kvd001>
- 180 Gauley, A. & Boden, S. A. Genetic pathways controlling inflorescence architecture and development in wheat and barley. *Journal of Integrative Plant Biology* **61**, 296-309 (2019). <https://doi.org/10.1111/jipb.12732>

- 181 Aukerman, M. J. & Sakai, H. Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its *APETALA2*-Like Target Genes. *The Plant Cell* **15**, 2730-2741 (2003). <https://doi.org/10.1105/tpc.016238>
- 182 Lee, D.-Y., Lee, J., Moon, S., Park, S. Y. & An, G. The rice heterochronic gene *SUPERNUMERARY BRACT* regulates the transition from spikelet meristem to floral meristem. *The Plant Journal* **49**, 64-78 (2007). <https://doi.org/10.1111/j.1365-313X.2006.02941.x>
- 183 Lee, D.-Y. & An, G. Two AP2 family genes, *SUPERNUMERARY BRACT* (*SNB*) and *OsINDETERMINATE SPIKELET 1* (*OsIDS1*), synergistically control inflorescence architecture and floral meristem establishment in rice. *The Plant Journal* **69**, 445-461 (2012). <https://doi.org/10.1111/j.1365-313X.2011.04804.x>
- 184 Zhang, Z. *et al.* Duplication and partitioning in evolution and function of homoeologous *Q* loci governing domestication characters in polyploid wheat. *Proceedings of the National Academy of Sciences* **108**, 18737-18742 (2011). <https://doi.org/10.1073/pnas.1110552108>
- 185 Miao, C., Wang, D., He, R., Liu, S. & Zhu, J.-K. Mutations in *MIR396e* and *MIR396f* increase grain size and modulate shoot architecture in rice. *Plant Biotechnology Journal* **18**, 491-501 (2020). <https://doi.org/10.1111/pbi.13214>
- 186 Lin, W., Gupta, S. K., Arazi, T. & Spitzer-Rimon, B. *MIR172d* Is Required for Floral Organ Identity and Number in Tomato. *Int J Mol Sci* **22**, 4659 (2021). <https://doi.org/10.3390/ijms22094659>
- 187 Yates, Andrew D. *et al.* Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Research* **50**, D996-D1003 (2022). <https://doi.org/10.1093/nar/gkab1007>
- 188 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. **35**, 1547-1549 (2018). <https://doi.org/10.1093/molbev/msy096>
- 189 Hall, B. G. Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution* **30**, 1229-1235 (2013). <https://doi.org/10.1093/molbev/mst012>
- 190 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004). <https://doi.org/10.1093/nar/gkh340>
- 191 Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275-282 (1992). <https://doi.org/10.1093/bioinformatics/8.3.275>

- 192 Krueger, F. *Trim Galore*,  
[https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (2019).
- 193 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525-527 (2016).  
<https://doi.org/10.1038/nbt.3519>
- 194 ZYMO RESEARCH. *ssDNA/RNA Clean & Concentrator™*,  
[https://files.zymoresearch.com/protocols/\\_d7010\\_d7011\\_ssdnarna\\_clean\\_concentrator.pdf](https://files.zymoresearch.com/protocols/_d7010_d7011_ssdnarna_clean_concentrator.pdf) (2023).
- 195 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).  
<https://doi.org/10.1093/bioinformatics/bty560>
- 196 Sun, F. *et al.* Whole-genome discovery of miRNAs and their targets in wheat (*Triticum aestivum* L.). *BMC Plant Biology* **14**, 142 (2014).  
<https://doi.org/10.1186/1471-2229-14-142>
- 197 Cunningham, F. *et al.* Ensembl 2019. *Nucleic Acids Research* **47**, D745-D751 (2019). <https://doi.org/10.1093/nar/gky1113>
- 198 Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348** (2015).  
<https://doi.org/10.1126/science.aaa6090>
- 199 vizgen. *MERSCOPE® Instrument User Guide*, [https://vizgen.com/wp-content/uploads/2024/05/91600001\\_MERSCOPE-Instrument-User-Guide\\_Rev-H.pdf](https://vizgen.com/wp-content/uploads/2024/05/91600001_MERSCOPE-Instrument-User-Guide_Rev-H.pdf) (2023).
