Investigating reproductive development in *Brachypodium distachyon* focussing on the YABBY family of transcription factors

Thesis submitted for the degree of

Doctor of Philosophy

by

Syabira Yusoff

Department of Genetics and Genome Biology

University of Leicester

July 2017

Abstract

Brachypodium, as a sister to the core pooids containing wheat, barley, oats and rye, represents a good model and point of comparison for the study of development and evolution in temperate cereals. Using a comparative cellular developmental and transcriptomic approach, we investigated regulation of key stages in grain development. This was achieved by generating a transcriptome incorporating several distinct developmental stages of *Brachypodium* grains; pre-anthesis ovaries, young grain (1-3 DAA), mid-length grain (3-8 DAA), full-length (8-15 DAA) and mature grain (15-24 DAA), mature grain (without embryo), germinating grains and seedlings stage. By looking at the differential expression of genes through grain development we identified clusters that coincide with the initiation of key developmental stages, such as the initiation of endosperm proliferation, cellularisation and differentiation, as well as the activation of specific metabolic pathways, such as starch and protein biosynthesis. Focus was given to members of the *YABBY* gene family that have an established role in promoting abaxial cell fate and as master regulators of reproductive development in eudicots, but with less clarity in grass species. Using *Brachypodium* as the model plant for cereal crops, the orthologues of *YABBY* genes in grasses were identified and subjected to detailed phylogenetic, expression and functional analyses using Bayesian Interference (BI) analyses, RT-PCR, transcriptomics, mRNA in situ hybridization (ISH) and RNAi. Based on several analyses, *YABBY6* was suggested as a novel candidate of transcription factors regulating seed development in *Brachypodium*. Metadata from Chapter 2 were used to extract similar expression genes of *YABBY* family members and potential motifs regulated in polarity networks involving *YABBY* genes were suggested.

Acknowledgements

"Look at a day when you are supremely satisfied at the end. It's not a day when you lounge around doing nothing; it's a day you've had everything to do and you've done it."

– **Margaret Thatcher**

God, thank you for giving me strength, braveness and the determination to overcome challenges, and always with me in this lonely journey.

Utmost appreciation to my supervisor Sinead Drea (Dr.), who supported me academically and emotionally throughout my PhD. Thank you for lighting the way, helping me grow and holding everything together. You are truly the best.

Former postdoc in my lab and a blossoming plant science lecturer in Cranfield University, Sofia Kourmpetli (Dr.), who helped me a lot academically and whilst I was adjusting to the life in the UK. Your advice and encouragement always made me feel on top of the world.

To my life motivation who keeps me going, my little brother and sister who always support me emotionally. Close friends whom never giving up on me, Amira, Nazwan, Ain, Myer, Siddiq, Luqman, Saira and Saikat. Love conquers everything. Thank you for being awesome and supportive. I wouldn't be here without all of your supports.

My second supervisor, Prof David Twell- thank you for your optimism and encouragement. It means a lot to me. My panel reviewer, Eamonn Mallon (Dr.) and James Higgins (Dr.) for allowing me to pass my probation review and providing inspiration during the process of my research. A special thanks to Rob Hammonds for allowing me to use his lab for my writing up period – this was much appreciated.

My work couldn't have been easier without the support of the greatest Technical and Administration Team in Genetics and Genome Biology Department; Heather, Manish, Yelena, Rachael, Neelam, Lisa, Lewis, Miranda and Ramesh.

My surrounding cheerleaders- Diyana, Nicky, Aisha, Hariza, Zalina, Sara, Cake and Coffee Clubs, Ultimate Frisbee Club (RED), walking and socializing group, lab 319 and 226. Thank you for making my years meaningful and memorable. I love you all with my entire heart. Finally, Malaysian government (MARA) for the financial support provided to me during my PhD studies.

Contents

Abstract ii
Acknowledgements iii
Contents iv
List of Figures ix
List of tables xi
Abbreviations xiii
Chapter 1 -Introduction 1
1.1 The Poaceae Family 1
1.2 *Brachypodium* as model plant 2
1.3 *Brachypodium* toolkit and resources 3
1.4 Evolutionary and Comparative Perspectives 7
1.5 The Importance of Transcription Factors (TFs) in Development 9
1.5.1 Transcription Factors 9
1.5.2 Gene regulation by TFs during plant development 9
1.6 *YABBY* gene family 10
1.6.1 Abaxial- Adaxial polarity by *YABBY* expression 11
1.6.2 History of *YABBY* genes in planta 12
1.6.3 *INO* genes expression across angiosperms 13
1.6.4 *DROOPING LEAF/CRABS CLAW* in angiosperms 14
1.6.5 *FILAMENTOUS FLOWER* expression in Angiosperms 15
1.6.5 *YABBY2* clade in crop improvement and secondary metabolites 15
1.7 Vegetative-reproductive Developmental Transition in Grasses 16
1.7.1 Plant Meristematic Development 16
1.7.2 Meristem-Inflorescence Transition in Grasses 17
1.8 Grain (caryopsis) Development 19
1.8.1 Fundamental features of *Brachypodium* grain development 19
1.8.2 Starch content in *Brachypodium* grains 22
1.8.3 Cell wall composition in *Brachypodium* grains 22
1.8.4 Protein storage in *Brachypodium* grains 23
1.9 Aims and Objectives 23

Chapter 2 – *Brachypodium* Grain Transcriptome 25
2.1 Introduction 25
2.2 Materials and Methods 27
2.2.1 Plant Material and RNA Isolation 27
2.2.2 Trancriptome Analyses 27
2.2.2.1 Workstation setup 27
2.2.2.2 RNA Sequencing and Data Analysis 28
2.2.2.3 Up and Down Regulation of differentially expressed genes 28
2.2.2.4 Gene Ontology Analysis 28
2.2.2.5 Clustering Analysis 29
2.2.3 mRNA *in situ* Hybridization 29
2.2.4 Toluidine Blue Staining and TUNEL Assay 29
2.3 Results and Discussion 30
2.3.1 Selection of developmental stages for transcriptomic analyses 30
2.3.2 Transcriptome sequencing and quality evaluation 32
2.3.3 Functional Classification of genes expressed in grain development 36
2.3.4 Identification of gene expression switches during key transition points 38
2.3.5 Categorization of gene clusters coinciding with key phases in Brachypodium grain development and germination 42
2.3.6 Identification of potential regulators of key developmental switches 44
2.3.7 TFs Domains and families distribution in *Brachypodium distachyon* transcriptome 46
2.3.8 Validation of TF expression patterns and correlation with function 48
2.4 Conclusion 51
Chapter 3 – Evolution and expression profiling of *YABBY* genes 52
3.1 Introduction 52
3.3 Materials and Methods 56
3.3.1 Plant material 56
3.3.2 Phylogenetic analysis 56
3.3.3 Tissue collection, disruption and storage 56
3.3.3.1 Genomic DNA isolation 57
3.3.3.2 Total RNA isolation 57
3.3.3.3 RNA quantification and cDNA synthesis 57
3.3.3.4 Reverse Transcription – Polymerase Chain Reaction (RT-PCR) 57

3.3.3.5 Gel electrophoresis 58
3.3.4 Tissue collection, fixation and processing 58
3.3.4.1 Sectioning and embedding 59
3.3.4.2 Toluidine Blue Staining 59
3.3.5 mRNA in-situ hybridization (ISH) 59
3.4 Results 64
3.4.1 Sequence identification and alignment 64
3.4.2 Phylogenetic analyses 67
3.4.3 Temporal and spatial pattern of *YABBY* genes 72
3.4.4 Novel expression of *YABBY* genes in the grain 80
3.5 Discussion 84
3.5.1 Evolution of *YABBY* genes in plants 84
3.5.2 *INO* and YABBY1 might have been lost during the evolution of grasses 84
3.5.3 *YABBY* expression is conserved in carpel and meristem formation 85
3.5.4 *YABBY* expression in mature carpel is new 86
3.5.5 *YABBY6* is a novel gene that might be involved in grain development 86
3.6 Conclusion 87
Chapter 4 – *YABBY* functional analysis and gene networks 88
4.1 Introduction 88
4.2 Materials and methods 90
4.2.1 Plant Material 90
4.2.2 Gateway cloning (DL) 90
4.2.3 Entry Clones 92
4.2.3.1 Design of attB site primers 92
4.2.3.2 Amplification and purification of PCR products 92
4.2.3.3 BP recombination reaction 93
4.2.3.4 Escherichia coli (E. coli) transformation 94
4.2.3.5 Plasmid miniprep and sequencing 94
4.2.4 Expression Clones 95
4.2.4.1 LR recombination reaction 95
4.2.5 Agrobacterium Transformation by Electroporation 95
4.2.6 Arabidopsis Plant transformation and genotyping (CRC) 96
4.2.6.1 Plant material 96
4.2.6.2 Floral dipping / transformation 96

4.2.6.3 Seed screening 97
4.2.7 Yeast-Two-Hybrid 97
4.2.7.1 PCR cloning 97
4.2.7.2 Selection of restriction enzymes 98
4.2.7.3 High-fidelity PCR 98
4.2.7.4 Double restriction enzyme digestions for DNA inserts and vectors 99
4.2.7.5 T4 Ligation 99
4.2.7.6 E.coli transformation, plasmid miniprep and sequencing 99
4.2.7.7 Yeast Transformation 100
Yeast Strain 100
Yeast Transformation 100
4.2.7.8 Autoactivation Experiments 100
Two-Hybrid Screening Using Yeast Mating 101
4.2.8 Promoter Analysis in DL genes 101
4.2.8.1 Gene set identification 101
4.2.8.2 Expression analysis 101
4.2.8.3 Motif enrichment analysis 101
4.3 Results and discussion 103
4.3.1 Generated clones 103
4.3.2 *BdDL* RNAi transgenic plants 106
4.3.2.2 *BdDL* expected phenotype 109
4.3.2.3 *BdDL* transcripts analysis 113
4.3.3 *Brachypodium FIL1* T-DNA insertion lines 115
4.3.4 Complementation of the Arabidopsis crc mutant with *BdDL* 118
4.3.5 YABBY Protein-Protein Interactions 121
4.3.6 Motif discovery of Polarity Networks 123
4.3.6.1 GO related themes based on similar expression of *YABBY* genes 123
4.4 Conclusions and Recommendations 130
Chapter 5- Conclusions and perspectives 131
5.1 The value of Brachypodium as a model system for grain development 132
5.2 Generating a transcriptome of grain development – its use in Brachypodium and beyond 132
5.3 The evolution of the *YABBY* gene family- new insights 134
5.4 Future directions for *YABBY* studies 134

Appendices **136**
Appendix 3.1 – Gene orthologs and its ID in phylogenetic tree 137
Appendix 3.2 – Plant materials used in RT-PCR and in-situ Hybridization 140
Appendix 3.3 – Primers used in YABBY evolution and expression profiling chapter (Chapter 3) 141
Appendix 3.4 – *YABBY* expression FPKM value 142
Appendix 4.1 – GATEWAY cloning vectors information 143
Appendix 4.2 – Primers used in Functional chapter (Chapter 4) 144
Appendix 4.3 – Yeast 2 Hybrid components and result interpretations 146
References **150**

List of Figures

Chapter 1

Figure 1.1: Phylogenetic analysis based on genomic comparison. 8
Figure 1.2: A schematic summarising and comparing *Brachypodium* and wheat grain development) 21

Chapter 2

Figure 2.1: *Brachypodium* developmental stages used for RNAseq analysis. 31
Figure 2.2: Global transcriptome quality..... 34
Figure 2.3: Number of genes expressed during *Brachypodium* grain development and germination...... 35
Figure 2.4: GO enrichment of *Brachypodium* grain transcriptome, in comparison to the whole transcriptome..... 37
Figure 2.5: Transcript clusters highlighting distinct gene expression profiles during Brachypodium grain development and germination.. 43
Figure 2.6: Clustering of all transcription factors expressed during *Brachypodium* grain development and germination. 45
Figure 2.7: Transcription factors families and domain distribution by InterProScan...... 47
Figure 2.8: Validation of transcriptomic data for selected genes. 50

Chapter 3

Figure 3.1: Morphological differences between eudicot and grass flowers..... 52
Figure 3.2: Chromosomal location and exon structure of Brachypodium *YABBY* genes..... 65
Figure 3.3: YABBY family protein alignment and conserved domains.. 66
Figure 3.4: The phylogeny of *YABBY* genes in plant lineages...... 69
Figure 3.5: Alignment of YABBY-like proteins in algae.. 70
Figure 3.6: Evolution of *YABBY* genes in grasses...... 71
Figure 3.7: *YABBY* genes expression patterns across developmental stages 74
Figure 3.8: Expression analysis of *Brachypodium YABBY* genes using RT-PCR... 75
Figure 3.9:Emerging spikelet and pre-anthesis ovary toluidine blue staining and in-situ hybridization histone expression...... 77
Figure 3.10: mRNA *in-situ* hybridation of *Brachypodium YABBY* genes at the emerging spikelet stage. 79
Figure 3.11: mRNA *in-situ* hybridization analysis of *YABBY* expression at the pre-anthesis stage..... 81
Figure 3.12: *YABBY* orthologues and expression profiling in wheat...... 82

Chapter 4

Figure 4.1. Gateway cloning workflow.. 89
Figure 4.2: Summary of *BdDL* RNAi lines...... 108
Figure 4.3: Phylogenetic footprinting of the *DROOPING LEAF* locus between rice, Brachypodium, maize and sorghum. 110
Figure 4.4: *BdDL* RNAi expression clone generated using GATEWAY® technology. 111
Figure 4.5: *BdDL* splice tvariants and DL_RNAi insert alignment. 112
Figure 4.6: *BdDL* spliced transcript expression proportion for each developmental stage. 114
Figure 4.7: *BdFIL* T-DNA insertion lines...... 116
Figure 4.8: Phenotype observation of crc mutant. 118
Figure 4.9: Sequence similarity between AtCRC, OsDL and BdDL proteins. 120
Figure 4.10: GO term enrichment in *YABBY* co-expressed genes 124
Figure 4.11:Motif discovery of polarity networks. 126
Figure 4.12: Identified promoter sequence motif with its associated genome ontology classifications...... 128
Figure 4.13: Identified promoter sequence motif with its associated genome ontology classifications..... 129

Appendix

Figure A.1: *Brachypodium* plant stages. 140
Figure A.2: Vectors information. 143

List of tables

Chapter 1

Table 1.1: Selected model plants 5
Table 1.2: Online resource for *Brachypodium* distachyon. 6

Chapter 2

Table 2.1: Summary of sequencing reads generated for each sample. 32
Table 2.2: GO enrichment analysis of differentially expressed genes in pairwise comparisons of grain development and germination stages.). 41

Chapter 3

Table 3.1: RT-PCR reactions and thermocycling program to investigate gene expression 58
Table 3.2: PCR reaction and thermocycling conditions for mRNA probe template synthesis 60
Table 3.3: In-vitro transcription of mRNA probe 60
Table 3.4: *In situ* Hybridisation Solution (IHS) composition 62
Table 3.5: *YABBY* gene distribution in available sequenced genomes of grasses... 67
Table 3.6 :Corresponding numbers and gene IDs presented in Figure 3.10. 83

Chapter 4

Table 4.1: Two-step PCR workflow summary for thermocycling PCR 1 and PCR2 93
Table 4.2: BP recombination reaction 94
Table 4.3: LR recombination reaction 95
Table 4.4: Agrobacterium transformation infiltration medium 96
Table 4.5: PCR reaction and thermocycling conditions for Y2H cloning. 98
Table 4.6: Double restriction enzyme digestion reactions 99
Table 4.7: Yeast transformation mixture 100
Table 4.8: Gene constructs generated in this project for functional analyses. 104
Table 4.9: Summary of T0 *BdDL* RNAi lines generated 106
Table 4.10: Characterisation of *BdFIL* T-DNA insertion lines. 117
Table 4.11: Available Arabidopsis mutant lines.. 118
Table 4.12: Protein-protein interactions 122

Appendix

Table A.1 : Gene orthologs and its ID in phylogenetic tree. 137
Table A.2: *YABBY* genes expression value from transcriptome 142

Table A.3: List of components needed in Y2H experiments 146
Table A.4: Yeast Growth Media & Supplements 147
Table A.5: Yeast mating and plates screening results..... 149

Abbreviations

µl	Microlitre
ABA	Abscisic acid
ANT	*AINTEGUMENTA*
At	*Arabidopsis thaliana*
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
Bd	*Brachypodium distachyon*
BLZ2	*BASIC LEUCINE ZIPPER FACTOR*
BM	Branch meristems
bp	Base pairs
cDNA	Complimentary DNA
CDS	Coding sequence
CRC	*CRABS CLAW*
DAA	Days after anthesis
DAG	Days after germination
DAS	Days after sowing
DL	*DROOPING LEAF*
DNA	Deoxyribonucleic acid
FIL	*FILAMENTOUS FLOWER*
FM	Floral meristem
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GA	gibberellin
GO	Gene ontology
HDZIP III	*Class III HOMOEODOMAIN LEUCINE ZIPPER*
IM	Inflorescence meristem
INO	*INNER NO OUTER*

ISH	*In-situ* Hybridization
Kb	Kilobase
KNOXI	*KNOTTED-LIKE HOMEODOMAIN*
L	Litre
LB Agar	Luria broth agar
Min	Minutes
mL	Mililitre
mRNA	Messenger RNA
NBT	Nitro blue tetrazolium
O2	*OPAQUE-2*
OE	Overexpression
Os	*Oryza sativa*
PCR	Polymerase chain reaction
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNAi	Ribonucleic interference
rpm	Revolutions per minute
RT	Room temperature
SAM	Shoot apical meristem
Sec	Seconds
SPA	*STORAGE PROTEIN ACTIVATOR*
SUP	*SUPERMAN*
TFBS	Transcription factor binding sites
TFs	Transcription factors
TOB	*TONGARI-BOUSHI1*
TPM	Transcript per Kilobase per Million
UTR	Un-translated region
UV	Ultraviolet
Y2H	Yeast-2-hybrid

YAB1	*YABBY 1*
YAB2	*YABBY 2*
YAB6	*YABBY 6*

Chapter 1 -Introduction

1.1 The Poaceae Family

Brachypodium is a genus in the Poaceae family. The Poaceae family is commonly called the grass family and contains many of the important cereal crops in cultivation (Kellogg and Buell, 2009). Cereals provide more than 50% of the worldwide crop production and are important renewable resources for food, animal feed, and industrial materials (http://faostat.fao.org/). The Triticeae tribe within the Pooideae subfamily of the grass family (Poaceae) includes the important crop genera *Triticum aestivum* (wheat), *Hordeum vulgare* (barley) and *Secale cereale* (rye). Since increasing crop production during the "green revolution" of 1970 to 1980s, the impact of climate change has been a new challenge in the agricultural sector. In fact, based on the United Nation Decade (2010-2020) report, "Desserts and the Fight Against Desertification", statistics show 24% of the land is degrading globally while nearly 20% of the degrading land is cropland (http://www.un.org). This means the cultivated area of crops around the world has been in decline for 25 years. In addition, the increase in the world population versus the decrease of available fertile land, are obliging us to adjust and find the solutions or alternatives to accommodate the problem.

An important approach to increasing the production of cereals and other grains is through advanced breeding and biotechnology tools such as genetic engineering. The world needs more resistant type crops such as drought tolerant, insect tolerant and herbicide tolerant to maintain the yield due to environmental challenge of commercial variety of cereals to equally increase the yield and decreased the post-harvest lost (Mochida and Shinozaki, 2010, Li and Olsen, 2016). This situation highlights the importance of genetics and genomics because half of the yield's increase in our major crops is largely the result of genetic improvement. Therefore, any information that can be obtained about genes and gene expression of our rich plant genetic resources is extremely valuable.