- 200 Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nature Methods* **19**, 1634-1641 (2022). <https://doi.org/10.1038/s41592-022-01663-4>
- 201 vizgen Post-processing Tool (2022).
- 202 Hie, B. L., Kim, S., Rando, T. A., Bryson, B. & Berger, B. Scanorama: integrating large and diverse single-cell transcriptomic datasets. *Nature Protocols* **19**, 2283-2297 (2024). <https://doi.org/10.1038/s41596-024-00991-3>
- 203 Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nature Methods* **19**, 171-178 (2022). <https://doi.org/10.1038/s41592-021-01358-2>
- 204 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).  
<https://doi.org/10.1186/s13059-017-1382-0>
- 205 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1 - 48 (2015).  
<https://doi.org/10.18637/jss.v067.i01>

- 206 Wen-jun, S. & Forde, B. G. Efficient transformation of *Agrobacterium* spp. by high voltage electroporation. *Nucleic Acids Research* **17**, 8385 (1989).  
<https://doi.org/10.1093/nar/17.20.8385>
- 207 Liu, Q., Wang, F. & Axtell, M. J. Analysis of Complementarity Requirements for Plant MicroRNA Targeting Using a *Nicotiana benthamiana* Quantitative Transient Assay *The Plant Cell* **26**, 741-753 (2014). <https://doi.org/10.1105/tpc.113.120972>
- 208 QIAGEN. *QIAprep® Miniprep Handbook* (2020).
- 209 invitrogen. *Qubit™ dsDNA HS Assay Kit*, [https://assets.thermofisher.com/TFS-Assets/LSG/manuals/Qubit\\_dsDNA\\_HS\\_Assay\\_UG.pdf](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/Qubit_dsDNA_HS_Assay_UG.pdf) (2022).
- 210 Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res* **44**, D67-72 (2016). <https://doi.org/10.1093/nar/gkv1276>
- 211 Chalmers, R., Sewitz, S., Lipkow, K. & Crellin, P. Complete nucleotide sequence of Tn10. *Journal of Bacteriology* **182** (2000). <https://doi.org/10.1128/jb.182.10.2970-2972.2000>
- 212 Halling, S. M., Simons, R. W., Way, J. C., Walsh, R. B. & Kleckner, N. DNA sequence organization of IS10-right of Tn10 and comparison with IS10-left. *Proceedings of the National Academy of Sciences* **79**, 2608-2612 (1982).  
<https://doi.org/10.1073/pnas.79.8.2608>
- 213 Kovařík, A., Matzke, M., Matzke, A. & Koukalová, B. Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. *Molecular Genetics and Genomics* **266**, 216-222 (2001).  
<https://doi.org/10.1007/s004380100542>
- 214 invitrogen. *Library Efficiency® DH5α™ Competent Cells*, <https://assets.fishersci.com/TFS-Assets/LSG/manuals/18263012.pdf> (2006).
- 215 Promega. *Wizard® SV Gel and PCR Clean Up System* (2024).
- 216 Roche. *rAPid Alkaline Phosphatase*, <https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/583/481/rap-ro.pdf> (2021).
- 217 New England Biolabs. Ligation Protocol with T4 DNA Ligase (M0202). *protocols.io* (2021).
- 218 New England Biolabs. High Efficiency Transformation Protocol (C2987I). *protocols.io* (2022).
- 219 Liu, Q. & Axtell, M. J. in *Plant Functional Genomics* Vol. 1284 *Methods in Molecular Biology* (eds J. Alonso & A. Stepanova) 287-303 (Humana Press, 2015).
- 220 Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S. & Mullineaux, P. M. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant

- transformation. *Plant Molecular Biology* **42**, 819-832 (2000).  
<https://doi.org/10.1023/a:1006496308160>
- 221 emmeans: Estimated Marginal Means, aka Least-Squares Means v. 1.10.3 (2024).
- 222 Hothorn, T., Bretz, F. & Westfall, P. Simultaneous Inference in General Parametric Models. *Biometrical Journal* **50**, 346-363 (2008).  