However, the size and complexity of the genomic and size of wheat, barley and rye have hindered the development of genomics and its application to produce Triticeae crops with improved composition and characteristics. As a result, scientists began to emphasize the

application and importance of a model species as a reference to the Triticeae. The establishment of *Arabidopsis thaliana* as a model plant is a good reference for all plant species, however, *Brachypodium distachyon* (henceforward referred to as *Brachypodium*) may help address outstanding gaps as a more targeted model for the Triticeae. *Brachypodium* has been described as a new model plant for grass functional genomics and in evolutionary terms diverged just prior to the clade of "core pooid" genera that contain the majority of important temperate cereals and forage grasses, such as wheat and barley (Catalan et al., 1997, Kellogg, 1998, Kellogg, 2001)

1.2 *Brachypodium* as model plant

The availability of model systems help scientists to explore the basic principles of their biology and translate this into the challenging and complex systems of more applied interest. The most powerful model plant system is *Arabidopsis thaliana* where the complete genome has been sequenced since 2000 (*Arabidopsis* Genome Initiative, 2000). In contrast to the use of flies, zebrafish and mice in animal biology, *Arabidopsis* was adopted as the general model system across all plant species. Although *Arabidopsis* has been established as a strong and powerful system for plant biology, it is not completely representative of grasses.

Shortcomings in *Arabidopsis* as a model for monocots and grasses are alleviated with the availability of *Oryza sativa* (rice) as a model system. It was introduced as a model system for grass species due to its small 400-600 Mb genomic size and importance as a staple crop. The sequenced rice genome was utilized in comparative genomics in order to understand the evolution of grass genes in other cereal crops such as wheat (Yan et al., 2006, Bruggmann et al., 2006). However, Jung et al. (2008) pointed out the big challenges of rice as the capacity of being a model species. The main drawbacks of rice are the prolonged life cycle and stringent growth requirements. Therefore, the consideration of other grasses as a model system is significant due to the demanding research in functional genomics for crop improvement.

Brachypodium, commonly known as "purple *false* brome", emerged as a model plant for both comparative and functional genomics of grass species (Huo et.al., 2009). *Brachypodium* is named for its features in Greek. Brachy for "short" and podion is "little

foot" - it refers to the small pedicels of the spikelets. The International *Brachypodium* Initiative (IBI, 2005) in 2005 and the genome sequencing proposal for *Brachypodium* was accepted by the Department of Energy's Joint Genome Institute (DOE-JGI) in 2006 greatly accelerating *Brachypodium* research.

Similar to *Arabidopsis*, *Brachypodium* has several features that recommend it as a model plant for functional genomic studies. It has a small physical size, rapid life cycle, short generation time, undemanding growth requirements, a small genome and the ability to be self-pollinated and cross-pollinated which is well suited to a model system for genetic analyses (Draper et al., 2001a).

Many reports have assessed *Brachypodium* as an emerging model plant, taking into account its own unique characteristics in addition to the points of comparison to other cereal crops (Brkljacic et al., 2011, Draper et al., 2001, Garvin, 2007, Girin et al., 2014, Guillon et al., 2012). As well as its use in understanding the biology of food-producing crops, it is aslo useful model plants for other grasses used in biofuel production (Bevan et al., 2010). In the last few decades significant efforts has been made to develop plant species as models for economically significant crops. The comparison details can be found in (**Table 1.1**). Compared to other species, *Brachypodium* is very highly similar as *Arabidopsis* which is about 15- 20 cm height, simple growth requirements, typical generation time (8-12 weeks) and growth densities (~1000 m^{-2}). It is a really practical plant to work based on its height and plant density requirements per metre square, is feasible in a very small and limited research facilities, which previous model such as rice cannot provide. The plant life cycle is the shortest in other grasses, allowing faster and rapid research progress.

1.3 *Brachypodium* toolkit and resources

As an emerging model plant species, *Brachypodium* must be accompanied by efficient and feasible genetic and genomic tools resources to be efficiently exploited by the research community. The huge efforts by International *Brachypodium* Initiative (IBI) accelerate the expansion of the information by collaborations with other institutes.

The genome has been sequenced and the latest assembly is publicly available on Phytozome website. The genome completion of *Brachypodium* assist the platform for

Bioinformatics resources that available publicly in several websites as we can see in the **Table 1.2** below. The germplasm resources for plant manipulation are readily available in the USDA National Plant Germplasm systems (NPGS). There are 30 *Brachypodium* accessions are available from NGPS and the provided website for further detail information about the accession. In fact, additional to the germplasm collection, TILLING populations, the information of the *Brachypodium* propagation and transformation service is the huge advancement of *Brachypodium* as model species for temperate grasses. More specific information and websites are listed in **Table 1.2**.

In terms of molecular resources, DNA clone libraries (cDNA, BAC and specialty plasmids) are available for Bd-21 accessions, with a total of 56-fold genome coverage (184,320 clones) (Febrer et al., 2010). cDNA libraries with different tissue samples (Vogel et al., 2006) have been created that could be used in yeast two-hybrid (Y2H) protein-protein interaction investigations.

During the authors research period, Phytozome Plant Genomic Resource resource was the most updated and very useful tools for basic genes and sequences queries such as keyword search, BLAST, BLAT, JBrowse, PhytoMine, BioMart and genome assemblies (available for download). However, the latest assembly of Wheat (TGAC v1) is not updated yet, hence for any comparative works related to wheat, Ensemble plants would be the best resource to refers to.

In this study, we enhanced the gene expression resources by generating a grain transcriptome with associated Gene Ontology (GO) classifications for further data manipulations (**Chapter 2**). This resource will help to narrow down and identify important, possibly novel, candidate genes and TFs important in grain development.

Table 1.1: Selected model plants –*Arabidopsis*, Rice, *Brachypodium* and crops –Barley, foxtail millet, foxtail green, maize, sorghum and wheat. Physical appearance, growth requirements, life-cycle, its genome assembly and photosynthesis type were listed and categorized based on species (Brkljacic et al., 2011).

Parameter	Arabidopsis	Barley	Brachypodium	Foxtail Millet	*M. sinensis*	Maize	Rice	*S. viridis*	Sorghum	Switchgrass	Wheat
Height (cm)	15–20	50–120	15–20	120–200	150–300	120–300	100	10–250	50–250	200–300	50–100
Density[a] (plants m^{-2})	2,000	80–120	1,000	50	3–4	4	36	1,000	50	6	50
Growth requirements[b] (controlled conditions)	Simple	Intermediate	Simple	Intermediate	Demanding	Demanding	Demanding	Simple	Demanding	Demanding	Intermediate
Efficiently crossed?	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Reproduction	Selfing	Selfing	Selfing	Selfing	Outcrossing	Outcrossing/ self-compatible	Selfing	Selfing	Selfing	Outcrossing	Selfing
Typical generation time (weeks)	8–12	10–20	8–12	11–13	12	8–15	12–24	6–8	13–18	26	10–20
Seeds per plant	>1,000	150–200	100–1,000	>10,000	>1,000	200–1,000	>1,000	>5,000	>1,000	>1,000	50–150
Transformation	Extremely easy	Efficient, but labor intensive	Highly efficient	Reported, but not efficient	Inefficient	Efficient, but labor intensive	Highly efficient	Efficient[c]	Inefficient	Efficient, but slow	Inefficient
Genome size (Mb)	119[d]	5,500	272[d]	515	5,000	2,300[d,f]	382[d]	515	700[d,e]	2,400	16,000
Assembled genome sequence	Finished genome sequence	Draft genome sequencing in progress	High-quality draft (finishing under way)	Draft genome	No	Draft genome[f]	Finished genome sequence	Draft genome sequencing in progress	Draft genome[d]	Sequencing in progress	Sequencing in progress
T-DNA resources	Extensive	None	10,000 lines available, 40,000 more planned	None	None	Transposon mutants are available	Extensive[g]	None	None	None	None
Cell wall type	Type 1	Type 2	Type 2	Type 2	Type 2	Type 2	Type 2	Type 2	Type 2	Type 2	Type 2
Photosynthesis	C_3	C_3	C_3	C_4	C_4	C_4	C_3	C_4	C_4	C_4	C_3

[a]High-density planting under laboratory conditions. [b]The difficulty of growing plants is dependent upon their size and the range of environmental conditions tolerated. Thus, small plants that tolerate varied conditions have simple growth requirements and large plants that need carefully controlled environmental conditions have demanding growth requirements. [c]Unpublished data (T. Brutnell). [d]Assembled genome size. [e]Approximately 20% of the genome was not assembled because of the repetitive nature. [f]Assembly consists of sequenced BACs many of which contain unordered genes because of the difficulty associated with assembling repetitive DNA. [g]While rice has extensive insertional mutant resources, the availability of the resources is constrained by quarantine restrictions and intellectual property concerns.

Table 1.2: Online resource for *Brachypodium distachyon*. Resources, institution and its URL were listed with brief descriptions about the resources (Brkljacic et al., 2011)

Resource	Institution	URL	Description
Arizona Genomics Institute	Arizona State University	http://www.genome.arizona.edu	BAC libraries
BrachyBase	Oregon State University	http://www.brachybase.org/	Genome sequence
BrachyBio	Boyce Thompson Institute for Plant Research	http://bti.cornell.edu/brachybio	TILLING population, resources for teachers
Brachypodium distachyon Information Resource	Oregon State University	http://www.brachypodium.org/	Central location for information
Brachypodium genome information	Munich Information Center for Protein Sequences	http://mips.helmholtz-muenchen.de/plant/brachypodium/	Genome sequence
Brachypodium resources	USDA-ARS, Western Regional Research Center	http://brachypodium.pw.usda.gov/	T-DNA lines, methods (crossing, mutagenesis, transformation), germplasm
BrachyTAG	John Innes Centre	http://www.brachytag.org/	T-DNA lines, protocols, transformation vectors, and service
CoGe	University of California, Berkeley	http://synteny.cnr.berkeley.edu/CoGe/	Comparative genomic tools
ELEMENT	Oregon State University	http://element.cgrb.oregonstate.edu/	Promoter searching tool
Garvin laboratory	USDA-ARS Plant Science Research Unit	http://www.ars.usda.gov/pandp/docs.htm?docid=18531	Germplasm, crossing method
GrainGenes	USDA-ARS, Western Regional Research Center	http://wheat.pw.usda.gov	Comparative genomics tools
Gramene	Cold Spring Harbor Laboratory	http://www.gramene.org	Comparative genomics tools
Iowa State University Plant Transformation Facility	Iowa State University	http://www.agron.iastate.edu/ptf/index.aspx	Transformation service
ModelCrop	John Innes Centre	http://www.modelcrop.org/	Genome sequence
NASC's International Affymetrix Service	University of Nottingham	http://affymetrix.arabidopsis.info/	Microarray service
Phytozome	JGI and Center for Integrative Genomics	http://www.phytozome.net/	Comparative genomic tools
PlantGDB	Iowa State University	http://www.plantgdb.org/	Comparative genomic tools
PlexDB	Iowa State University	http://www.plexdb.org/	Expression data (Brachypodium data are expected in mid 2011)
QuantPrime	University of Potsdam	http://www.quantprime.de/	Design oligonucleotides for quantitative PCR
TILLING database	Unité de Recherche en Génomique Végétale	http://urgv.evry.inra.fr/UTILLdb	TILLING collection
USDA NPGS	USDA	http://www.ars-grin.gov/npgs/	Germplasm
WMD3 Web MicroRNA Designer	Max Planck Institute for Developmental Biology	http://wmd3.weigelworld.org/	Tools for designing artificial microRNAs

1.4 Evolutionary and Comparative Perspectives

From an evolutionary perspective, *Brachypodium* has consistently been placed between rice and wheat in the phylogenetic tree (Figure 1.1) that has been evaluated many times as *Brachypodium* data accummulated (Vogel et al., 2006, Vogel and Bragg, 2009). More specifically, *Brachypodium* diverged just prior to the clade of "core pooid" genera that contain the majority of important temperate cereals, such as wheat and barley, and the forage grasses such as *Lolium* and *Festuca* (Catalan et al., 1997, Kellogg, 1998, Kellogg, 2001) In addition, comparative analyses have been done focused on cell walls, grain development and root growth, validating *Brachypodium*'s utility as a model system (Wang et al., 2011, Opanowicz et al., 2011, Hands and Drea, 2012).

In addition to the *Brachypodium distachyon* species itself, other species within the genus are also being exploited. An update on the classification of other *Brachypodium* species was reported by (Catalan et al., 2014, Catalan et al., 2012) designating *Brachypodium hybridum*, *Brachypodium stacei* and *Brachypodium sylvaticum*. The paper describes features of the genus that highlight its relevance in ecological and evolutionary biology i.e., the capacity of different related *Brachypodium* species survived in a wide range of dynamic habitats thereby making it a very useful, undomesticated, versatile model species to understand the effects of abiotic stress in grasses. Another effort to characterize closely related species to *Brachypodium* was effort from Hands et al. (2012) where he included several species of *Brachypodium* and revealed distinct differences between grains of the *Brachypodium* genus and other Triticeae species. Aspects of *Brachypodium*s's grain biology will be examined in more detail in section 1.9 below.

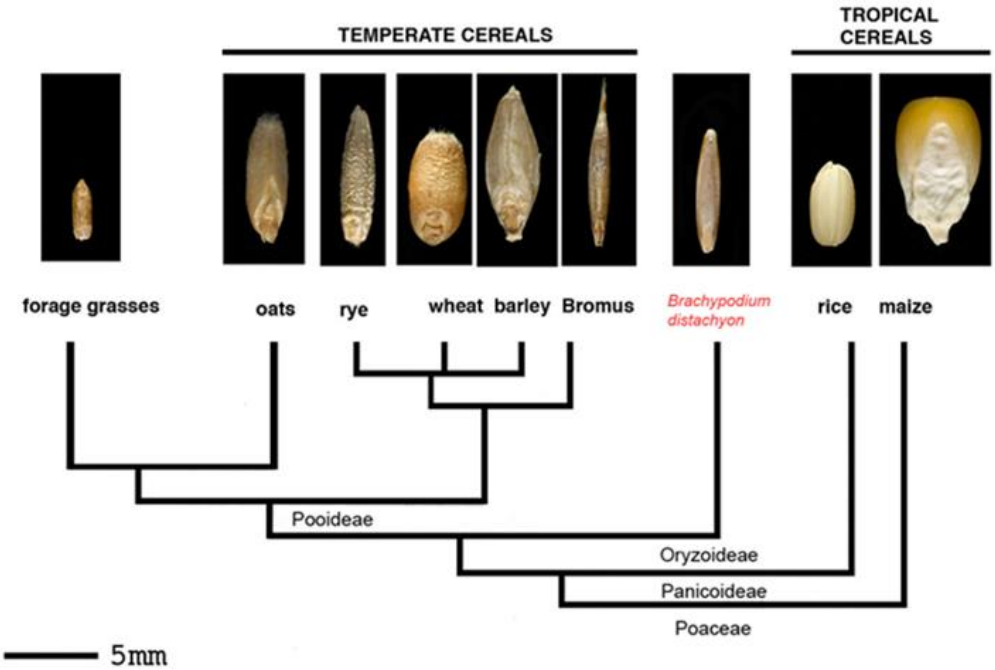


Figure 1.1: Phylogenetic analysis based on genomic comparison. *Brachypodium distachyon* (red) were the species focused in this research. Barr 5mm; seed size. (Opanowicz et al., 2008).

1.5 The Importance of Transcription Factors (TFs) in Development

1.5.1 Transcription Factors

Transcriptional regulatory networks coordinate myriad biological processes in both plants and animals. It is directed by the Transcription factors (TFs) recognizing their cognate cis-regulatory elements in the promoter of their target genes, and regulating the gene's expression.

TFs are proteins that have played a central role in evolution and are major regulators of development in living organisms (Shan et al., 2009). More TFs were discovered in other plants using the genomic approach and classified into different families based on their DNA binding domain (Krogan et al., 2012, Roque et al., 2016, Reinheimer and Kellogg, 2009, Wang et al., 2015, Reyes et al., 2004, Agalou et al., 2008, Kaufmann et al., 2010, Xiong et al., 2005, Latchman, 1997). In addition, TFs can be divided into number of functional classes, with some proteins belonging to more than one class. The first class is activators and repressors, which is the major class of TF. These proteins will bind specifically to DNA sequences that can be found only in certain promoters to promote gene-specific regulation. The second class of TF are coactivators and corepressors. These groups of proteins are not able to bind directly on their own, and need protein-protein interaction to bind with promoter-specific, activators and repressors. The third class consists of the general TFs, which are pivotal components in Pol II (Polymerase II) transcription- initiation complex and the final one is the TFs involved in remodelling DNA that facilitate the binding of other proteins to the promoter (Singh, 1998).

TFs consisting DNA-binding domain and protein-protein interaction domain which interacts with cis-regulatory elements of its target genes and facilitates dimerizations between TF and its regulators, respectively (Wray et al, 2003). In addition, recent finding from Hong et.al. (2003) suggesting the classical floral homeotic genes AGAMOUS (AG) binds to its cis- regulatory elements at the second intron of the gene to infer functional activity of AG, contemplating the capacity of TFs binds to another region in the genome.

1.5.2 Gene regulation by TFs during plant development

Linking TFs to specific aspects of plant development became possible in the early 1990s in model species *Arabidopsis* and in *Antirrhinum majus* where the genetics were tractable and mutant collections were available (Coen and Meyerowitz, 1991). The first classic

identification of TFs were from the *MADS-box* family, responsible for fundamental aspects of flower development in plants (Pinyopich et al., 2003, Gu et al., 1998, Ng and Yanofsky, 2001).

Gene regulations by TFs are pivotal for plant development, differentiation, cell growth, and response to environmental signals (Kaufmann et al., 2010, Riechmann, 2002). During plant development, integration of a wide range of transcription complexes regulate spatial and temporal gene expressions, controlling several biological processes such as flower development, cell fate specification and the establishment in leaf polarity (Benfey and Weigel, 2001, Schiefelbein, 2003).

The interactions between TFs and their DNA binding sites are an integral part of gene regulatory networks (Badis et al., 2009). Interestingly, some TFs to recognize multiple different motifs, thus making it challenging to understand how each protein interacts in the network. In **Chapter 4** we used Y1H (yeast one-hybrid) and Y2H to investigate the protein-protein and protein-DNA networks involved in YABBY-mediated regulation of development.

1.6 YABBY gene family

The *YABBY* family is a group of plant specific TF genes with small numbers of family members (five genes in *Arabidopsis* and eight in rice). The proteins encoded by *YABBY* genes contain two conserved domains, namely zinc finger in the N-terminal region and the helix-loop-helix like *YABBY* domain, in the C-terminal region (Bowman, 2000).

The *YABBY* TF family members are involved in the specification of abaxial polarity in lateral organs such as leaves and floral organs in angiosperms (Eshed et al., 1999). In eudicots such as *Arabidopsis*, this TF family includes *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER* (*FIL*), and *INNER NO OUTER* (*INO*) that have important functions in development of reproductive organs. *CRC* in *Arabidopsis* regulates carpel and nectary development (Bowman and Smyth, 1999, Lee et al., 2005). *DROOPING LEAF* (*DL*) in rice, the orthologue of *CRC* in *Arabidopsis*, is essential in flower development and has additional roles in midrib formation in the leaf (Yamaguchi et al., 2004). Both *CRC* and *DL* sharing the common ancestral function in carpel development regulation (Yamada et al., 2011). *FIL* in *Arabidopsis* controls abaxial/adaxial patterning

and maintenance of flower meristem activity and the expression of *YAB1* and *YAB3* was found in the abaxial region of leaf correlating with the leaf differentiation in *Arabidopsis* (Kumaran, 2002).

1.6.1 Abaxial- Adaxial polarity by *YABBY* expression

Leaves of higher plants exhibit varying degrees of asymmetry along their abaxial/abaxial (lower/upper) axis. This asymmetry is thought to reflect inherent positional differences in the developing organ relative to the shoot apical meristems (SAM) (Timmermans et al., 2004). As mentioned earlier, the optimization of leaf abaxial/adaxial polarity can affect the efficiency of photosynthesis. The abaxial (away from) the meristem or lower surface is to optimize gas exchange, and the adaxial (close to) the meristem or top surface of the leaf is for the light capture (Yamaguchi et al., 2012, Timmermans et al., 2004). The bottom surface (abaxial) consists of sugar-bearing phloem strands, loosely packed spongy mesophyll cells and a high density of stomata while the upper surface (adaxial) is made of water-conducting xylem stands and densely packed of palisade mesophyll cells (Gifford and Foster, 1989).

Ideally, the polarity is established in the meristem and is maintained throughout organ development to coordinate proper outgrowth and patterning of the leaf, so this could be the source of a signal required for proper abaxial/adaxial development of the leaf (Juarez et al., 2004, Bowman et al., 2002, Stahle et al., 2009). Plus, the sign of asymmetry in early stage of leaf development is portrayed by the expression of abaxial-adaxial specific genes in meristems or young organ primordial (Sessions and Yanofsky, 1999).

The classical conceptual framework was developed during 1920s to 1970s describing the interaction between lateral-organ primordia and SAM, and this framework is still widely accepted. The meristem signals to the primordia to establish their abaxial/adaxial polarity which is a prerequisite for the formation of a leaf blade (Reinhardt et al., 2005). In 1955, Sussex discovered this signal by conducting microsurgical isolation of young potato incipient primordia, resulting in formation of radialized primordia that lack of a leaf blade (Sussex, 1955). Sussex also showed that isolated leaf primordia that had already begun their morphological differentiation developed autonomously into phenotypically normal leaves. This indicates that abaxial and adaxial domains are established during the transition from leaf anlage to leaf primordium. Subsequently, several studies have

provided insight into the molecular genetic mechanisms which polarity is established in lateral organs. The establishment of leaf abaxial/adaxial polarity is underpinned by the control of TF and small regulatory RNAs through a series of conserved and overlapping pathways (Bowman et al., 2002, Eshed et al., 2001, Husbands et al., 2009, Kidner and Timmermans, 2007, McConnell and Barton, 1998).

Experiments using *Antirrhinum* with mutations in the MYB transcription factor *PHANTASTICA* (Waites et al., 1998, Waites and Hudson, 1995) and the overexpression of the *YABBY* and *KANADI* families in *Arabidopsis* mutant (Bowman et al., 2002) showed how the polarity is established in the lateral organs. In all cases, adaxial tissues were converted into abaxial. The findings strongly support the previous mechanistic concept (Sussex classical microsurgical work) by which the abaxial/adaxial patterning is dependent on the signal from SAM. In addition, they pull in together where the separation of primordia from the meristem caused the loss of lateral leaflets and dorsoventral patterning defects.

We can say that leaf polarity established as early as meristem initiation by the interaction between SAM and leaf primordium. This mechanism is controlled by specific TF gene families (e.g. *YABBY*, *HD-ZIPIII* and *KNOXI*), to establish an asymmetrical architecture of abaxial/adaxial that is crucial for leaf photosynthesis and functions. In our project, we will focus on the expression of *YABBY* transcription factors and the interactions involved in protein networking in the context of the development of the floral structure and the grain specifically.

1.6.2 History of *YABBY* genes *in planta*

In comparative development or evo-devo studies, a standard approach is to examine key developmental regulatory genes (TFs) in species with distinctive morphologies and compare at sequence, gene expression and protein function levels. This candidate gene approach is employed here with a focus on the *YABBY* family of genes and their roles in patterning the grass flower and fruit. *YABBY* genes encode plant-specific transcription factors (TF) containing zinc finger (C2-C-X20-C2-C) and homeobox domains (later termed as *YABBY* domain). To date, five clades has been identified in angiosperms, i.e. *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER*(*FIL*)/*YABBY3* (*YAB3*), *INNER NO OUTER* (*INO*), *YABBY*2 (*YAB2*) and *YABBY5* (*YAB5*) but *YAB5* is missing in grasses

lineage, deduced from previously published phylogenies in rice (Toriba et al., 2007) and confirmed here (**Chapter 3**).

1.6.3 *INO* genes expression across angiosperms

INNER NO OUTER (INO) is a reproductive *YABBY* associated with both polarity determination and outer integument initiation during ovule development in *Arabidopsis*, while also promoting the abaxial expression like other *YABBY* genes (Villanueva et al., 1999). *INO* gene expression through mRNA detection was initially accumulates only on the abaxial side of ovule primordia at the initiation site of outer integument then were restricted to the layer of outer integument (Villanueva et al., 1999, Meister et al., 2002). In *Arabidopsis*, *INO* has been suggested as a positive regulator of its own expression where a strong *ino* mutant phenotype completely lacks outer integuments and the absence of integument growth (Villanueva et al., 1999).

In 2003, further analysis was carried out by Yamada et al. (2003) in basal-most angiosperms of Nymphaeceae family, *Nymphaea alba* and *Nymphaea colorata*. The *INO* homolog was isolated and the expression patterns were examined. From the report, the abaxial expression in ovule was conserved along angiosperms where the similar abaxial expression of *NaINO* and *NcINO* can be detected in the outer integument. In contrast, the initial detection of *NaINO* signals were different to *Arabidopsis*, where the initial expression was not only restricted to the outer integument initiation. In fact, the *NaINO* expressions were detected in five to six outer epidermal cells of the outer integument, both sides of inner integument and at the nucellus tip cells.

In addition, another paper reported by McAbee et al. (2005) describing *INO* expression pattern among *Impatiens* species indicated that the role of *INO* in the integument growth and polarity of the outer integument has been conserved along angiosperm divergence. With that in mind, Skinner et.al (2016) also concluded the expression localization, sequence identity of INO orthologs is conserved in tomato, but it was not able to complement Arabidopsis INO mutant.

To date, there is little if any detail of *INO* expression in grasses. However, some work has been done by Toriba et al. (2007) analysing the organ specific expression pattern in rice of Os*YABBY7* (orthologue to the *INO*) using real-time PCR. Os*YABBY7* expression was found in leaf sheath, inflorescence (at the stage of floral organ differentiation to floral

organ development) and in mature flower in rice before heading. The expression levels increased along with the flower development and no Os*YABBY7* expression was detected after heading. We assumed that the expression in the outer integument would be the same with the *Arabidopsis* as the *INO* expression was only expressed during the reproductive development and cannot be detected at post heading stage corresponding to the post-anthesis stage in *Arabidopsis*. In fact, *INO* expression level was lower in comparison to other genes in rice. Hence, the detailed expression analysis of *BdINO* in *Brachypodium* will help to understand more about *INO* genes expression and evolution before/in grasses.

1.6.4 *DROOPING LEAF/CRABS CLAW* in angiosperms

The ancestral functions of *CRC* in eudicots are carpel specification and floral meristem (FM) initiation (Yamada et al., 2011, Orashakova et al.,2009, Bowman and Smyth, 1999, Fourquin et al., 2007). The most studied *YABBY* gene in grasses is *DROOPING LEAF*, ortholog of *CRABS CLAW* (*CRC*). In *Arabidopsis*, *CRC* is required for carpel formation and the expression is very limited to abaxial region in carpel and nectaries in *Arabidopsis* flowers. In addition, the *crc* mutant fails to fuse at the apex with reduced amount of style tissue and ovules (Bowman and Smyth, 1999). In contrast to *Arabidopsis*, *DROOPING LEAF* in rice is expressed in the entire carpel and in the middle of undifferentiated cells of leaves, which is very important to regulate carpel identity and midrib formation in rice (Yamaguchi et al., 2004). The spatial expression pattern of *DL* orthologs in several grasses (maize, wheat and sorghum) suggested the function in carpel identity might be conserved in grasses but no explicit functional analyses was performed (Ishikawa et al., 2009).

Genetic regulation involving *DL* genes has been reported by Nagasawa et al. (2003) where he suggested that *DL* and *SUPERWOMAN1* (B gene) was acting mutually and antagonistically to control carpel identity based on mutant analysis. Later on, another published report by (Li et al., 2011) suggested that *OsDL* and *OsMADS13* may function in the same pathways to specify the identity of carpel and floral meristem based on double mutants analysis in rice. *OsDL* and *OsETTIN2* were reported involved in awn development based on comparative RNAi and expression analysis between *Oryza sativa ssp, indica* and *Oryza sativa ssp. japonica*. This research emphasizing about the awn characteristics that might loss during domestication or breeding program in rice (Toriba and Hirano, 2014).

In my opinion we are still lacking detail about *DL* gene regulation in rice and therefore in grasses generally. Hence any further analysis investigating the mutant phenotype, genetic regulations such as protein-protein and protein-DNA will be useful to discover more novel candidates involves in carpel identity. Efforts to investigate this are described in **Chapters 3** and **4**.

1.6.5 *FILAMENTOUS FLOWER* expression in Angiosperms

FILAMENTOUS FLOWER (*FIL*) is another member of the *YABBY* gene family known as "Vegetative *YABBY*" together with other two subfamilies, which are *YABBY*2 and *YABBY*5. These vegetative *YABBY*s were known to promote abaxialization in lateral organs of *Arabidopsis* (Siegfried et al., 1999, Bowman et al., 2002). There are two paralogues in *Arabidopsis* (*FIL* and *YAB3*), three in rice and five paralogues in maize (Yamada et al., 2011, Vosnakis et al., 2012, Tanaka et al., 2012, Eckardt, 2010).

To date, there are several reports describing spatio-temporal expression of *FIL* genes in plants using mRNA in-situ hybridisation method. For instance, in *Arabidopsis*, the expression signal can be detected in the abaxial region of leaf primordia (Chen et al., 1999). In the same report stated that the *fil* mutant lacks of flower clusters i.e. not all pedicels develop a flower at the end, while the double mutant giving more severe defect where the pedicels are completely flowerless. In fact, Chen et al. (1999) said the *FIL* is responsible for early establishment of correct formation and number of flowers in *Arabidopsis*.

In rice and maize, the expression of *FIL* is mainly detected in nonpolar manner in immature organs containing meristems, organ primordia then were temporally expressed in abaxial region of developing vascular bundle, which develops into phloem tissue (Liu et al., 2007, Zhao et al., 2006). However, the expression pattern is not conserved in the monocot lineage. In the maize *zyb9* and *zyb14*, known as the *FIL* orthologues in maize, the expression signal was restricted to the adaxial incipient and leaf primordia region, but not related to the polarity specification (Juarez et al., 2004).

1.6.5 *YABBY2* clade in crop improvement and secondary metabolites

YABBY 2 clade is a subclade in *YABBY* genes family, which has been suggested to be a sister clade to *YABBY*5 in *Arabidopsis*. However, in grasses, *YABBY5* is missing and *YABBY*2 genes are duplicated in this lineage (Eckardt, 2010, Finet et al., 2016, Toriba et

al., 2007, Yamada et al., 2011). The importance of *YABBY2* in plant development was revealed with the genes association with important traits in tomato and sorghum – fruit size and shattering. The *FAS* gene (*YAB2* orthologue) is an important regulator for fruit shape size in tomato (Cong et al., 2008) and its orthologues in sorghum are responsible for shattering effects during its parallel domestication from various geographical locations (Lin et al., 2012).

Interestingly, in addition to traits in crop improvement, *YABBY* genes have been reported as having roles in secondary metabolism in aromatic plants. *YABBY*5 was highly expressed in glandular trichomes in the leaf which is important for secondary metabolites in mint plant whereas *YABBY*5 in other eudicots species remain unknown or classified as redundant functions (Wang et al., 2016).

1.7 Vegetative-reproductive Developmental Transition in Grasses

1.7.1 Plant Meristematic Development

Plant development depends on the activity of various types of meristems generated sequentially at key transition points. The meristems comprise a central region of pluripotent cells and a peripheral region that generates organs such as leaves and floral organs throughout the plant cycle, passing through inflorescence development during the reproductive phase. Inflorescence development controls plant reproduction, in terms of the number of branches produced, the pattern and timing of their production, dictating the number of flowers and the vascular bundles entering the inflorescence (Kellogg et al., 2013). Generally, the shoot apical meristem (SAM) is established during embryogenesis, with unspecified cell structures which eventually convert to an inflorescence meristem (IM) and differentiate to form flowers and branches. However, this simple chronology does not apply to grasses, neither the IM nor the branch meristems (BM) are ever converted directly to floral meristems.

In cereals, only one seed can be produced from each floret, hence controlling the potential number of seed by its unique floret arrangement. In addition, the arrangement of the vasculature in inflorescences is important for the plant to supply the optimum amount of water and photosynthates to the developing seeds. The inflorescence architecture controls

both seed numbers and size which is the most central economic parameters in the cereal grain production, and both together determine the yield. Hence, efforts to describe and understand more about plant meristems and inflorescence is valuable for a new model plant, such as *Brachypodium*. The detail discussion about plant meristem in grasses will be summarised in the next section.

1.7.2 Meristem-Inflorescence Transition in Grasses

In grasses, several types of meristem are responsible for the complex structure of inflorescence. Initially the SAM converts from its vegetative state, producing leaves on its flanks, to form an IM. This process marks the transition from the vegetative phase to the reproductive phase. Then, it initiates the branch meristem (BM) forming the main axis of the inflorescence. Eventually, the BM differentiates into spikelet meristem (SM). And the spikelet consists of one to several florets and two glumes that enclose them, while each floret consists of floral organs (carpel, stamen and lodicule) and two outer organs (palea and lemma) enclosing them. This differentiation process is generally occurring in grasses but different species have different types of meristem differentiations, modifications and inflorescence architectures as described by Bommert et al. (2005). In rice, IM produces several primary branch meristems (pBM) before terminating. Then the BM initiates secondary branch meristem (sBM) before differentiated into SM and ultimately formed single FM. A slightly different orientation happens in maize. Maize has two types of inflorescence, ear and tassel. The ear is a female flower while tassel is a male flower. First, the tassel will arise at the apex of the plant and initiate several LM. Unlike the tassel, the ear will arise from the axil of vegetative leaf and will not form any LM to maximize the flower packing. Both the ear and tassel then were forming the transient spikelet pair meristem (SPM) before differentiating into two SMs. Finally, the SM will form bracts and two FM. However, the lower flower in the ear will be aborted (Bommert et al., 2005, Thompson and Hake, 2009, Tanaka et al., 2013). In fact, the commercial temperate cereals have shown different chronology and inflorescence architecture. In barley (*Hordeum Vulgare*) and wheat (*Triticum aestivum*), both the inflorescence and spike are unbranched unlike in rice and maize. However, the inflorescence in barley is indeterminate and produces lateral spikelet meristems that form three florets (Houston et al., 2013). While the determinate inflorescence meristem development in wheat produces a limited number of lateral spikelet meristems where each

of it will differentiate into multiple florets (Derbyshire and Byrne, 2013). The same report discussed the inflorescence meristem development of *Brachypodium* and the genes related to its development. *Brachypodium* composed of two or three lateral spikelets where each spikelet contains an average of 11 florets arranged in distichous phyllotaxy along the central rachis. The same floret arrangement in early inflorescence architecture in *Brachypodium* has been discussed by Kellogg et al. (2013).

Flowers are the reproductive structure produced by angiosperms for propagation in plants. It is typically made up of attractive petals with protective modified layers of sepal surrounding variable numbers of stamens and carpels producing pollen and ovules which are important for double fertilization in plants (Berger et al., 2008, Dumas and Rogowsky,2008). However, Poaceae or Gramineae are monocotyledonous flowering plants known as grasses and have a distinctive morphological structure (Bommert et al., 2005). Instead of petal and sepal as in the common dicot flowers, grasses have palea/lemma, lodicules, stamens and carpels (Schmidt and Ambrose, 1998, Goto et al., 2001).

The distinctive floral morphology of grasses also encompasses the unique fruit form, the caryopsis or grain. Caryopses in grasses generally consists of embryo, endosperm and pericarp, where the pericarp is the fruit component and embryo/endosperm, the seed. Pericarp in grasses is composed of epicarp, mesocarp and endocarp layers (Brillouet, 1987). Wheat pericarp has been reported to have many functions such as protection, photosynthesis, mineral accumulation, synthesis and degradation of starch (Xiong et al., 2013). During early stage of wheat caryopsis, the pericarp contains rich chloroplasts and chlorophyll, which can synthesise and supply abundant materials for seed respiration and growth. The storage starch in pericarp layer were used for amylose synthesis. Pericarp cells will die rapidly after the storage starch is fully consumed (Zhou et al., 2009).

1.8 Grain (caryopsis) Development

1.8.1 Fundamental features of *Brachypodium* grain development

Since *Brachypodium* been introduced nearly a decade ago, it has gained popularity amongst cereal researchers as an interesting species in itself as well as a model for comparative approaches to gain insights into grain development in cereal crops.

The grain is a single seeded fruit unique to grasses. It consists of three components- pericarp layers (fruit), embryo and endosperm (seed) (Esau, 1960). As in all angiosperms, grain development, including both the maternal and filial (embryo and endosperm) components, initiates upon double fertilization and develops into the mature grain accumulating carbohydrates, proteins and lipid during maturation. Seminal studies investigating grain development in barley and maize revealed the similarity of fundamental steps from early syncytial divisions, through cellularization and tissue specification (Becraft, 2001, Olsen, 2001, Olsen et al., 1992). The endosperm has been particularly well-studied as it is the main reservoir for starch and protein storage. Hence, relatively abundant studies have been reported about the endosperm development in cereal crops such as wheat and barley, with its similar developmental mechanisms and related gene expression (Abebe et al., 2010, Freeman and Palmer, 1984, Gubatz et al., 2007, Opanowicz et al., 2011, Sreenivasulu et al., 2010, Sreenivasulu et al., 2006, Tanackovic et al., 2014, Trafford et al., 2013, Wobus et al., 2005, Bates, 1943, Caley et al., 1990, Drea et al., 2005).

However, the evolutionary perspectives of grain biology are more complete with the rise of *Brachypodium* as a model plant in its taxonomic position sister to the core pooids. There is a large volume of published studies describing the properties and potential of *Brachypodium* as model for grain development (Hands et al., 2012, Pellny et al., 2012, Yu et al., 2015, Guillon et al., 2012, Opanowicsz et al., 2011, Trafford et al., 2013, Ritchie et al., 2007, Drea et al., 2005) and cell wall composition in comparison to other cereal crops.

Figure 1.2 below summarises the grain development stages in *Brachypodium* in comparison with those of wheat. *Brachypodium* grain development takes about 24 days to maturation stage while wheat takes about 35 days after anthesis (DAA). Between 1-2

DAA the endosperm is in the syncytial stage, followed by 2-4 DAA when cellularization occurs - the grain fresh weight increases dramatically during this phase (Guillon et al., 2012). From six to about 20 days is the grain filling process and finally maturation stage. Desiccation is the last stage where the desiccated seed comprises 20 % embryo and 5 % of water content (Guillon et al., 2012) which is lower than mature cereal grains (13-15 %).

Comprehensive analyses in *Brachypodium* endosperm has been carried out and identified several distinct characterizations of *Brachypodium* endosperm compared to wheat (Opanowicz et al., 2011). Endosperm in wheat consisting four major cell types transfer cells, aleurone cells, starchy endosperm cells and embryo surrounding region cells. In *Brachypodium*, the aleurone is not regionally differentiated as in wheat, the central endosperm and nucellar epidermis accumulated with thick cell walls which more closely related in terms of composition to barley but about twice as thick as in wheat (Hands and Drea, 2012, Opanowicz et al., 2011). Anatomically and within the endosperm, *Brachypodium* has an irregular peripheral aleurone and lacks a typical modified aleurone or transfer cell layer (Opanowicz et al., 2011). In surrounding maternal tissues it has a prominent and persistent nucellar epidermis and shows an early maturation of the integument layers (Opanowicz et al., 2011). Overall, it is also notable for its flat, narrow (though not short) grain and the presence of an unpronounced concave ventral/adaxial indentation rather than a crease as in other cereals, most notably, wheat.