<https://doi.org/10.1002/bimj.200810425>
- 223 multcompView: Visualizations of Paired Comparisons v. 0.1-10 (2024).
- 224 Yao, E. *et al.* GrainGenes: a data-rich repository for small grains genetics and genomics. *Database* **2022**, baac034 (2022).  
<https://doi.org/10.1093/database/baac034>
- 225 New England Biolabs. Q5® Site-Directed Mutagenesis (E0554). *protocols.io* (2022).
- 226 Roche. *Restriction Endonuclease EcoR V*,  
<https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/243/804/ecorvroubul.pdf>
- 227 TSL SynBio. *QUICK GUIDE PART 1: TYPE IIS CLONING WITH THE STANDARD PLANT SYNTAX AND THE GOLDEN GATE MoClo PLASMIDS*,  
<https://synbio.tsl.ac.uk/uploads/a94d80d88a4d03acbe8b53da74fdcf49.pdf>
- 228 New England Biolabs. *NEBcloner*, <https://nebcloner.neb.com/#/> (2024).
- 229 Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* **6**, e16765 (2011). <https://doi.org/10.1371/journal.pone.0016765>
- 230 Smedley, M. A., Hayta, S., Clarke, M. & Harwood, W. A. CRISPR-Cas9 Based Genome Editing in Wheat. *Current Protocols* **1**, e65 (2021).  
<https://doi.org/10.1002/cpz1.65>
- 231 Chen, Y. *et al.* The *T. ispahanicum* elongated glume locus *P2* maps to chromosome 6A and is associated with the ectopic expression of *SVP-A1*. *Theoretical and Applied Genetics* **135**, 2313-2331 (2022). <https://doi.org/10.1007/s00122-022-04114-y>
- 232 Hillis, D. M. & Bull, J. J. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology* **42**, 182-192 (1993). <https://doi.org/10.1093/sysbio/42.2.182>
- 233 Simon, C. An Evolving View of Phylogenetic Support. *Systematic Biology* **71**, 921-928 (2022). <https://doi.org/10.1093/sysbio/syaa068>
- 234 Yi, X., Zhang, Z., Ling, Y., Xu, W. & Su, Z. PNRD: a plant non-coding RNA database. *Nucleic Acids Research* **43**, D982-D989 (2015).  
<https://doi.org/10.1093/nar/gku1162>

- 235 Xin, M. *et al.* Transcriptome Comparison of Susceptible and Resistant Wheat in Response to Powdery Mildew Infection. *Genomics, Proteomics & Bioinformatics* **10**, 94-106 (2012). <https://doi.org/10.1016/j.gpb.2012.05.002>
- 236 Kenan-Eichler, M. *et al.* Wheat Hybridization and Polyploidization Results in Deregulation of Small RNAs. *Genetics* **188**, 263-272 (2011). <https://doi.org/10.1534/genetics.111.128348>
- 237 Mueth, N. A., Ramachandran, S. R. & Hulbert, S. H. Small RNAs from the wheat stripe rust fungus (*Puccinia striiformis* f.sp. *tritici*). *BMC Genomics* **16**, 718 (2015). <https://doi.org/10.1186/s12864-015-1895-4>
- 238 Pandey, R., Joshi, G., Bhardwaj, A. R., Agarwal, M. & Katiyar-Agarwal, S. A Comprehensive Genome-Wide Study on Tissue-Specific and Abiotic Stress-Specific miRNAs in *Triticum aestivum*. *PLOS ONE* **9**, e95800 (2014). <https://doi.org/10.1371/journal.pone.0095800>
- 239 Nagy, T. *et al.* [Letter to the Editor] Comparison of Small RNA Next-Generation Sequencing with and Without Isolation of Small RNA Fraction. *BioTechniques* **60**, 273-278 (2016). <https://doi.org/10.2144/000114423>
- 240 Houston, K. *et al.* Variation in the interaction between alleles of *HvAPETALA2* and microRNA172 determines the density of grains on the barley inflorescence. *Proceedings of the National Academy of Sciences* **110**, 16675-16680 (2013). <https://doi.org/10.1073/pnas.1311681110>
- 241 Sakai, H. *et al.* Rice Annotation Project Database (RAP-DB): An Integrative and Interactive Database for Rice Genomics. *Plant and Cell Physiology* **54**, e6-e6 (2013). <https://doi.org/10.1093/pcp/pcs183>
- 242 Cheng, C. *et al.* Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal* **89**, 789-804 (2017). <https://doi.org/https://doi.org/10.1111/tpj.13415>
- 243 Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* **31**, 8-22 (2022). <https://doi.org/10.1002/pro.4218>
- 244 Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genetics* **39**, 1278-1284 (2007). <https://doi.org/10.1038/ng2135>
- 245 Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science* **280**, 585-590 (1998). <https://doi.org/10.1126/science.280.5363.585>
- 246 Percival, J. *The wheat plant; a monograph.* (Duckworth, 1921).
- 247 Wiwart, M. *et al.* Can Polish wheat (*Triticum polonicum* L.) be an interesting gene source for breeding wheat cultivars with increased resistance to *Fusarium* head

- blight? *Genetic Resources and Crop Evolution* **60**, 2359-2373 (2013).  
<https://doi.org/10.1007/s10722-013-0004-2>
- 248 Biffen, R. H. Mendel's Laws of Inheritance and Wheat Breeding. *The Journal of Agricultural Science* **1**, 4-48 (1905). <https://doi.org/10.1017/S0021859600000137>
- 249 Mendel, G. in *Meeting of the Brünn Natural History Society* 3-47 (Natural History Society of Brünn, 1865).
- 250 Engledow, F. L. The inheritance of glume-length and grain-length in a wheat cross. *Journal of Genetics* **10**, 109-134 (1920). <https://doi.org/10.1007/bf02983276>
- 251 Watanabe, N., Yotani, Y. & Furuta, Y. The inheritance and chromosomal location of a gene for long glume in durum wheat. *Euphytica* **91**, 235-239 (1996).  
<https://doi.org/10.1007/BF00021076>
- 252 Kane, N. A. *et al.* *TaVRT-2*, a Member of the *StMADS-11* Clade of Flowering Repressors, Is Regulated by Vernalization and Photoperiod in Wheat. *Plant Physiology* **138**, 2354-2363 (2005). <https://doi.org/10.1104/pp.105.061762>
- 253 Howe, K. L. *et al.* Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Research* **48**, D689-D695 (2020).  
<https://doi.org/10.1093/nar/gkz890>
- 254 Winfield, M. *Evolution of Wheat*,  
[https://www.cerealsdb.uk.net/cerealgenomics/WheatBP/Documents/DOC\\_Evolution.php](https://www.cerealsdb.uk.net/cerealgenomics/WheatBP/Documents/DOC_Evolution.php)
- 255 Zuckerkandl, E. & Pauling, L. in *Evolving Genes and Proteins* (eds Vernon Bryson & Henry J. Vogel) 97-166 (Academic Press, 1965).
- 256 Force, A. *et al.* Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**, 1531-1545 (1999).  
<https://doi.org/10.1093/genetics/151.4.1531>
- 257 Gregis, V. *et al.* Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in *Arabidopsis*. *Genome Biology* **14**, R56 (2013). <https://doi.org/10.1186/gb-2013-14-6-r56>
- 258 Hartmann, U. *et al.* Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. *The Plant Journal* **21**, 351-360 (2000).  
<https://doi.org/10.1046/j.1365-313x.2000.00682.x>
- 259 Li, D. *et al.* A Repressor Complex Governs the Integration of Flowering Signals in *Arabidopsis*. *Developmental Cell* **15**, 110-120 (2008).  
<https://doi.org/10.1016/j.devcel.2008.05.002>

- 260 Jang, S., Torti, S. & Coupland, G. Genetic and spatial interactions between *FT*, *TSF*  
and *SVP* during the early stages of floral induction in Arabidopsis. *The Plant*  
*Journal* **60**, 614-625 (2009). <https://doi.org/10.1111/j.1365-313X.2009.03986.x>
- 261 Schoof, H. *et al.* The Stem Cell Population of *Arabidopsis* Shoot Meristems Is  
Maintained by a Regulatory Loop between the *CLAVATA* and *WUSCHEL* Genes.  