According to Trafford (2013), the low rate of cell poliferation in *Brachypodium* could be because of reduced expression of cell-cycle-related genes *cyclin A3* and *CDKB1*. In comparison, during the similar period of cellularization to maximum fresh weight in Barley, the cell cellularization increased 2-fold and the areas increased 15-fold (Trafford et al 2013).

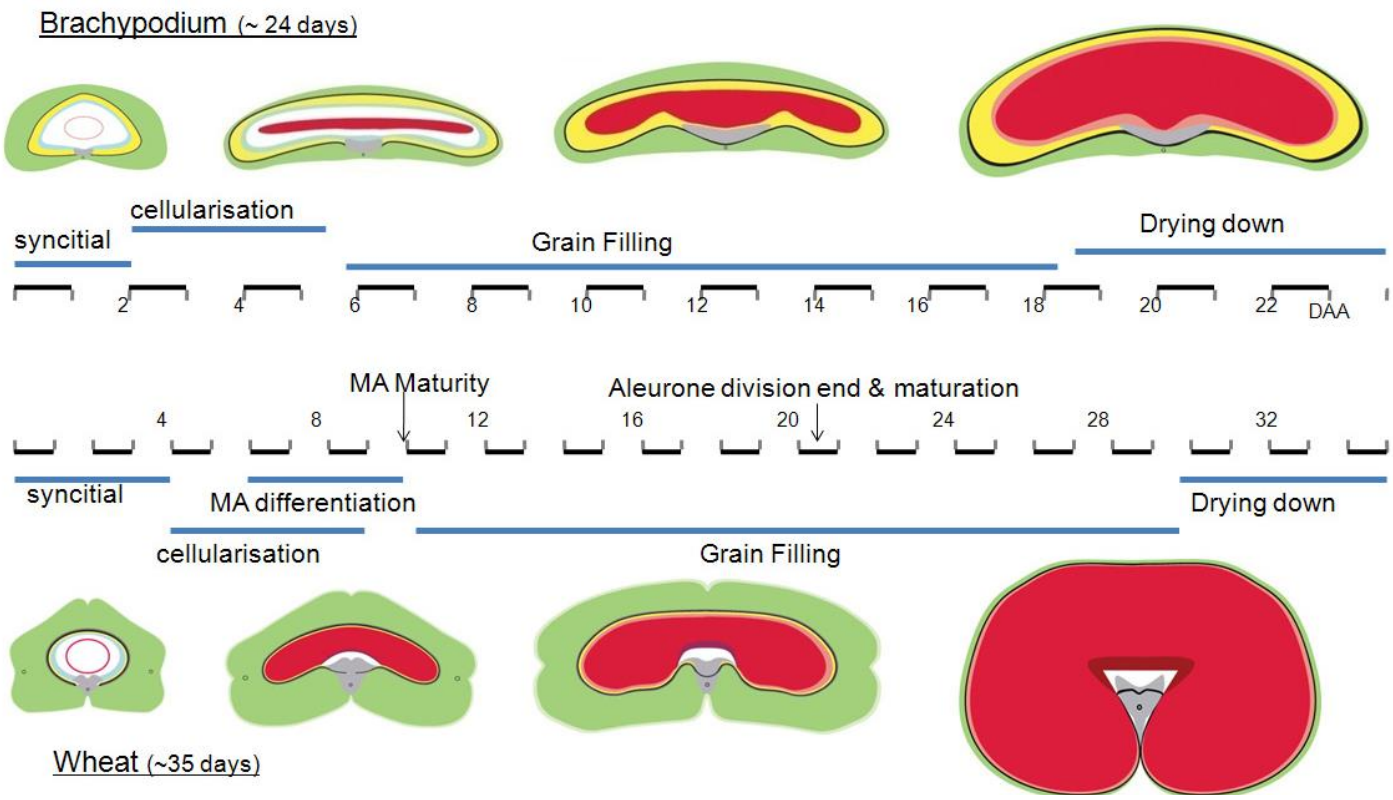


Figure 1.1: A schematic summarising and comparing *Brachypodium* and wheat grain. developmentpericarp, green; integuments; black; nucellar epidermis, yellow; nucellar lysate, blue; nucellar projection, grey; aleurone, pink; central endosperm, red; MA, modified aleurone. DAA -days after anthesis (Opanowicz et al., 2011)

1.8.2 Starch content in *Brachypodium* grains

Starch content is another important feature for grain quality and a number of articles published trying to explain in depth about starch development, composition and biological and chemical pathways involved during grain development (Comparot-Moss and Denyer, 2009). At an early development stage, the starch accumulated in the pericarp layer and started to accumulate in endosperm after cellularization (Chen et al., 2014, Guillon et al., 2012, Trafford et al., 2013).

A comprehensive report by Chen et al. (2014) explained starch granule development in *Brachypodium* in comparison to wheat, Chinese spring (CS) and *Aegilops peregrine*. In this paper, she pointed out that starch accumulation in *Brachypodium* is very high in pericarp layer and it appeared about 8 days after anthesis (DAA) and classified as B-granules with the size less than 10 µm (Wilson et al., 2006). With this in mind, the structure of *Brachypodium* starch granule is similar to barley granule, with smooth surface, but internal part in *Brachypodium* does not possess pores, channels or growth rings (Tanackovic et al., 2014). Meanwhile in wheat, A-granules grew rapidly around 6- 8 DPA, then slowly growth at 8-12 DAF accumulating bigger size of starch granules, ranging between 10-50 µm.

Brachypodium has a very distinctive feature with a very low starch content (6-12 %) compared to 50-70 % starch composition in other cereal grains (Tanackovic et al., 2014). The low starch content in *Brachypodium* grain was possibly related to decrement of expression of starch synthesis genes (Chen et al., 2014, Trafford et al., 2013).

1.8.3 Cell wall composition in *Brachypodium* grains

One of the most prominent feature of *Brachypodium* is the composition of cell walls in its grain. According to Guillon et al. (2012), the cell wall posess about 52 % (w/w) of its endosperm compared to other cereals with 2-7 %. Arabinoxylan and (1,3;1,4)-β-glucan are the component of endosperm cell wall polysaccharides (Burton and Fincher, 2014).

The thick cell wall in *Brachypodium* account for ~52 % (w/w) of grain endosperm compared to other cereals with 2 -7 % (Guillon et al., 2012). *Brachypodium* grains consists of about 80 % (w/w) (1,3;1,4)-β-glucan in the endosperm cell walls (Guillon et al., 2012, Trafford et al., 2013) which is comparable to barley at about 70 % (1,3;1,4)-β-glucan and 20 % arabinoxylan. This is in contrast to wheat with 70 % arabinoylan and only 15 % of (1,3;1,4)-β-glucan (Feng et al., 2009). Guillon et al. (2012) also suggested that the starch accumulation is seven times slower than the cell wall deposition. The higher cell wall composition (1,3;1,4)-β-glucan in

Brachypodium grain also has been suggested as storage compound in *Brachypodium* which provides energy during germination stage in plant cycle (Opanowicz et al., 2011) .

1.8.4 Protein storage in *Brachypodium* grains

Protein storage has been classified into three different fractions, albumins, globulins and prolamins (Shewry and Halford, 2002). The classifications were based on its solubility in water. Globulins, which were widely distributed in flowering plants are soluble in dilute saline solutions, while prolamins (alcohol-soluble), consists of proline and glutamine were restricted to grasses (Larre et al., 2010). With that in mind, storage protein accounts about 50 % of the total protein in mature cereal grains and have crucial impacts on livestock and human consumptions based on their functional properties in food processing. As an example, the composition of prolamins storage protein in wheat is a major contributor in determining the properties of flour dough for end-user consumption (Shewry and Halford, 2002).

Storage protein in *Brachypodium* consists predominantly of globulins, with minor percentages in gliadins and glutenins (Larre et al., 2010, Laudencia-Chingcuanco and Vensel, 2008). The general protein content in *Brachypodium* is extremely high at 28 % compared to barley cultivars, 12 % (Tanackovic et al., 2014), making it closer in composition to that of rice and oat. More knowledge needs to be acquired in terms of the biological and chemical properties of cereal grains to improve and sustain productions for human consumptions and the variation in *Brachypodium's* grain composition makes it a valuable point of reference and a vehicle for understanding the basis of the variation.

1.9 Aims and Objectives

The work described in this thesis has two main aims:

1) To generate resources to enhance our understanding of *Brachypodium* grain development
2) To perform an in-depth analysis of the roles (actual and potential) of the YABBY family of TFs in *Brachypodium* reproductive development, especially grain development, using the resources generated here and by other researchers.

Both these aims were addressed in a comparative approach to maximise how the data generated can also inform our understanding of development in species other than *Brachypodium*.

To address these aims we generated a comprehensive grain development transcriptome in *Brachypodium* and **Chapter 2** describes the time points, generation and analyses of the transcriptome with its data quality, gene ontology (GO) classifications, and validation by comparing the data with several published genes involved during grain development. In the analyses, the TFs expressed at transition points were identified. **Chapter 3** focusses on the phylogenetics and expression profiling and of *YABBY* genes in *Brachypodium*. The final results chapter (**Chapter 4**) describes experiments to investigate functional conservation/diversification of the *YABBY* genes detailing analyses of RNAi lines, complementation experiments, protein-protein interactions and identification of potential DNA sequence motifs involved in polarity network.

Chapter 2 – Brachypodium Grain Transcriptome

2.1 Introduction

Comparative transcriptomic analyses are particularly powerful in research when there is detailed characterization of morphological/anatomical features and biochemical/metabolic processes in the species being compared. This detailed knowledge now exists for many aspects of *Brachypodium*, including grain morphology, development and composition, making it a powerful reference point for comparison to cultivated grain crop species which have strong points of variation (Hands and Drea, 2012; Wu *et al.*, 2016). Another feature that makes it an effective outgroup for comparative studies is that is outside, but sister, to the core pooids (incorporating the triticeae, hordeinae and aveninae subtribes) (Hands et al., 2012, Catalan et al., 1997). Features taken as being characteristic of Poaceae grains were largely based on studies of cultivated members of the family while basic analyses of *Brachypodium* grain structure and contents have highlighted potential variation and points needing further analyses across the Poaceae before a consensus of grain structure, development and composition can be reached.

Notable points of comparison and divergence in *Brachypodium*'s grain biology are its low starch content, its globulin-dominant storage protein composition and its very thick cell walls throughout the endosperm, which are particularly rich in beta-glucans (Trafford et al., 2013, Opanowicz et al., 2011, Laudencia-Chingcuanco and Vensel, 2008, Larre et al., 2010, Hands et al., 2012, Guillon et al., 2012, Guillon et al., 2011). Anatomically and within the endosperm, *Brachypodium* has an irregular peripheral aleurone and lacks a typical modified aleurone or transfer cell layer (Opanowicz et al., 2011). In surrounding maternal tissues it has a prominent and persistent nucellar epidermis and shows an early maturation of the integument layers (Opanowicz et al., 2011). Overall, it is also notable for its flat, narrow (though not short) grain and the presence of an unpronounced concave ventral/adaxial indentation rather than a crease as in other cereals, most notably, wheat.

In terms of composition and, in contrast to wheat and barley, where prolamins are the dominant form of storage protein, *Brachypodium* grains are composed mainly of globulins (Laudencia-

Chingcuanco and Vensel, 2008, Larre et al., 2010). Prolamin storage proteins in cereal grains are regulated by a conserved bZIP transcription factor, the orthologues called SPA (wheat), O2 (maize), *RISBZ1* (rice) and *BLZ2* (barley) (Schmidt et al., 1990, Albani et al., 1997, Onate et al., 1999, Onodera et al., 2001). Less is known about the regulation of globulin storage proteins in cereals though the dimerization partner and paralogue of *RISBZ1*, *REB*, has also been shown to bind to the alpha-globulin promoter in rice (Nakase et al., 1997). The regulation of globulins has been studied more extensively in the eudicots, particularly the Fabaceae tribe, where embryos accumulate globulins during seed development. Studies have shown that there are some common elements in terms of the regulatory motifs found in storage protein gene promoters in monocots and eudicots (Fauteux and Stromvik, 2009). Common promoter motifs, such as the ACGT core sequence, are found in genes expressed in grain and seed maturation in monocots and dicots (Vicente-Carbajosa and Carbonero, 2004).

Two of the most striking features identified in the endosperm component of the *Brachypodium distachyon* grain composition are the low starch content and the thick cell walls in the central endosperm. The low starch content was explained as being due to reduced synthesis and an alternative reallocation of carbohydrate reserves to the cell walls (Trafford et al., 2013). The reduction in the transcription of key starch biosynthetic enzymes, such as *SSI* and *SBEI*, was accompanied by a decrease in cell proliferation and expansion, relative to barley. The endosperm cell walls are distinctive in terms of their thickness and composition and detailed analyses have produced a comprehensive chemical profile (Francin-Allami et al., 2016, Guillon et al., 2011, Opanowicz et al., 2011). Less is known about the regulation of carbohydrate metabolism (starch and cell wall) though recent studies in rice and maize have revealed that the *RISBZ1/O2* proteins are also involved in regulating starch synthesis genes (Wang et al., 2013, Zhang et al., 2016).

The transition from grain maturation to grain germination involves a profound switch in grain metabolism directed by ABA (maturation) and GA (germination). The two hormones act in a coordinated but contrary manner that ultimately promotes maturation and dormancy while suppressing germination and then the reverse (Gubler et al., 2005, Rubio-Somoza et al., 2006). When the grain is mature and dry before imbibition and subsequent germination, the only living tissues in the grain are the peripheral aleurone and the embryo (Hands et al., 2012, Opanowicz et al., 2011). Imbibition prompts GA-induced activation of amylases in the aleurone that act to hydrolyse the starch in the endosperm to provide the energy for germination. The relative paucity of starch in *Brachypodium* leads to the suggestion that the cell walls would provide the

energy reserves required for germination (Opanowicz et al., 2011). The germination process has also been shown to be regulated by conserved transcription factors in *Brachypodium distachyon* and barley (González-Calle et al., 2014).

In this study, we apply a detailed knowledge of the distinct features of grain development and morphogenesis in *Brachypodium distachyon* to generate a comprehensive transcriptome of grain development, while also taking into account the developmental transitions bordering grain development from fertilization to germination. The study goes on to focus on transcription factors (TFs) as a subclass of all genes expressed to illustrate the potential of this resource in identifying putative regulators of grain biology at different stages.

2.2 Materials and Methods

2.2.1 Plant Material and RNA Isolation

Brachypodium distachyon grains (Bd-21 accession) were germinated on moist filter paper after stratification at 4°C for 48h. Five-day old seedlings were transferred to 9cm square pots with a 2:2:1 multipurpose compost: vermiculite: sand mix and grown under controlled environment conditions with a 18h photoperiod at 20°-22°C and light intensity of 180-200 umol/m/s. Tissue samples were collected from eight distinct developmental stages; pre-anthesis ovaries, young grains (1-3 DAA), middle length grains (3-8 DAA), full length grains (8-15 DAA), mature grains (15-20 DAA), mature grains without embryo, germinating grains and seedlings at 3-4 days after germination (**Figure 2.1**). Tissues were frozen in liquid nitrogen immediately after collection and were kept at -80°C until further analyses.

Total RNA was isolated from frozen tissue samples using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) following the manufacturer's instructions. Three biological replicates were used for each stage of development. The RNA samples were treated with DNase I (New England Biolabs) in order to eliminate DNA contamination and their concentration was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

2.2.2 Trancriptome Analyses

2.2.2.1 Workstation setup

LINUX and R environment was used in ALICE account for High Performance Computer (HPC) provided by the University of Leicester. The *Brachypodium*.v3.1 gene annotation assembly file was downloaded from Phytozome12 website (http://phytozome.jgi.doe.gov/) as the latest genome reference.

2.2.2.2 RNA Sequencing and Data Analysis

Random primed cDNA libraries were created and sequenced by GATC Biotech (https://www.gatc-biotech.com). At least 30 million 50 bp single reads were produced per sample. Raw data were analysed following The Basic Tuxedo suite following Trapnell et al. (2012). Adapter sequences were trimmed and the read quality was assessed using FastQC. Tophat along with Bowtie short read aligner were used to map each sample's reads to the Brachypodium v3.1 genome assembly available from Phytozome v12 (https://phytozome.jgi.doe.gov/pz/portal.html). Transcripts assembly and differential expression analysis was performed using Cufflinks and Cuffdiff, respectively. The CummeRbund package was used to explore and visualise the differential expression analysis data (Trapnell et al., 2012).

2.2.2.3 Up and Down Regulation of differentially expressed genes

The gene features and their log 2-fold change values (L2FC) were extracted from CummeRbund. Genes with positive L2FC values between two developmental stages were considered as up-regulated, while those with negative ones as down-regulated. A 2-fold up- and down-regulation cut-off was used to select gene candidates for further analyses.

2.2.2.4 Gene Ontology Analysis

Two gene sets were created; one consisting of all the genes expressed at least in one of the analysed samples ('whole transcriptome') and the subset of all the genes expressed at least in one of the grain developmental stage samples (YG, ML, FL, MG; 'grain transcriptome'). The protein sequences of the 'whole transcriptome' gene set were retrieved from Phytozome v.12 through Biomart and were used for further analyses. A BLAST analysis, followed by InterProScan, Gene Ontology (GO) mapping and GO annotation were performed using the Blast2GO PRO software, for the functional annotation of the genes. Multiple testing of the statistical contrast for the selected gene sets by the Fisher's exact test (provided in Blast2GO PRO software) was performed against the whole transcriptome GO terms with the rate of false positives; FDR = 0.05. Then the enriched GO terms were narrowed down to more specific GO terms or its higher level of GO terms (options available in the Blast2GO software). Transcription factors from the whole transcriptome GO terms were filtered using GO terms related to; GO:0003677 - "DNA binding", GO:0003700 - "TF activity, sequence specific DNA binding", GO:0006355 -"regulation of transcription, DNA-dependent".

2.2.2.5 Clustering Analysis

K-means clustering (K=25) was used to identify gene clusters according to similar expression profiles. Jensen-Shannon distance was used to measure the similarity between two probability distributions and to identify genes expressed at key stages of grain development. The gene lists from each cluster were extracted and used for further GO term analysis in order to get an insight into the transcriptome profiling of each developmental stage.

2.2.3 mRNA in situ Hybridization

Grain and seedling tissues were fixed overnight in FAA (3.7% formaldehyde, 5% acetic acid, 50% ethanol) at 4°C and then transferred to 70% ethanol for further processing. Fixed samples underwent a series of dehydration and infiltration steps before embedded in paraffin as described by Drea et al. (2005). 10-12um thickness sections were cut and slides were prepared as previously described (Opanowicz et al., 2011). Slides were de-waxed and hybridised with the mRNA probes as previously described (Drea et al., 2005). The primers used to generate the BdBLZ1 probe were:

BdBLZ1 F
(5'-TACGCCGAGTCAGCATTTCC-3')
BdBLZ1 RT7
(5'-GAATTGTAATACGACTCACTATAGGGGGGGATCTCCTCCACCGAGAA-3').

2.2.4 Toluidine Blue Staining and TUNEL Assay

Paraffin sections were prepared as above and toluidine blue histological analysis was performed as described by Hands et al. (2012).

For the TUNEL (Terminal deoxynucleotidyl transferase (TdT) dUTP Nick-End Labeling) assay, slides with paraffin sections were first de-waxed and rehydrated as for the mRNA *in situ* hybridisation protocol (Drea et al., 2005). Permeabilization and labelling were performed using the *In-situ* Cell Death Detection kit (Roche) following the manufacturer's instructions. Positive and negative control slides were used for each assay. Hybridised slides were mounted and viewed using a fluorescence microscope.