*Cell* **100**, 635-644 (2000). [https://doi.org/10.1016/S0092-8674\(00\)80700-X](https://doi.org/10.1016/S0092-8674(00)80700-X)
- 262 Gregis, V., Sessa, A., Colombo, L. & Kater, M. M. *AGL24*, *SHORT VEGETATIVE*  
*PHASE*, and *APETALA1* Redundantly Control *AGAMOUS* during Early Stages of  
Flower Development in *Arabidopsis*. *The Plant Cell* **18**, 1373-1382 (2006).  
<https://doi.org/10.1105/tpc.106.041798>
- 263 Theißen, G., Melzer, R. & Rümpler, F. MADS-domain transcription factors and the  
floral quartet model of flower development: linking plant development and  
evolution. *Development* **143**, 3259-3271 (2016). <https://doi.org/10.1242/dev.134080>
- 264 Lee, H. *et al.* Genetic framework for flowering-time regulation by ambient  
temperature-responsive miRNAs in Arabidopsis. *Nucleic Acids Research* **38**, 3081-  
3093 (2010). <https://doi.org/10.1093/nar/gkp1240>
- 265 Mateos, J. L. *et al.* Combinatorial activities of *SHORT VEGETATIVE PHASE* and  
*FLOWERING LOCUS C* define distinct modes of flowering regulation in  
*Arabidopsis*. *Genome Biology* **16**, 31 (2015). <https://doi.org/10.1186/s13059-015-0597-1>
- 266 Kane, N. A. *et al.* *TaVRT2* represses transcription of the wheat vernalization gene  
*TaVRN1*. *The Plant Journal* **51**, 670-680 (2007). <https://doi.org/10.1111/j.1365-313X.2007.03172.x>
- 267 Xie, L. *et al.* *TaVrt2*, an SVP-like gene, cooperates with *TaVrn1* to regulate  
vernalization-induced flowering in wheat. *New Phytologist* **231**, 834-848 (2021).  
<https://doi.org/10.1111/nph.16339>
- 268 Adamski, N. *et al.* *Wheat Training*, <http://www.wheat-training.com/> (2021).
- 269 Pallotta, M. A. *et al.* in *10th International Wheat Genetics Symposium*. 789-791  
(Istituto Sperimentale per la Cerealicoltura, 2003).
- 270 3cr. *PACE® GENOTYPING MASTER MIX USER GUIDE*,  
<https://3crbio.com/document/pace-user-guide/>
- 271 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis.  
*Nature Methods* **9**, 676-682 (2012). <https://doi.org/10.1038/nmeth.2019>
- 272 R: A language and environment for statistical computing (R Foundation for  
Statistical Computing, , Vienna, Austria, 2020).
- 273 RStudio: Integrated Development for R (RStudio, PBC, Boston, Massachusetts,  
2020).

- 274 ZYMO RESEARCH. *Direct-zol™ RNA Miniprep*,  
[https://files.zymoresearch.com/protocols/\\_r2050\\_r2051\\_r2052\\_r2053\\_direct-zol\\_rna\\_miniprep.pdf](https://files.zymoresearch.com/protocols/_r2050_r2051_r2052_r2053_direct-zol_rna_miniprep.pdf) (2023).
- 275 invitrogen. *SuperScript® IV Reverse Transcriptase*,  
[https://assets.thermofisher.com/TFS-Assets/LSG/manuals/SSIV\\_Reverse\\_Transcriptase\\_UG.pdf](https://assets.thermofisher.com/TFS-Assets/LSG/manuals/SSIV_Reverse_Transcriptase_UG.pdf) (2015).
- 276 Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* **25**, 402-408 (2001). <https://doi.org/10.1006/meth.2001.1262>
- 277 Hayta, S. *et al.* An efficient and reproducible *Agrobacterium*-mediated transformation method for hexaploid wheat (*Triticum aestivum* L.). *Plant Methods* **15**, 121 (2019). <https://doi.org/10.1186/s13007-019-0503-z>
- 278 Lazo, G. R., Stein, P. A. & Ludwig, R. A. A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Bio/Technology* **9**, 963-967 (1991). <https://doi.org/10.1038/nbt1091-963>
- 279 Hayta, S., Smedley, M. A., Clarke, M., Forner, M. & Harwood, W. A. An Efficient *Agrobacterium*-Mediated Transformation Protocol for Hexaploid and Tetraploid Wheat. *Current Protocols* **1**, e58 (2021). <https://doi.org/10.1002/cpz1.58>
- 280 SIGMA. *Spectrum™ Plant Total RNA Kit*,  
<https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/323/177/strn250bul.pdf>
- 281 Promega. *RQ1 RNase-Free Dnase Usage Information* (2018).
- 282 Roche. *LightCycler® 480 SYBR Green I Master Method Sheet*, <<https://elabdoc-prod.roche.com/eLD/api/downloads/b26b3645-075e-e511-f1bd-00215a9b3428?countryIsoCode=XG>> (2015).
- 283 Bio-Rad. *Clarity and Clarity Max ECL Western Blotting Substrates kit (Bio-Rad)*,  
<https://www.bio-rad.com/sites/default/files/webroot/web/pdf/lsr/literature/D085075.pdf>
- 284 Cho, H. J. *et al.* SHORT VEGETATIVE PHASE (SVP) protein negatively regulates miR172 transcription via direct binding to the pri-miR172a promoter in *Arabidopsis*. *FEBS Letters* **586**, 2332-2337 (2012). <https://doi.org/10.1016/j.febslet.2012.05.035>
- 285 Lee, J. H., Jin, S., Kim, S. Y., Kim, W. & Ahn, J. H. A fast, efficient chromatin immunoprecipitation method for studying protein-DNA binding in *Arabidopsis* mesophyll protoplasts. *Plant Methods* **13**, 42 (2017). <https://doi.org/10.1186/s13007-017-0192-4>

- 286 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).  
<https://doi.org/10.7554/eLife.21856>
- 287 Hyde, L. *et al.* Identification, characterization, and rescue of CRISPR/Cas9 generated wheat *SPO11-1* mutants. *Plant Biotechnology Journal* **21**, 405-418 (2023). <https://doi.org/10.1111/pbi.13961>
- 288 Lee, Y.-S., Lee, D.-Y., Cho, L.-H. & An, G. Rice *miR172* induces flowering by suppressing *OsIDS1* and *SNB*, two AP2 genes that negatively regulate expression of *Ehd1* and florigens. *Rice* **7**, 31 (2014). <https://doi.org/10.1186/s12284-014-0031-4>
- 289 Lauter, N., Kampani, A., Carlson, S., Goebel, M. & Moose, S. P. *microRNA172* down-regulates *glossy15* to promote vegetative phase change in maize. *Proceedings of the National Academy of Sciences* **102**, 9412-9417 (2005).  
<https://doi.org/10.1073/pnas.0503927102>
- 290 Chuck, G., Meeley, R. & Hake, S. Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes *ids1* and *sid1*. *Development* **135**, 3013-3019 (2008). <https://doi.org/10.1242/dev.024273>
- 291 Chuck, G., Meeley, R., Irish, E., Sakai, H. & Hake, S. The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. *Nature Genetics* **39**, 1517-1521 (2007).  
<https://doi.org/10.1038/ng.2007.20>
- 292 Zhu, Q.-H., Upadhyaya, N. M., Gubler, F. & Helliwell, C. A. Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (*Oryza sativa*). *BMC Plant Biology* **9**, 149 (2009). <https://doi.org/10.1186/1471-2229-9-149>
- 293 Aksimentiev, A. Thread, read, rewind, repeat: towards using nanopores for protein sequencing. *Nature* **633**, 553-534 (2024). <https://doi.org/10.1038/d41586-024-02664-3>
- 294 Huang, K., Baldrich, P., Meyers, B. C. & Caplan, J. L. sRNA-FISH: versatile fluorescent *in situ* detection of small RNAs in plants. *The Plant Journal* **98**, 359-369 (2019). <https://doi.org/10.1111/tpj.14210>