2.3 Results and Discussion

2.3.1 Selection of developmental stages for transcriptomic analyses

Based on previous detailed analyses of *Brachypodium* grain development (Opanowicz et al., 2011, Guillon et al., 2012), we selected five stages of grain development for transcriptomic processing; pre-anthesis ovaries, young grains at 1-3 DAA, middle length grains at 3-8 DAA, full length grains at 8-15 DAA and mature grain at 15-20 DAA (**Figure 2.1a-j**). The first two stages represent a key developmental point for the morphogenesis phase, followed by storage and maturation phases across middle length to full length grain, finally switching to desiccation in the mature grain (Sreenivasulu et al., 2010). A mature grain stage where the embryo has been removed was also included in the analysis, in order to allow identification of transcripts that are primarily derived from the grain endosperm rather than the embryo itself at this stage of development. This would allow a better understanding of the processes undergoing in the grain endosperm and also provide an indirect profile of the mature embryo transcriptome. Finally, a germinating grain stage and a young seedling stage (**Figure 2.1 k-l**) have also been included in our analysis to allow further insights into the germination process and early vegetative development. Germination is a complex process with several biochemical changes occurring in the grain at that stage (Graeber et al., 2010, Yu et al., 2015), so our aim was to capture the transcriptomic profile of this very dynamic stage of development. The young seedling sample serves a dual purpose; a purely vegetative sample for comparison purposes with the grain transcriptome, but also an insight into the early vegetative development straight after grain germination.

Grain stages are described in terms of size rather than days after anthesis (DAA) as we have found this to provide a stronger correlation with developmental characteristics than counting days post anthesis (DPA), due to the higher asynchrony in grain development in the *Brachypodium* spikelets compared to cultivated species. This approach to staging allows a more direct comparison with other cereal grains, as while the time from anthesis to maturity varies between species, the key developmental transitions are conserved (Hands and Drea, 2012, Opanowicz et al., 2011).

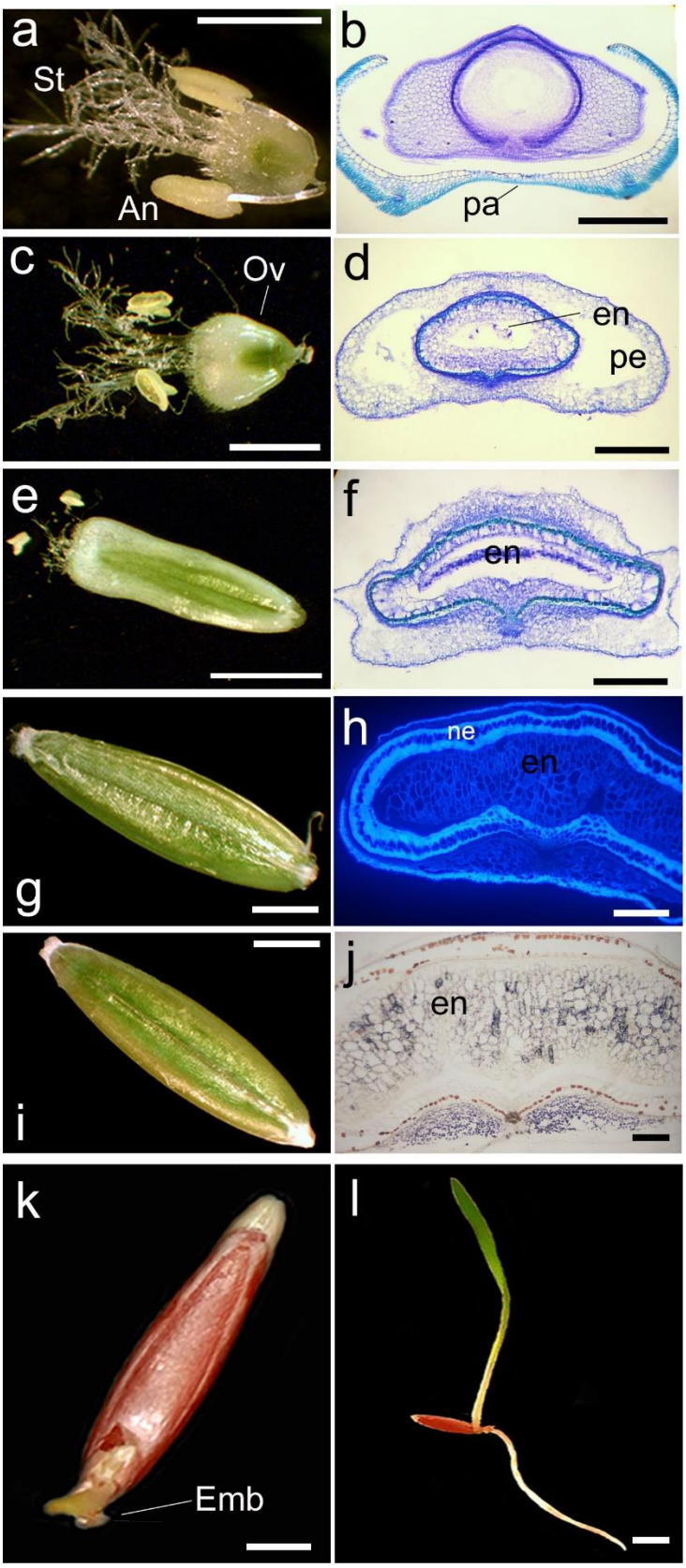


Figure 2.1: ***Brachypodium*** **developmental stages used for RNAseq analysis. a-b.** Pre-anthesis ovaries (PrAnOv). **c-d.** Young grain (YG; 1-3 DAA). **e-f.** Mid-length grain (ML; 3-8DAA). **g-h.** Full-length grain (FL; 8-15 DAA). **i-j.** Mature grain (MG; 15-20 DAA). **k.** Germinating grain (Germ). **l.** Young seedling (SDL; 3-4 days after germination). **b, d** and **f** are toluidine blue stained paraffin sections of pre-anthesis ovary, young grain and mid-length grain, respectively. **h** is Calcoflour white stained section of full length grain showing the persistent nucellar epidermis and **j** is an iodine-stained section of a mature grain showing the low starch levels. An, anther; emb, embryo; en, endosperm; ne, nucellar epidermis; ov, ovary; pa, palea; pe, pericarp; st, stigma. Scale bars (a,c,e,g,i,k) 1mm; (b,d,f,h,j,l) 100µm.

2.3.2 Transcriptome sequencing and quality evaluation

A total of twenty-four RNA samples were analysed consisting of three biological replicates for each of the selected developmental stages described above. All samples were aligned and more than 95 % of input reads were mapped to the latest *Brachypodium* assembly v3.1. **Table 2.1** below summarises the samples that were sequenced, along with the numbers of raw reads generated and the percentages of aligned reads per sample.

Table 2.1: Summary of sequencing reads generated for each sample.

Sample ID	No of reads per sample	Total no of reads per stage	No of mapped reads	% of mapped reads
PrAsOV_0	37890719		37212967	98.2
PrAsOV_1	39566274	119384897	30003556	95.8
PrAsOV_2	41927904		41026168	97.8
YG_0	33632910		32893752	97.8
YG_1	42667262	117800369	41769487	97.9
YG_2	41500197		40635650	97.9
ML_0	42457844		41658412	98.1
ML_1	36312991	106994839	35637367	98.1
ML_2	28224004		27377868	97
FL_0	43178528		42391452	98.2
FL_1	40837378	120893413	40125431	98.3
FL_2	36877507		35642851	96.5
MG_0	41232924		39973308	96.8
MG_1	39621516	119265643	38688283	96.7
MG_2	38411203		37259653	96.9
Germ_0	46008046		45214460	98.2
Germ_1	43785888	131464527	43168187	98.1
Germ_2	41670593		40856581	98
SDL_0	40331265		39785935	98.2
SDL_1	37011613	118988240	36388026	98.1
SDL_2	41645362		41045080	98.4
MG-Emb_0	40214918		38977405	96.9
MG-Emb_1	32245286	38486285	31536469	97.8
MG-Emb_2	42998651		42344412	98.5

*based on mapped reads only

PrAsOv, pre-anthesis ovaries; YG, young grain (1-3 DAA); ML, mid-length grain (3-8 DAA); FL, full-length grain (8-15 DAA); MG, mature grain (15-20 DAA); MG-Emb, mature grain without the embryo; Germ, germinating grain; SDL, seedling 3-4 days after germination.

Numbers 0-2 represent the three independent biological replicates for each developmental stage.

A dendrogram and a dispersion plot were generated to assess the global quality of the generated transcriptomic data (**Figure 2.2**). The three biological replicates for each stage were trifoliously clustered together in the dendogram, confirming that they are more similar compared to other biological samples. Higher branch points indicate less similarity between samples based on Jensen Shannon distance.

A count vs dispersion plot was used to visualize the full model fit. This plot allows the evaluation of the gene distribution for each independent sample. All analysed samples showed an overall good dispersion of transcripts, indicating that the generated data were of good quality and could be used for downstream analyses.

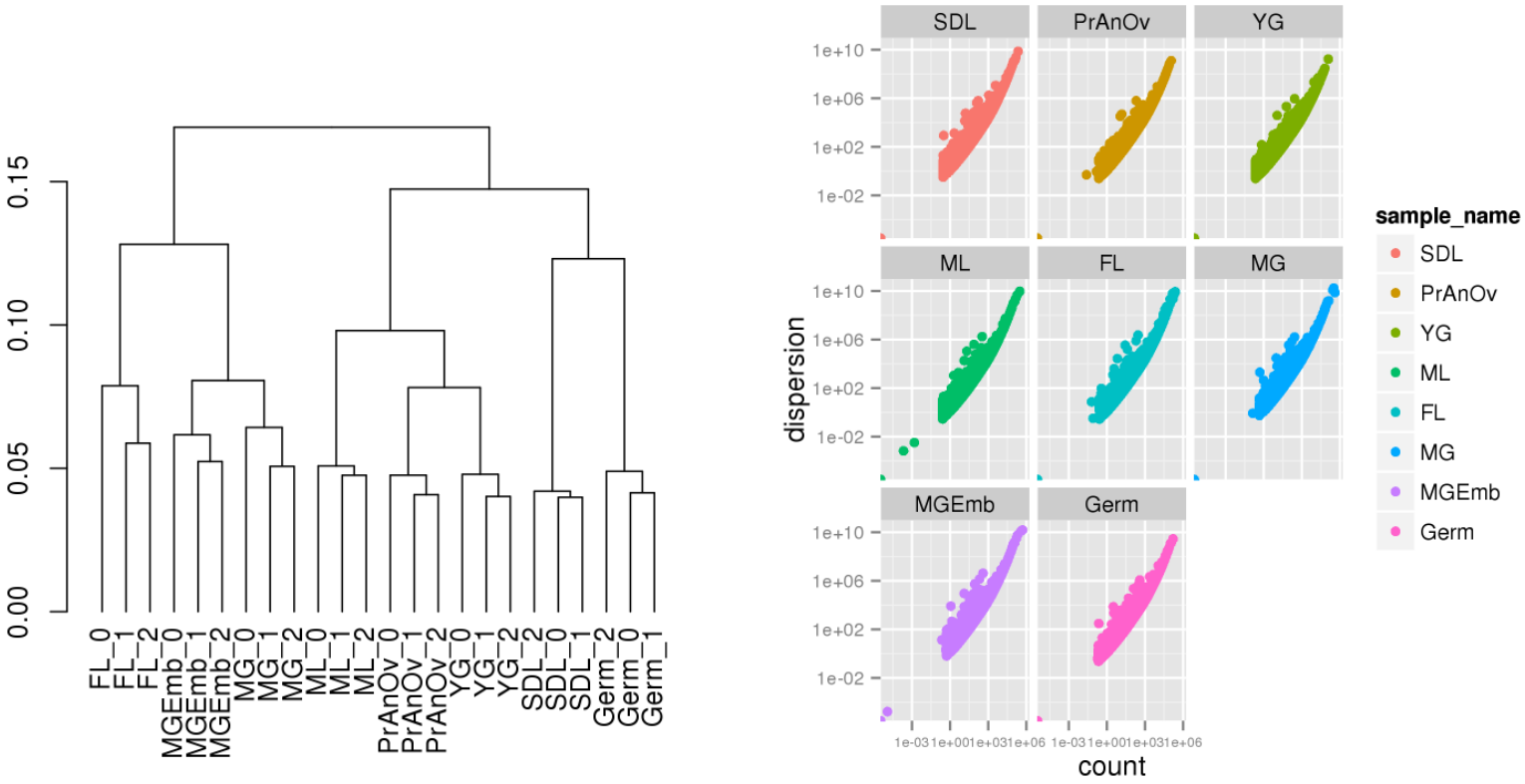


Figure 2.2: Global transcriptome quality. a. Hierarchical clustering of samples used for RNAseq analysis. b. Dispersion plots for each sequenced tissue. PrAnOv, pre-anthesis ovaries; YG, young grain (1-3 DAA); ML, mid-length grain (3-8 DAA); FL, full-length grain (8-15 DAA); MG, mature grain (15-20 DAA); MGEmb, mature grain without the embryo; Germ, germinating grain; SDL, seedling 3-4 days after germination. Numbers 0, 1 and 2 indicate the three biological replicates used for each developmental stage.

A total number of 32563 genes were identified to be expressed across all analysed samples, while the grain transcriptome specifically (including samples from young grain to mature grain only) comprised of 31867 genes -80 % from total identified genes. High number of sharing genes between grain only and whole transcriptome was expected at the beginning of my analysis. Grain development involving different processes such as rapid cellularization, cellular elongation, differentiation, maturations and desiccations are similar in processes during general plant development. Hence, sharing similar genes is not surprising and this idea was supported by Ostergaard (2009).

The Venn diagram in **Figure 2.3** illustrates the number of overlapping and/or unique transcripts between the grain, germination and seedling stages. 26461 transcripts were shared by all samples, 2869 transcripts were unique to the grain transcriptome, 74 were only detected in the germination stage and 134 transcripts were seedling-specific (**Data S2.1**).

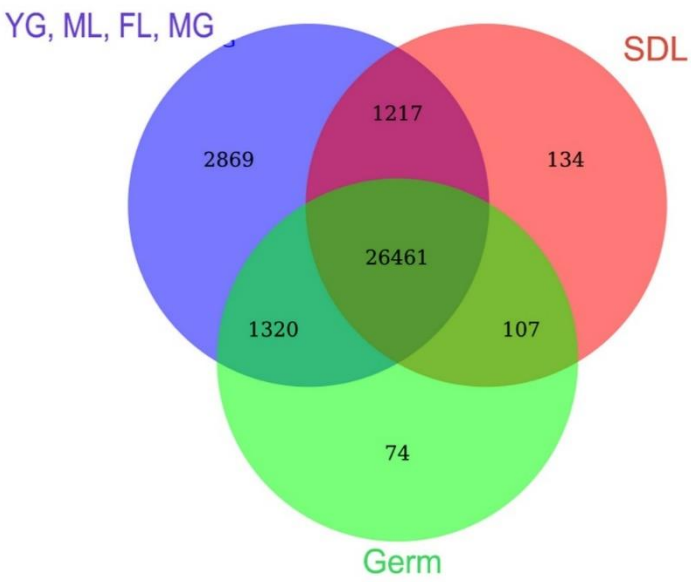


Figure 2.3: Number of genes expressed during *Brachypodium* grain development and germination. a. Venn diagram of the number of genes expressed during grain development (YG, ML, FL, MG), germination (Germ) and young seedling development (SDL).

2.3.3 Functional Classification of genes expressed in grain development

In order to shed light into the *Brachypodium* grain transcriptome, a GO analysis was performed for two different sample sets; 'grain transcriptome' (grain development samples only; YG, ML, FL, MG), and 'whole transcriptome' (all samples, including pre-anthesis, germination and seedling stage; PrAnOV, YG, ML, FL, MG, MG-E, Germ, SDL).

GO analysis revealed that most GO terms are similar (in type and number of sequences) for both sample sets, but there are three categories of Biological process terms that are unique to the grain transcriptome; 'single-organism developmental process', 'anatomical structure development' and 'multicellular organismal process' (**Figure 2.4**). Narrowing down to the grain-only transcriptome slightly shifts the developmental perspective. GO terms unique to the grain-only transcriptome relate to the development of a specific anatomical structure – in this case the grain, but common to the ontogeny of any plant structure (Ostergaard, 2009).

The most abundant GO terms appear to be related with 'protein binding' and 'hydrolase activity' in the molecular function category. 'Protein binding' is important for biological regulation consisting of GO subclasses positive-negative regulation of protein binding, homo-heterodimerization, receptor binding. 'Hydrolase' GO terms are associated with the catalysis of the hydrolysis of various bonds in plants, including enzymes with GTPase and ATPase activity, endonuclease activity, phosphatase activity and microtubule motor activity.

Genes expressed exclusively in grain tissues were associated with several GO terms, such as regulation of gene expression, proteolysis, cellular protein modification process and carboxylic acid metabolic process. Germination -specific genes were associated with post-embryonic development, cellular catabolic process and response to light stimulus. Seedling stage consist of carboxylic acid metabolic process, DNA templated transcription, regulation of cellular process and oxidation-reduction process.

Within the 2869 grain-specific transcripts, we identified several interesting gene families such as Expansins (IPR007118) and cytochrome P450 (IPR001128 & IPR002401). Expansins are important for plant development, as they mediate cell wall extension without degrading the cell wall itself, which might be ruptured under internal pressure during growth (Chebli and Geitmann, 2017). Cytochrome P450 family genes are generally involved in the synthesis of structural molecules such as lignin, cutin or suberin, and the synthesis or catabolism of other components of hormonal signalling and plant defence mechanisms (Pinot and Beisson, 2011, Zhang et al., 2011).

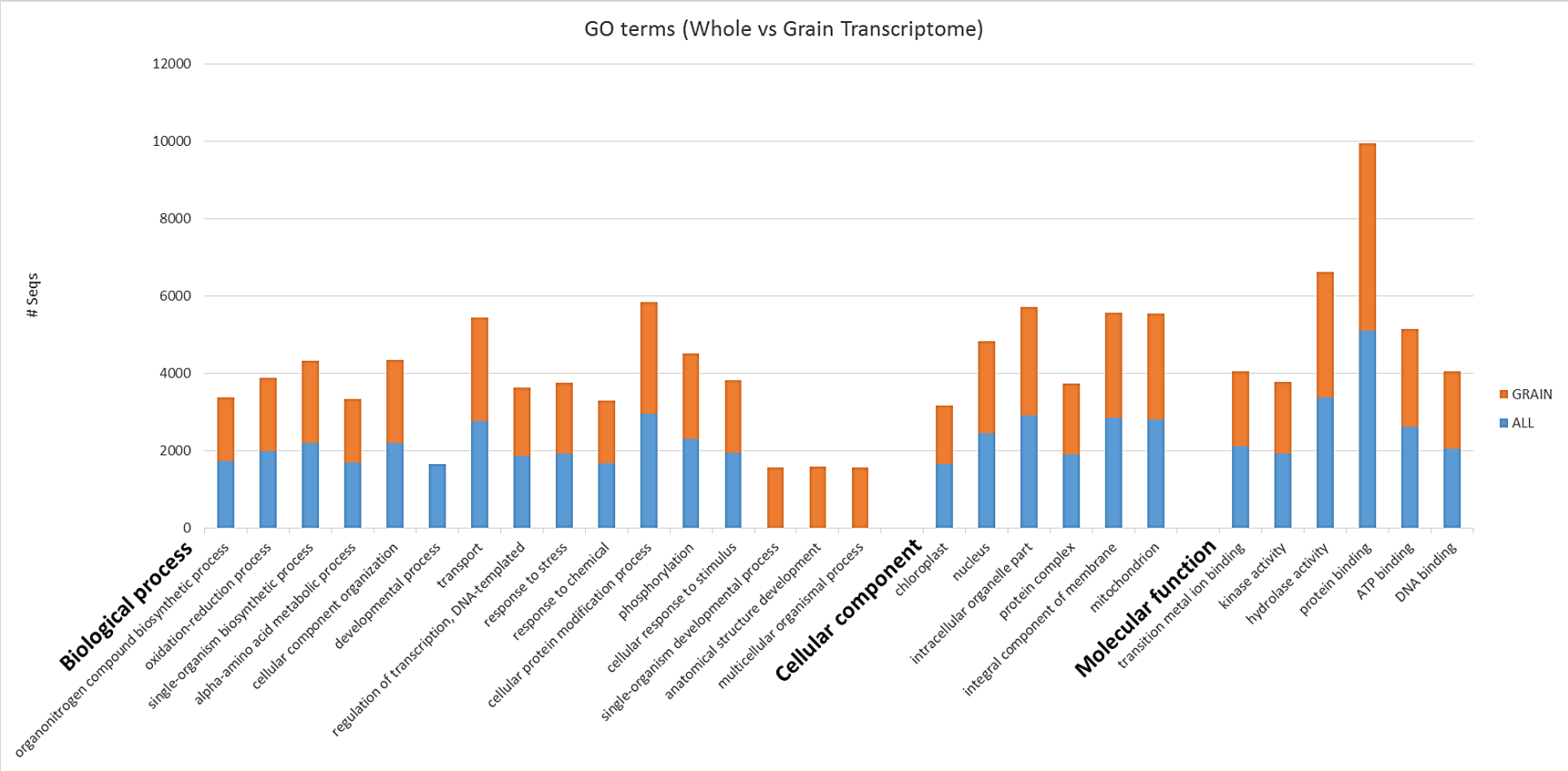


Figure 2.4: GO enrichment of *Brachypodium* grain transcriptome, in comparison to the whole transcriptome. The x-axis indicates the GO annotations of three main categories; Biological process, Cellular Component and Molecular Function, and the y-axis indicates the number of Unigenes.

2.3.4 Identification of gene expression switches during key transition points

To investigate the type of biological and metabolic processes that occur during key transitions in the development of the *Brachypodium* grain, as well as during germination and early seedling development, we performed pairwise developmental stage comparisons and isolated the up- and down-regulated genes that showed a higher than 2 LFC (log2 fold change) in their expression. The top five GO terms of these up-and down-regulated genes are shown in **Table 2.2**, and the full dataset analyses were attached in supplemental **Data S2.3**.

Many useful insights have emerged from this analysis. Firstly, sucrose metabolism appears to be upregulated at each stage early on (from pre-anthesis to young grain, and from there to mid-length grain), while it is then downregulated until the mature grain stage. This is consistent with recent work on wheat, showing that sucrose metabolism-related enzymes reach their highest activity at 14 DAP, while steadily decreasing after that (Wang et al., 2016b). In barley, sucrose levels in the grain follow the same pattern, peaking at 8 DAF and subsequently decreasing (Wobus et al., 2005). Sucrose metabolism at the early stages of grain development plays a key role in fuelling the cells with the necessary energy for the rapid cellularisation. The expression of genes related to the sucrose metabolism seems to also be upregulated during germination, consistent again with fulfilling the energy requirement for seed germination (Yu et al., 1996).

The starch metabolic process appears to be upregulated after fertilization (PrAn-YG), but interestingly, downregulated during the grain filling stage (FL-MG). At the early stages of grain development, starch is mainly accumulated in the pericarp, while during the grain filling and maturation, the endosperm cells are the main starch reservoirs. *Brachypodium* endosperm is particularly starch-poor in comparison to other cereal grains (Opanowicz et al., 2011, Trafford et al., 2013). Our results are consistent with recent research that suggested that the observed low starch content in *Brachypodium* was due to the reduced starch biosynthetic capacity during grain filling (Trafford et al., 2013). The same authors also demonstrated that the endosperm cell enlargement during grain maturation, was mainly due to cell-wall thickening, rather than starch accumulation as in barley and wheat. Our results support this hypothesis, as we observe an upregulation of the hemicellulose metabolic process during the grain filling stage (FL-MG), consistent with the deposition of cell wall component materials. During germination, the carbohydrate catabolic process appears to be upregulated, while there is an observed downregulation of starch biosynthesis correlating with the breakdown of carbohydrate reserves to fuel the germination process.

Another interesting point that confirms what is already known about grain development in *Brachypodium* is that the regulation of the mitotic cell cycle is upregulated at the mid-length stage, while cell proliferation and cytokinesis by cell plate formation are downregulated during the grain maturation (FL-MG). During the mid-length stage grain elongation occurs very rapidly, which could be explained by the enhanced regulation of mitotic divisions. At the grain filling stage though, endosperm cellularisation has already been completed, which is consistent with the downregulation of cell proliferation (Evers and Millar, 2002, Drea et al., 2005). At that stage, response to heat seems to be also upregulated, which could suggest a conditioning of the developing grain in preparation for the desiccation process.

Photosystem II assembly and the thylakoid membrane organisation appear upregulated during the early grain development (YG-ML) (Morrison, 1976). This period coincides with the change of colour in the grain, from white-pale green to dark green, indicating the accumulation of chloroplasts in the pericarp and the increasing photosynthetic capacity of the developing grain. These processes are also upregulated in early seedling development, which would be expected considering the increasing photosynthetic capacity of the developing seedling. It is also possible that genes involved in plastid biogenesis generally, particularly amyloplast biogenesis, would be active at this developmental stage as the endosperm develops (Pogson et al., 2015).

By the time the *Brachypodium* grain reaches final maturity and has desiccated the embryo comprises about 20% of the grain and has attained morphological maturity by 18-26 DAF (Guillon et al., 2012). To allow some discrimination between the endosperm and embryo components of the developing grain we compared intact mature grains (MG), with grains that have their embryo removed (MG-Emb). Our GO enrichment analysis did not identify any upregulated genes (at more than 2-log-fold change) (**Data S2.3**). Some of the top downregulated GO terms though included ´cell cycle´, ´cell division´ and ´cell wall organisation´. This downregulation is associated with transcripts that are involved in biological processes relevant to the embryo component of the mature grain. At the maturation stage, the only living tissues in the grain are the aleurone (comprising only a small percentage of the overall grain tissue) and the embryo (Hands et al., 2012; Opanowicz et al., 2011). Within the list of genes with differential expression (**Data S2.3**) there are 11 Late Embryogenesis Proteins (LEAs) showing down-regulation on removal of the embryo and two farinin storage protein genes with higher expression in the sample with the embryo removed. This enrichment in the embryo and endosperm components would be expected for these genes which play key roles

in embryo maturation and endosperm protein accumulation (Leprince et al., 2017, Subburaj et al., 2016).

Overall, our results are in agreement with what has already been observed in *Brachypodium* and other cereals in terms of grain development and germination and provide further support to suggested mechanisms related to carbohydrate reserves and tissue organisation of the grain.

Table 2.2: GO enrichment analysis of differentially expressed genes in pairwise comparisons of grain development and germination stages. The top five GO terms that show an up- or down-regulation of more than 2-log fold change are presented. PrAnOV, pre-anthesis ovary; YG, young grain (1-3 DAA); ML, mid-length grain (3-8 DAA); FL, full-length grain (8-15 DAA); MG, mature grain (15-20 DAA); MG-Emb, mature grain without the embryo; Germ, germinating grain; SDL, young seedling (3-4 days after germination).

Stage comparison	Up regulated >2.0LFC	Down regulated >2.0LFC
PrAnOV - YG	oxidation-reduction process starch metabolic process sucrose metabolic process negative regulation of catalytic activity glactose metabolic process	auxin-activated signaling pathway cellulose metabolic process amide transport bicarbonate transport borate transport
YG - ML	pentose-phosphate shunt rRNA processing sucrose metabolic process photosystem II assembly thylakoid membrane organisation	oxidation-reduction process defense response cell proliferation regulation of DNA replication DNA methylation
ML - FL	regulation of trancription, DNA -templated regulation of mitotic cell cycle negative regulation of petidase activity coumarin biosynthetic process	sucrose metabolic process protein autiphosphorylation transmembrane receptor protein serine plant organ morphogenesis fatty acid biosynthetic process
FL-MG	regulation of transcription response to inorganic substance response to heat alcohol metabolic process hemicellulose metabolic process	strach metabolic process sucrose metabolic process cell proliferation microtubule-based movement cytokinesis by cell plate formation
MG - Germ	cell wall organisation purine nucleotide metabolic process purine ribonucleoside monophosphate metabolic process sucrose mebolic process carbohydrate catabolic process	pentose-phosphate shunt starch biosyntetic process response to inorganic substance rRNA processing photosystem II assembly
Germ - SDL	isopentenyl diphosphate biosynthetic process thylakoid membrane organization pentose-phosphate shunt rRNA processing photosystem II assembly	regulation of transcription, DNA-templated response to heat response to acid chemical
MG – MG-Emb		regulation of transcription, DNA-templated Cell wall organization Cell cycle Negative regulation of cellular process Cell division

2.3.5 Categorization of gene clusters coinciding with key phases in *Brachypodium* grain development and germination

To allow a better exploration of the generated RNAseq data, differentially expressed genes were clustered based on their temporal expression patterns, using a k-means co-expression clustering analysis. We detected 25 distinct co-expression clusters and we grouped them in five categories (**Figure 2.5 a-e**). Category a included the highest number of transcripts and genes that are constitutively expressed across all surveyed developmental stages. The second group of clusters (b) included genes that appear to peak early in grain development, mainly at the young grain stage (1-3 DAA), or mid- to full- length stage (3-15 DAA) and have very low expression levels during the rest of the grain development. Group c includes gene clusters that show a sharp decrease in their expression at some point during development, while group d genes are possibly involved in grain filling and/or embryo development, as they seem to be expressed mainly during the late stages of grain development. Finally, the last category (e) includes the clusters of genes that possibly have a function in seed germination and/or early seedling development. The full lists of genes and their GO annotations for each gene cluster can be found in **Data S2.4**.

Clustering analysis is a powerful computational tool, allowing the prediction of a functional role for unknown genes that are in the same cluster with well-characterised genes, as clusters tend to be significantly enriched for specific functional categories (Augen, 2004). To date, there are still thousands of genes with unknown function in the *Brachypodium* genome. Clustering analysis could help to shed light in their role in grain development. For example, in the first cluster of category b (**Figure 2.5**), 31 of the 271 genes have no known function. Similarly, 34 out of the 431 genes from the fist cluster of category d are also of unknown function. The two clusters show significantly different expression profiles, with the first one including genes that have their peak of expression at the young grain stage (1-3 DAA) and the second one at the full-grain stage (8-15DAA) and relatively low expression levels at all other stages of development. Looking at the GO enrichment for each cluster, ´cytoplasmic vesicle´ appears to be the most abundant GO term in both clusters. However, the second most abundant GO term for cluster d is ´plastid´ (6.8%), which in cluster b only constitutes 2.9% of the GO terms. This is not surprising, as during the filling stage the grain endosperm cells get filled with amyloplasts for the storage of starch and proteinoplasts for the storage and processing of proteins, while the still green pericarp cells are rich in chloroplasts (Bechtel and Wilson, 2003). At the early stages of grain development, the presence of such plastids is limited. However, at that stage, one of the most abundant functional terms appears to be ´carbohydrate metabolic process´, which is

consistent with the high-energy demand of the young developing grain and the need for carbohydrates as energy sources.

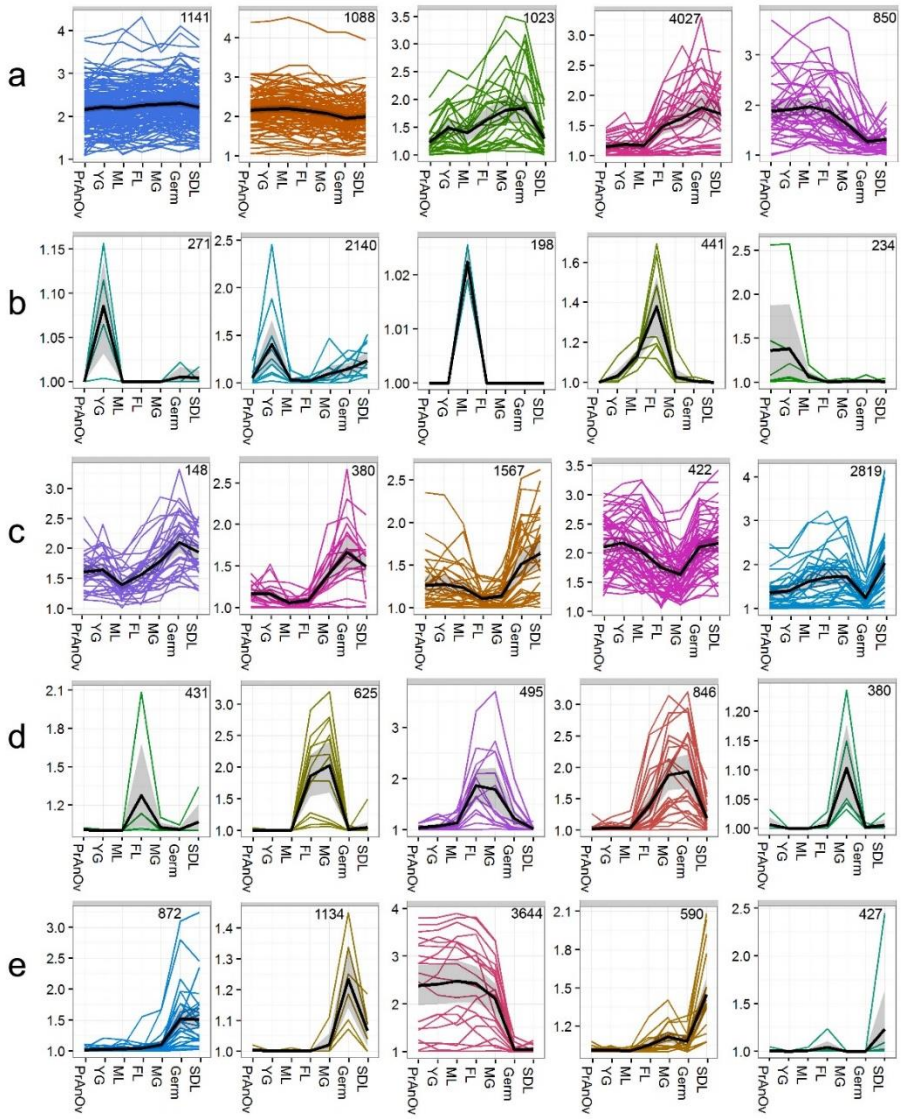


Figure 2.5: Transcript clusters highlighting distinct gene expression profiles during *Brachypodium* grain development and germination. x axis: samples; y axis: median-centered log2(FPKM). The coloured lines represent individual transcript profiles, while the black line shows the average expression value per cluster. The number of transcripts in each cluster is indicated at the top right corner. PrAnOv, pre-anthesis ovaries; YG, young grain (1-3 DAA); ML, mid-length grain (3-8 DAA); FL, full-length grain (8-15 DAA); MG, mature grain (15-20 DAA); Germ, germinating grain; SDL, seedling 3-4 days after germination.

2.3.6 Identification of potential regulators of key developmental switches

Our aim was to identify TFs that might be involved in the regulation of key phases in grain development and germination. For that purpose, we filtered the GO terms of the whole *Brachypodium* transcriptome to select all TFs that are expressed in at least one stage of development. That resulted in 709 TFs that were then clustered according to their expression pattern (**Figure 2.6**). By clustering the expression patterns of all TFs in the grain transcriptome, we were able to identify groups of TFs with very distinct patterns that could provide an indication about their function. For example, the TFs in cluster 2 show a dip in expression during grain development post-anthesis and include several homeobox-containing orthologues. These TFs are central regulators of meristem and early stages of lateral organ development (Hay and Tsiantis, 2010) and so might be downregulated during the differentiation of the grain. TFs of clusters 5 and 6 seem to be highly expressed during the grain filling stage, but show very low expression levels in all other stages, suggesting that they might play a role in the regulation of storage proteins, starch or the thickening of the endosperm cell walls. Indeed, cluster five contains the orthologue of *BLZ2* which regulates storage protein gene expression in barley (Onate et al., 1999). Clusters 7 and 8 show a rather similar pattern but slightly shifted towards germination, which could indicate that they are also involved in aspects of embryo development. Cluster 7 contains the *VP1* (*VIVIPAROUS1*) orthologue, a B3 TF important in seed development in eudicots and recently shown also to regulate maturation and germination in barley grains (Abraham et al., 2016). Cluster eight contains *AP2/ERF* (*APETALA2 / ETHYLENE RESPONSE FACTOR*) orthologues, which are abiotic stress-induced and associated with grain maturation and drying (Licausi et al., 2013). Similarly, TFs included in the clusters 9-12, are possibly involved in germination and/or early seedling development, based on their expression pattern. Cluster 10 contains the orthologue of *LHY* (*LATE ELONGATED HYPOCOTYL*), a central regulator of the circadian clock in *Arabidopsis* that acts to repress the transition to flowering (Fujiwara et al., 2008). A full list of the TFs included in each cluster can be found in **Data S2.5**.

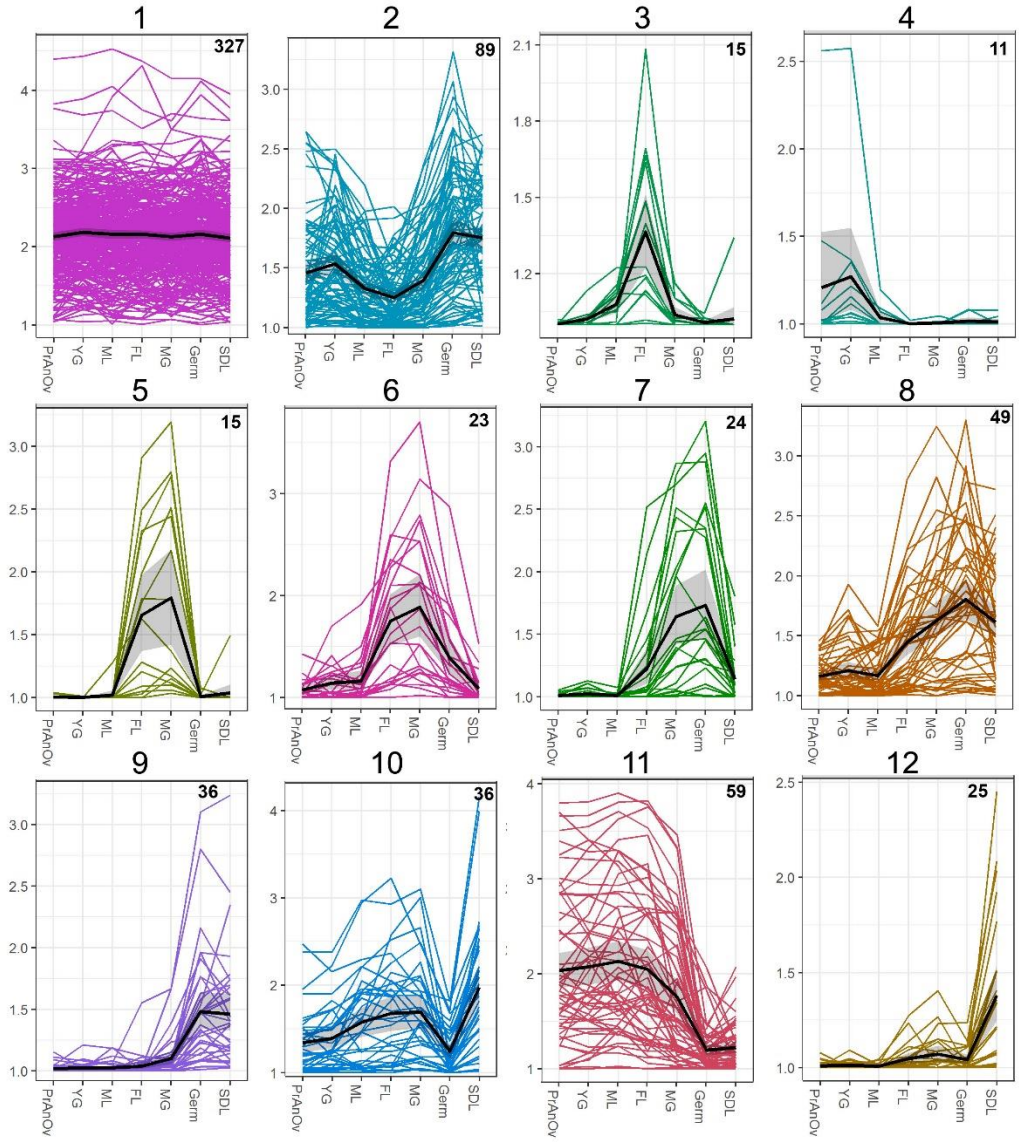


Figure 2.6: Clustering of all transcription factors expressed during *Brachypodium* grain development and germination. x axis: samples; y axis: median-centered log2(FPKM). The coloured lines represent individual transcript profiles, while the black line shows the average expression value per cluster. The clusters are numbered 1-12 and the number of transcripts in each cluster is indicated at the top right corner. PrAnOv, pre-anthesis ovaries; YG, young grain (1-3 DAA); ML, mid-length grain (3-8 DAA); FL, full-length grain (8-15 DAA); MG, mature grain (15-20 DAA); Germ, germinating grain; SDL, seedling 3-4 days after germination.

2.3.7 TFs Domains and families distribution in *Brachypodium distachyon* transcriptome

From the whole transcriptome, GO terms that related to transcription factors were filtered; GO:0003677 -"DNA binding", GO:0003700-"TF activity, sequence specific DNA binding", GO:0006355-"regulation of transcription, DNA-dependent". TFs were filtered by both domains and by family. As TF families are defined by their DNA-binding domains, there should be correlations between the two classifications, but as can be seen in **Figure 2.7** there are differences. The definition of domains can be more generic than specific family annotations and so YABBY family TFs could be included in the zinc-finger domain containing category. Also, while MADS-box domain proteins are well-represented in figure 2.7A, they are absent from the families classification (**Figure 2.7B**). The full list of Domains and Families and its respective genes were attached in **Data S2.6.**

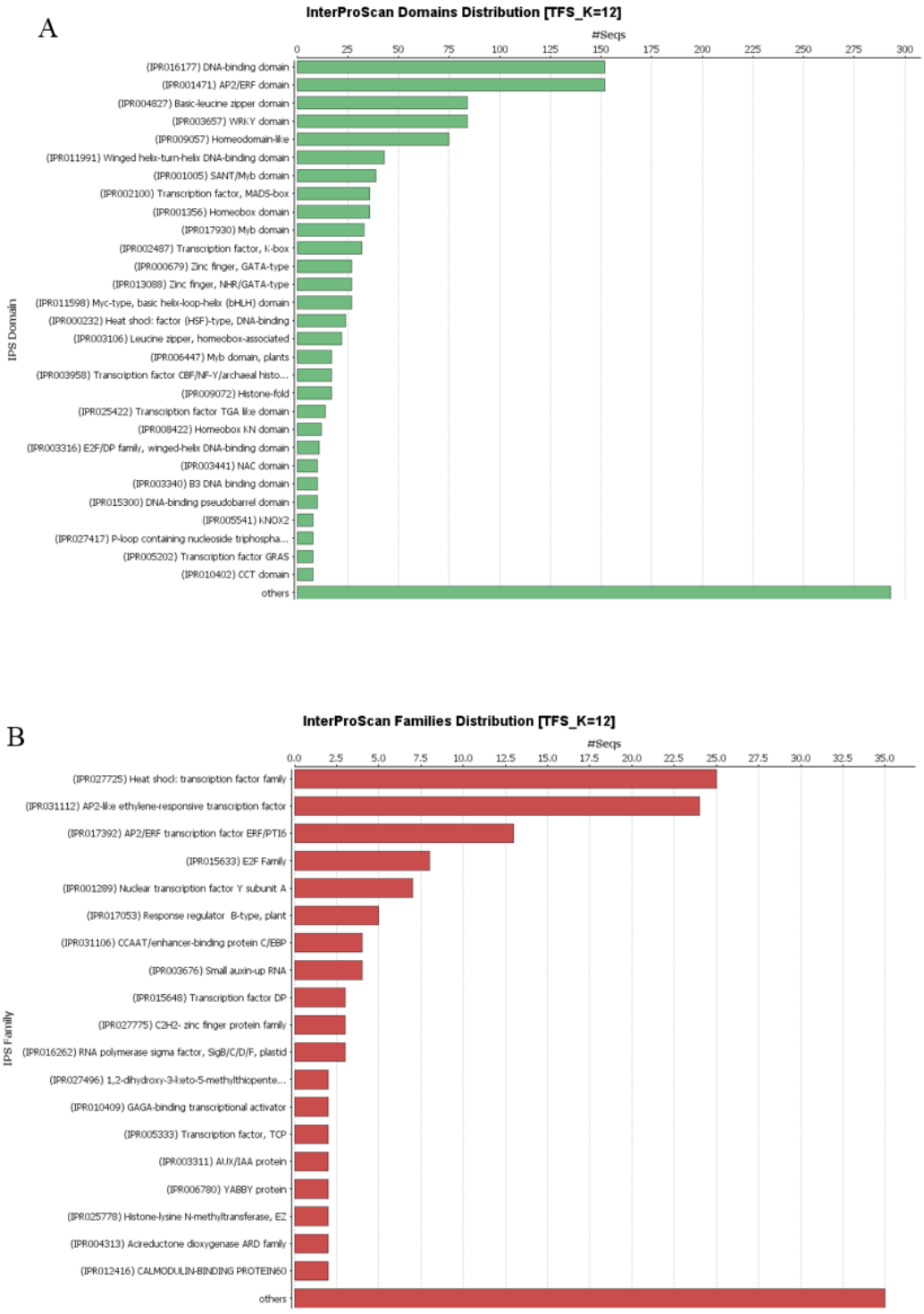


Figure 2.7: Transcription factors families and domain distribution by InterProScan. (A) Domains distributions showing the domains distributions such as AP2/ERF, WRKY, MADS-box, Myb and KNOX. (B) Families distributions showing the list of expected Families such as AP2 and *YABBY*. Y axis is the family name, and the X axis is the number of sequences associated with family types.

2.3.8 Validation of TF expression patterns and correlation with function

To validate the quality of our transcriptome and to illustrate its value as a resource, we took characterised TFs and examined their expression patterns relative to published data (**Figure 2.8**). We selected TFs involved, or proposed to be involved, in distinct processes throughout grain development: *MADS29* shown to be involved in nucellar tissue degeneration in rice (Yin and Xue, 2012); 13 single repeat *MYB* TFs related to the specifier of endosperm transfer cells in maize *ZmMRP1* (Gomez et al., 2002, Gomez et al., 2009), the *Brachypodium* orthologue of which has been shown to have a weaker expression during grain development compared to maize (Hands et al., 2012); *BLZ1* and *BLZ2* are *bZIP* TFs that regulate storage protein and starch metabolism (Onate et al., 1999, Vicente-Carbajosa et al., 1998); *DOF* TFs shown to regulate germination (González-Calle et al., 2014). Our transcriptomic data for the selected TFs matched published reports and also provided extra information. The *BdMADS29* expression was high but peaked later in grain development than in rice (Yin and Xue, 2012). This agrees with the developmental program in the *Brachypodium* grain as its nucellus persists (Opanowicz et al., 2011; Hands et al., 2012). To complement the transcriptomic data, TUNEL assays were performed on sections of developing *Brachypodium* grains to time the degradation of nucellar tissue. **Figure 2.8 (b-e)** shows strong signal in the nucellus in ML and FL grains which is in agreement with the gene expression data.

Previous analysis tested the expression of the *ZmMRP1* orthologue (Bradi1g72300) and showed little or no expression during grain development (Hands et al., 2012) – a result matching the transcriptomic data (**Figure 2.8a**). In addition, the transcriptome allowed testing of the remaining 12 single repeat MYBs to determine if *Brachypodium* had any related genes showing the distinct expression profile reminiscent of *ZmMRP1* during grain development (Gomez et al., 2002). Based on the expression profiles of the *MRP1*-like paralogues, none are likely candidates which fits with the lack of the transfer cells in *Brachypodium* grains.

BLZ1/2 and their orthologues regulate transcription of storage protein genes during grain filling in barley, wheat, maize and rice (Albani et al., 1997, Onate et al., 1999, Onodera et al., 2001, Schmidt et al., 1990). Previous expression analyses showed that while *BdBLZ1* was expressed throughout grain development as in the other species, BdBLZ2 was expressed at low levels in maturing grains in contrast to the high expression levels in the species with predominantly prolamin storage proteins (**Figure 2.8e**).

Finally, *BdDOF24* and *BdGAMYB* were shown to peak in expression levels at germination correlating with *CathepsinB* expression (González-Calle et al., 2014) – this matched transcriptomic data obtained in this analysis.

The examples above highlight the high quality and validity of this new transcriptomic resource and its value in complementing studies not only in *Brachypodium* but in other cereal crops as well. We envisage that it will be a useful tool for researchers working in various areas of seed/grain biology and germination.

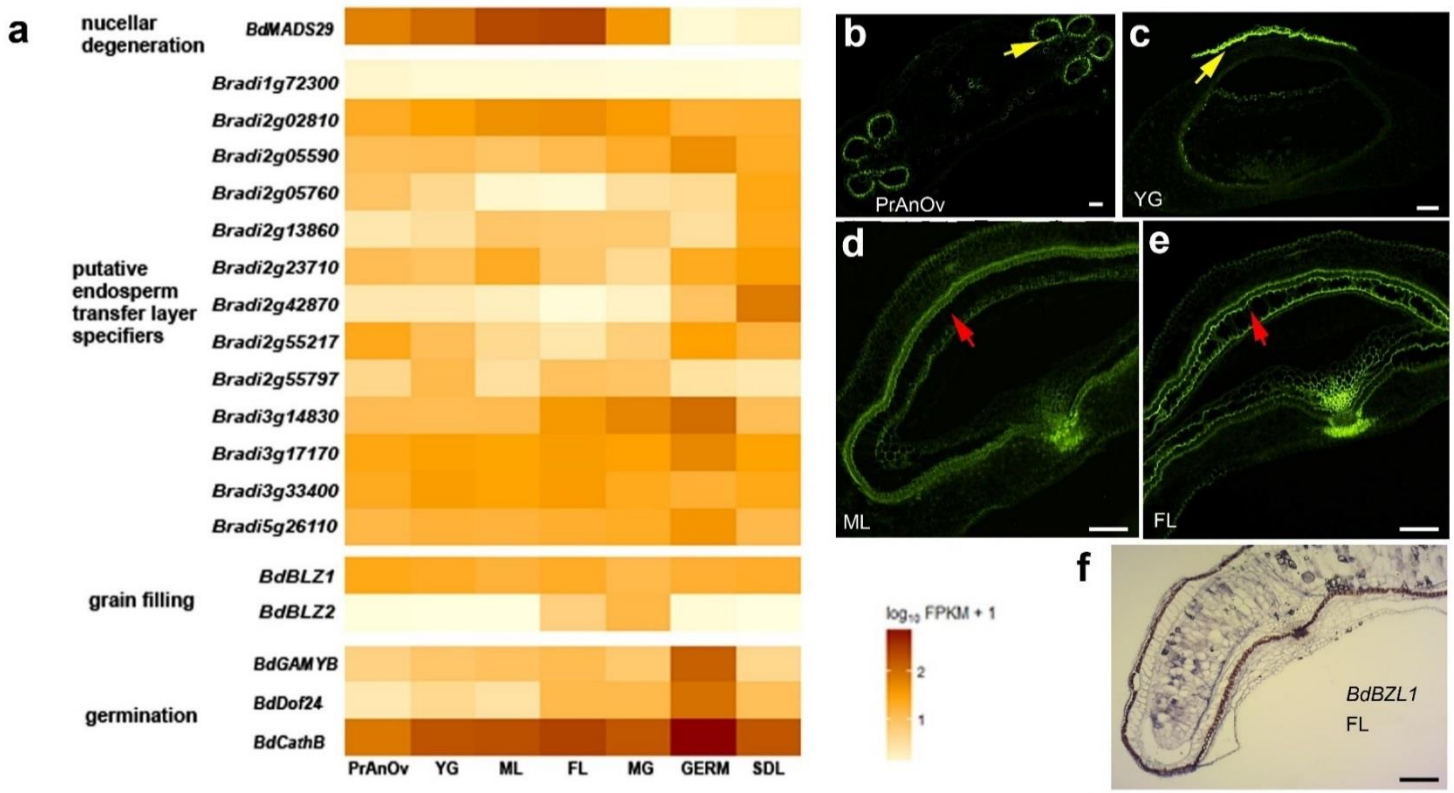


Figure 2.8: Validation of transcriptomic data for selected genes. a. Heatmap of RNA-seq data for 19 genes correlated with characterised functions in grain development. b-e. TUNEL assay of *Brachypodium* grain development, from pre-anthesis ovaries to full-length grain. Note the tapetal cells of the anther in b and palea in c – yellow arrows; persistent nucellar tissue visible throughout grain development - red arrows in d and e. f. In situ hybridisation of full-length grain section with a BLZ1 mRNA probe, showing BdBLZ1 expression throughout the endosperm. PrAnOv, pre-anthesis ovaries; YG, young grain (1-3 DAA); ML, mid-length grain (3-8 DAA); FL, full-length grain (8-15 DAA); MG, mature grain (15-20 DAA); GERM, germinating grain; SDL, young seedling (3-4 days after germination). Scale bars: 50 µm

2.4 Conclusion

The value of transcriptomic analyses are greatly enhanced when based on a detailed knowledge of the underlying biology of the subject, in this case the development of the grain of *Brachypodium distachyon*. The species, relatively recently developed as a model system for temperate cereals, has been subject to several informative studies involving development, biochemistry and evolution generally and grain biology more specifically (Catalan et al., 2016, Guillon et al., 2011, Larre et al., 2010, Laudencia-Chingcuanco and Vensel, 2008, Opanowicz et al., 2011, Trafford et al., 2013). The grains of *Brachypodium* have been shown to have features not considered typical of cereal grains, such as low starch content and thick cell walls in the central endosperm, while also sharing features with a range of well-studied species e.g., storage protein profile and cell wall composition with varied similarities to oat, rice, wheat and barley. With this underlying knowledge transcriptomic analyses such as the one described here can be an invaluable resource for the species, genus, family and beyond.

Chapter 3 – Evolution and expression profiling of YABBY genes

3.1 Introduction

Flowers are the reproductive structure produced by angiosperms for propagation in plants. They are typically made up of attractive petals with protective modified layers of sepals surrounding variable numbers of stamens and carpels producing pollen and ovules which are important for double fertilization in plants (Berger et al., 2008, Dumas and Rogowsky, 2008). However, Poaceae or Gramineae are monocotyledonous flowering plants, known as grasses, and have a distinctive morphological structure (Bommert et al., 2005). Instead of petals and sepals as in the common dicot flowers, grasses have palea/lemma and lodicules and glumes (**Figure 3.1**). (Schmidt and Ambrose, 1998, Goto et al., 2001).

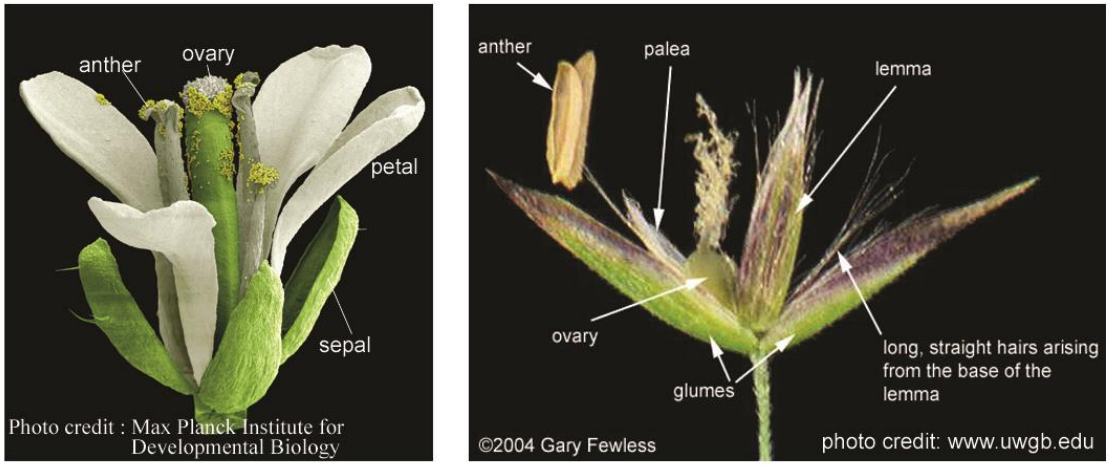


Figure 3.1: Morphological differences between eudicot and grass flowers. Left: flower of *Arabidopsis thaliana* (eudicot). Right: grass floret. Individual floral organs are indicated.

Within the Poaceae, *Brachypodium distachyon* has been developed as a model system over the last decade (Draper et al., 2001). As with *Arabidopsis*, it has a small physical size, small genome (~355 Mbp), short life cycle (8-12 weeks), is self-pollinated and has undemanding growth requirements in contrast to rice and wheat (Vogel and Bragg, 2009, Vogel, 2008). Consequently, much research has been reported to support *Brachypodium* resources and features as a model for Triticeae (Bevan et al., 2010, Steinwand et al., 2013,

Garvin et al., 2008, Huo et al., 2009, Christensen et al., 2010, Garvin, 2007, Olsen et al., 2006, Parker et al., 2008).

The distinctive floral morphology of grasses also encompasses the unique fruit form, the caryopsis or grain. Caryopses in grasses generally consist of the embryo, endosperm and pericarp tissues. The pericarp is the fruit component and embryo/endosperm, the seed. The pericarp in grasses is composed of three distinct layers, the epicarp, mesocarp and endocarp (Brillouet, 1987). Wheat pericarp has been reported to have many functions such as protection, photosynthesis, mineral accumulation, synthesis and degradation of starch (Xiong et al., 2013). During the early stages of wheat caryopsis development, the pericarp contains high numbers of chloroplasts and is rich in chlorophyll, generating valuable energy resources for seed respiration and growth. The stored starch in the pericarp layer is used for amylose synthesis. The pericarp cells die rapidly after the storage starch is fully consumed (Zhou et al., 2009).

In comparative developmental studies, a classic approach is to examine key regulatory genes (transcription factors) in species with distinctive morphologies and compare them at sequence, gene expression and protein function levels. This candidate gene approach is also employed here with a focus on the *YABBY* gene family and their role in patterning the grass flower and fruit. *YABBY* genes encode plant-specific transcription factors (TF) containing zinc finger (C2CX20C2C) and homeobox domains (termed as *YABBY* domain hereafter). So far, five clades have been identified in angiosperms; *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER* (*FIL*)/*YABBY3* (*YAB3*), *INNER NO OUTER* (*INO*), *YABBY2* (*YAB2*) and *YABBY5* (*YAB5*). However, previous studies in rice have shown that *YAB5* is missing in the grasses lineage (Toriba et al., 2007).

Phylogenetic analysis across plant lineages, within and beyond the angiosperms, is essential to understand the evolution of *YABBY* genes in plants. To date, extensive phylogenetic analyses have been undertaken in flowering plants (Yamada et al., 2011, Bartholmes et al., 2012, Eckardt, 2010), including the recent discovery of *YABBY* orthologs in gymnosperms (Finet et al., 2016). However, systematic and comprehensive analysis to date is very limited and currently published studies are lacking important information in specific groups of plants, such as the cereal crops including wheat, rye, rice, barley, and maize. This study is using *Brachypodium* as the model to infer the

evolution of the *YABBY* gene family to focus on the specific impact of the family's evolution within the grasses.

Gene expression and functional analyses to date have revealed that expression patterns and functions of some *YABBYs* are conserved across plant lineages (Yamada et al., 2011, Orashakova et al., 2009, Bowman and Smyth, 1999, Fourquin et al., 2007). However, a comprehensive investigation of *YABBY* genes throughout grain development and functional data on these genes in grasses is scarcely reported.

As an example, the ancestral functions of *CRC* in eudicots are carpel specification and floral meristem (FM) initiation (Yamada et al., 2011, Orashakova et al., 2009, Bowman and Smyth, 1999, Fourquin et al., 2007). The most studied *YABBY* gene in grasses, *DROOPING LEAF (DL)*, is an orthologue of *CRABS CLAW* (*CRC*). *CRC* in *Arabidopsis* is required for carpel formation and its expression is limited to the abaxial region of the carpel and the nectaries. In addition, the *crc* mutant carpels fail to fuse at the apex with reduced amount of style tissue and ovules (Bowman and Smyth, 1999). In contrast to *Arabidopsis*, *DROOPING LEAF* in rice is expressed in the entire carpel (in a non-polar manner) and also in the middle of undifferentiated cells of leaves, indicating its importance in the regulation of carpel identity and midrib formation in rice (Yamaguchi et al., 2004). The spatial expression pattern of *DL* orthologues in several grasses (maize, wheat and sorghum) suggested its function in determining carpel identity might be conserved in grasses but no explicit functional analysis has been performed to date (Ishikawa et al., 2009).

Several studies on eudicot *FIL* genes demonstrated that *FIL* is expressed in a polar manner and is important for plant organ identity (Chen et al., 1999, Bonaccorso et al., 2012, Vosnakis et al., 2012) In rice, three *FIL* homolog genes have been identified due to genome duplication events that have occurred in the grasses; *OsFIL1*, *OsFIL2* and *OsFIL3*. *OsFIL1,* also known as *TONGARI-BOUSHI,* is expressed in organ primordia in a non-polar manner and the mutant shows homoeotic transformation of carpels to stamens (Tanaka et al., 2012). *OsFIL3* is expressed in the cells that subsequently differentiate into sclerenchyma cells, while *OsFIL2* is hypothesized to regulate development of vasculature in rice (Liu et al., 2007, Toriba et al., 2007). The expression pattern and localization of

YABBYs in grasses is therefore different compared to eudicots. (Ishikawa et al., 2009, Nakayama et al., 2010, Yamaguchi et al., 2012, Yamaguchi et al., 2004).

Interestingly, in addition to polarity determination, *YABBY* genes have been reported to play a role in secondary metabolism in aromatic plants. *YABBY5* was highly expressed in leaf glandular trichomes in mint plants, which are important for secondary metabolites, whereas *YABBY5* in other eudicot species remains unknown or classified as having redundant functions with other *YABBY* genes (Wang et al., 2016). In addition, the regulatory region of *YABBY* genes brings a significant contribution during the evolution of crops domestication. As an example, the *FASCIATED (FAS)* gene is an important regulator of fruit shape size in tomato (Cong et al., 2008), while its orthologues in sorghum were responsible for shattering effects during its parallel domestication in various geographical locations (Lin et al., 2012).

Recent developments in the field of genomics and the relatively low cost of sequencing technologies have made it easier for scientists to use reverse genetics approaches to study the function of potential candidate genes. Gene expression analyses using RNA-seq help to identify co-expression networks and their gene ontology (GO) classifications. It has been demonstrated, that co-expressed genes are more likely to have similar biological functions (Alloco et al., 2004, Eisen et al., 1999).

In the present study, the evolution of *YABBY* genes is discussed in detail, with an emphasis in grasses. The spatial and temporal expression profile of *YABBY* genes is also determined and is compared to available gene expression databases for other grasses. Based on preliminary findings, it is suggested that one of the *YABBY* genes, *YAB6*, is an important regulator of grain development in grasses.

3.3 Materials and Methods

3.3.1 Plant material

Brachypodium distachyon grains (Bd-21 accession) were germinated on moist filter paper after stratification at 4°C for 48h. Five-day old seedlings were transferred to 9 cm square pots with a 2:2:1 multipurpose compost: vermiculite: sand mix and grown under controlled environment conditions with a 18h photoperiod at 20°-22°C and light intensity of 180-200 umol/m/s.

3.3.2 Phylogenetic analysis

The YABBY orthologues of *Brachypodium* and other species were obtained by BlastP search using known rice, maize, and *Arabidopsis* YABBY proteins as queries against the relevant selected proteomes available in Phytozome 12, Ensemble plants and Congenie websites. The orthologoes of the genes listed in **Appendix 3.1**.

Multiple sequence alignment (MSA) were performed using Mafft (L-INS-i) (Nuin et al., 2006) and edited manually using the Geneious software (Geneious 9.0.2) (Kearse et al., 2012). Prior to Bayesian analysis, aligned gene sequences were exported to MEGA 6 to determine the overall mean distance. A p-distance <0.8 was used in order to ensure that a reliable tree was computed. The Jones-Taylor-Thornton (JTT) amino acid substitution model was used with a discrete Gamma distribution (+G).

3.3.3 Tissue collection, disruption and storage

All plant material used (**Appendix 3.2**) for DNA or RNA extraction was snap frozen in liquid nitrogen straight after collection and stored at -80°C. Frozen tissue samples were typically ground using a Silamat amalgam mixer (Ivoclar Vivadent, UK) for 12 sec with acid-washed glass beads (SIGMA), or for more robust spikelet samples, using a pestle and mortar.

Samples were collected from different plant and grain stages; vegetative stage (2 days after germination, DAG) mature leaf, very young inflorescence (13-18 days after sowing (DAS), emergent spikelet (25-28 DAS), pre-anthesis floret (28-32 DAS), young grain (2-

3 days after anthesis, DAA), middle length (4-8 DAA), full length (9-15 DAA) and mature grain (18- 20 DAA). Anthesis in *Brachypodium* occurs around 32-34 DAS.

3.3.3.1 Genomic DNA isolation

Genomic DNA extraction was performed using approximately 300 mg of young leaf tissue using a typical phenol:chloroform extraction protocol (Kang and Yang, 2004). Ground tissues were incubated in 400 µl extraction buffer at 50°C for 20-30 min, followed by a phenol:chloroform extraction and precipitation using isopropanol. DNA pellets were washed with 70% ethanol and re-suspended in sterile water for direct use in PCR reactions.

3.3.3.2 Total RNA isolation

Total RNA was extracted from frozen ground tissues using TriSure (Bioline) according to manufacturer's instructions. For grain samples that contain more starch, the Spectrum Plant Total RNA extraction kit (Sigma) was used. For the majority of RNA samples extracted, DNase treatment was performed using DNase I (NEB), in 1x reaction buffer (10mM Tris-HCl, 2.5mM $MgCl_2$, 0.5mM $CaCl_2$, pH7.6), with a 10 min incubation at 37°C and enzyme deactivation at 75°C for 10 min.

3.3.3.3 RNA quantification and cDNA synthesis

RNA quantification was performed using NanoDrop 2000/2000c (Thermo Scientific). Approximately 700 ng RNA was used in 20 µl cDNA synthesis reactions using Tetro Reverse Transcriptase (Bioline) and a poly (T) primer 5' GACTCGAGTCGACATCGA (T_{50}), according to manufacturer's instructions.

3.3.3.4 Reverse Transcription – Polymerase Chain Reaction (RT-PCR)

The synthesized cDNA from various tissue samples was used with gene-specific primers in 10 µl reactions, as detailed in **Table 3.1**. The sequences of the primers used can be found in **Appendix 3.3.**

Table 3.1: RT-PCR reactions and thermocycling program to investigate gene expression

PCR reaction	Solutions	Quantity (µl)
	PCR mix	5.0 µl
	Sterile water	3.0 µl
	10 µM Forward primer (F)	0.5 µl
	10 µM Reverse primer (R)	0.5 µl
	cDNA	1.0 µl

PCR program		Temperature (°C)	Time (min)
Initial denaturation		94	05:00
Denaturation		94	00:30
Annealing	30 cycles	55	00:30
Extension		72	00:30
Final extension		72	06:00
Hold		10	HOLD

3.3.3.5 Gel electrophoresis

PCR products were run on 1% agarose (Molecular grade, Bioline) gel with 0.5 µg/ml EtBr, and electrophoresis was performed at 95 V (Enduro Power Supplies, 300V) for one hour alongside 100 bp DNA ladders (NEB). Gels were examined using a benchtop UV transilluminator (UVP Biodoc-IT Imaging System). The picture was calibrated using the "Touchman" calibration software and printed using a SONY Digital Graphic Printer, UP-D987.

3.3.4 Tissue collection, fixation and processing

Plant tissue samples were excised and immediately immerged in FAA (3.7% formaldehyde, 5% acetic acid, 50% ethanol). Samples were vacuum infiltrated in 3x5 min cycles, and fixed overnight at 4°C with continuous agitation. Tissues were then washed with 70% EtOH and stored at 4°C. Samples were processed as follows; 70% ethanol (EtOH) for 1 hour at 4°C; 85% EtOH for 1 hour at 4°C; 95% EtOH for 1 hour at 4°C; 100% EtOH for 1 hour at 4°C; 100% EtOH for 1 hour at room temperature (RT); 25% of Histoclear (Agar Scientific) (H/C) in EtOH for ½ hour at RT; 50% H/C in EtOH for 1/2 hour at RT; 75% H/C in EtOH for ½ hour at RT; 100% H/C overnight (OV) at RT; 100% H/C for 1 hour at 60°C; 50% H/C in paraffin for 4 hours at 60°C; 100% paraffin for OV

at 60°C; 100% fresh paraffin for 8 hours at 60°C and a final change with 100% paraffin for OV at 60°C.

3.3.4.1 Sectioning and embedding

Processed samples were embedded in paraffin using Tissue Embedding Molds (Polyscience, USA). The blocks were cut into a trapezoid shape with the crease of samples facing the shortest edge. A Bright 5030 (Bright instrument Co., England) microtome was used for sectioning. Sections were normally cut in a thickness of 10-14 µm, organized sequentially on poly-lysine-coated slides (Grace Biolab, Supplied by Stratech Scientific, Soham, UK) on a drop of distilled water and were dried down at 38°C.

3.3.4.2 Toluidine Blue Staining

Glass slides containing sample paraffin sections were washed twice with 100% histoclear for 20 min to clear the paraffin. Tissue rehydration was performed using a series of 100%, 80% and 50% EtOH for 5 min each. Rehydrated sections were rinsed with sterile water and soaked in 0.05% toluidine blue in 0.1M Phosphate buffer (pH 6.8) for 30 sec. The slides were then allowed to completely dry, and were subsequently mounted with Entellan (proscitech) and left overnight in the fume hood to dry.

3.3.5 mRNA *in-situ* hybridization (ISH)

3.3.5.1 RNA probe synthesis

The probe template was produced by PCR using cDNA as template and gene-specific primers. Primers were designed amplifying unique region of each genes, prioritizing the UTR region, if not available, then the non—conserved region of the genes. The reverse primers incorporated a T7 promoter region for subsequent *in-vitro* transcription. A Proof Reading Polymerase was used (Q5, NEB) in a reaction and PCR cycle as listed in **Table 3.2**. The PCR product was purified using a PCR Purification kit (supplied by Bioline), according to manufacturers' instruction and was used as a template for the *in-vitro* transcription of the mRNA probe described below.

Table 3.2: PCR reaction and thermocycling conditions for mRNA probe template synthesis

PCR reaction (50 µl)	Volume
H_2O	21.5 µl
5x Q5 Reaction Buffer	10.0 µl
GC Enhancer	10.0 µl
Forward primer with enzyme attachment (10 µM)	2.0 µl
Reverse primer with enzyme attachment (10 µM)	2.0 µl
cDNA template	3.0 µl
dNTPs (10 µM)	1.0 µl
Q5 High-Fidelity DNA Polymerase	0.5 µl

PCR Themocycling Conditions

Step		TEMPERATURE (∘C)	TIME (min)
Initial denaturation		98	00:30
Denaturation		98	00:10
Annealing	30 cycles	55	00:20
Extension		72	00:30
Final Extension		72	02:00
Hold		10	∞

3.3.5.2 In-vitro transcription

Probe transcription and DIG-labelling was performed in 20 µl reactions as shown in **Table 3.3**, using as template the purified PCR product described above.

Table 3.3: *In-vitro* transcription of mRNA probe

20 µl reaction mixure for in-vitro transcription	
Solutions	Quantity (µl)
H2O (DEPC)	5.0 µl
Txn NTP mix (10x)	2.0 µl
Dig-UTP	2.0 µl
RNAse inhibitor	1.0 µl
T7 polymerase	2.0 µl
10x Buffer (Txn 10x)	2.0 µl
Probe template (Purified DNA)	6.0 µl

The assembled reactions were incubated at 37°C for 2 hours before 2 µl 4M LiCl and 66 µl of ice-ethanol were added. The tubes were left at -20 °C overnight and were centrifuged at 11.4 k for 10 min at 4 °C. The supernatant was removed and the pellet was air-dried and re-suspended in 30-50 µl of TE buffer. RNA probes were stored at -80 °C until used.

3.3.5.3 Dot blot Probe Testing

In order to confirm that the probes were successfully transcribed and provide a crude estimate of "strength", a dot blot was performed. The probes were diluted to 1:100 (0.5-50 µl) using DEPC-treated water (dH2O). A 1 µl dot for every probe was applied on a suitable size of hybond membrane, and the membrane was baked at 50 °C for a few minutes until it completely dried to fix it. The membrane was then soaked in 1xTBS buffer and drained before adding the blocking solution (0.1 g milk powder in 10ml TBS) and leaving it in gentle agitation for 30 min. The drained membrane was then soaked in 1:5000 dilution of Anti-DIG antibody solution (Sigma Aldrich) (1 µl to 5 ml TBS) for 30 min. The antibody solution was then drained and the membrane was rinsed with 1xTBS for 5 min. AP buffer was then added for 5 min and drained before the final developing solution was added (15 µl NBT- Nitroblue Tetrazoluim salt, supplier by Sigma Aldrich (0.2 mM) and 15 µl BCIP - 5-bromo-4-chloro3-indolyl phosphate (0.2 mM) in 10ml AP Buffer). The colour was allowed to develop in darkness. The development progress was checked after 15 min and regularly thereafter to gauge strength of probes and speed of development relative to the controls.

3.3.5.4 Slide Pre-treatment

Slide pre-treatment was performed following the protocol of Drea et.al, (2005) as follows: paraffin clearing with histoclear (twice) for 10 min, rehydration with EtOH (100%, 80%, and 50%) for 5 min each, respectively; slides were rinsed with PBS buffer twice for 5 min. Proteins adhered to RNA were removed using 2 µg/ml Proteinase K (Sigma Aldrich) for 30 min at 37 °C. Proteinase K activity was stopped by using a Glycine PBS (10 %) solution for 2 min, and subsequently rinsing with PBS for 5 min. The slides were then neutralized by TEA buffer and Acetic anhydride stirred for 5 min before the final rinse with PBS for 5 min. Sections were dehydrated again using the same series of EtOH as for rehydration in reverse order and were left to air-dry at RT. Pre-treated slides were stored in air tight container at 4 °C until hybridisation.

3.3.5.5 Hybridisation and Washing

Hybridisation and washing steps were performed using the protocol described by Drea et al. (2005). In this project, 180 µl of hybridisation solution (HS) (**Table 3.4**) was used for each slide. Firstly, the appropriate needed volume was mixed by vortex (Labnet) in eppi tubes then subjected to denaturation at 80°C for 2-3 min and transferred immediately on ice. The HS was carefully applied on the slides (arranged in a hybridisation box) and covered with a glass cover slip (22x50 mm, 0.13-0.17 mm thickness). The hybridisation box was left at 50-55 °C overnight under high humidity.

Table 3.4: *In situ* Hybridisation Solution (HS) composition

Hybridisation Solution	Ratio	Quantity (µl) for 1 slides
10x in situ salts	10	22.5
Deionized formamide	40	45
50 % Dextran sulfate	20	45
50 x Denhardts solution	2	4.5
tRNA	1	2.25
H2O	7	15.75
		total = 180 µl

After the overnight hybridisation, the glass cover slips were carefully separated from the slides by dipping into pre-warmed (55 °C) SSC buffer. The slides were then transferred into suitable slide containers, covered with pre-warmed SSC and left at 55 °C for 1 hour with gentle agitation. This step was repeated 3 times with fresh SSC solution. The slides were then covered in 1 % blocking solution for at least 45 min at RT. The same conditions were applied to the next change of BSA solution and lastly the antibody solution. The formula for blocking and antibody solution as described below;

Blocking solution = 1 % solution in TBS (0.1 g milk powder in 10 ml TBS)

Antibody solution = **1:3000** antibody dilution in 1xTBS + 0.05%Triton (3 µl antibody + 0.1g BSA + 30 µl Triton to 15 ml TBA)

3.3.5.6 Slide Development

Slide development was conducted after the hybridisation and washing steps. The slides were soaked in AP buffer for environment equilibration for 2-3 min at RT. The slides were then soaked in developing solution (15 µl NBT- Nitroblue Tetrazoluim salt, supplier

by Sigma Aldrich (0.2 mM) and 15 µl BCIP - 5-bromo-4-chloro3-indolyl phosphate (0.2 mM) in 10ml AP Buffer) and were left in darkness. The developing time varied among different genes; hence the adequate time and rational judgement were crucial to decide the suitable time to stop the developing process. Reactions were allowed to run to a point where sections appeared slightly "over-developed" as the following rinsing step would wash off a small amount of the colour product.

3.3.5.7 Finishing and Mounting

The developed slides were rinsed with sterile water for several times. Then the slides were soaked for 3- 4 mins in EtOH (50%, 70% and 100%, respectively). A careful observation needed to make sure the EtOH was not over-washing the slides, and the time needed sometimes was less than described in the reference protocol. Subsequently, the slides were left to completely dry and were mounted with a xylene-based mounting medium (e.g. Entallan) and left to dry overnight in a fume hood. Pictures were taken using a compound microscope Gx L3200B, Gx camera and Image Focus software.

3.4 Results

3.4.1 Sequence identification and alignment

To identify the *Brachypodium* YABBY proteins, the known consensus protein domains from rice and *Arabidopsis* were used as queries to perform a BLAST search against the *Brachypodium* proteome database (http://blast.*Brachypodium*.org). Eight non-redundant proteins were identified and their exon structures and chromosome locations are shown in **Figure 3.2**. Amongst the YABBY proteins, DL and the three FILs consist of seven exons, while INO, YAB1, YAB6 of six.

All YABBY proteins from *Brachypodium*, rice, maize, wheat, *Arabidopsis* and tomato were aligned. Two conserved domains were identified and showed high similarity across monocots and eudicots; a zing finger domain and a homeobox domain (**Figure 3.3A**). The YABBY zinc finger domain has been shown to play a role in DNA binding and protein-protein interactions. It is a C2C2 type domain and consists of 48-50 amino acids in both monocots and dicots (**Figure 3.3B**).

The second YABBY domain is similar to the High Mobility Group (HMG) functioning as transcription factors in eukaryotes. This domain consists of a helix-loop-helix structure and is repeated as an array in the conserved domain. This domain is highly conserved in YABBY proteins across monocots and dicots (**Figure 3.3C**).

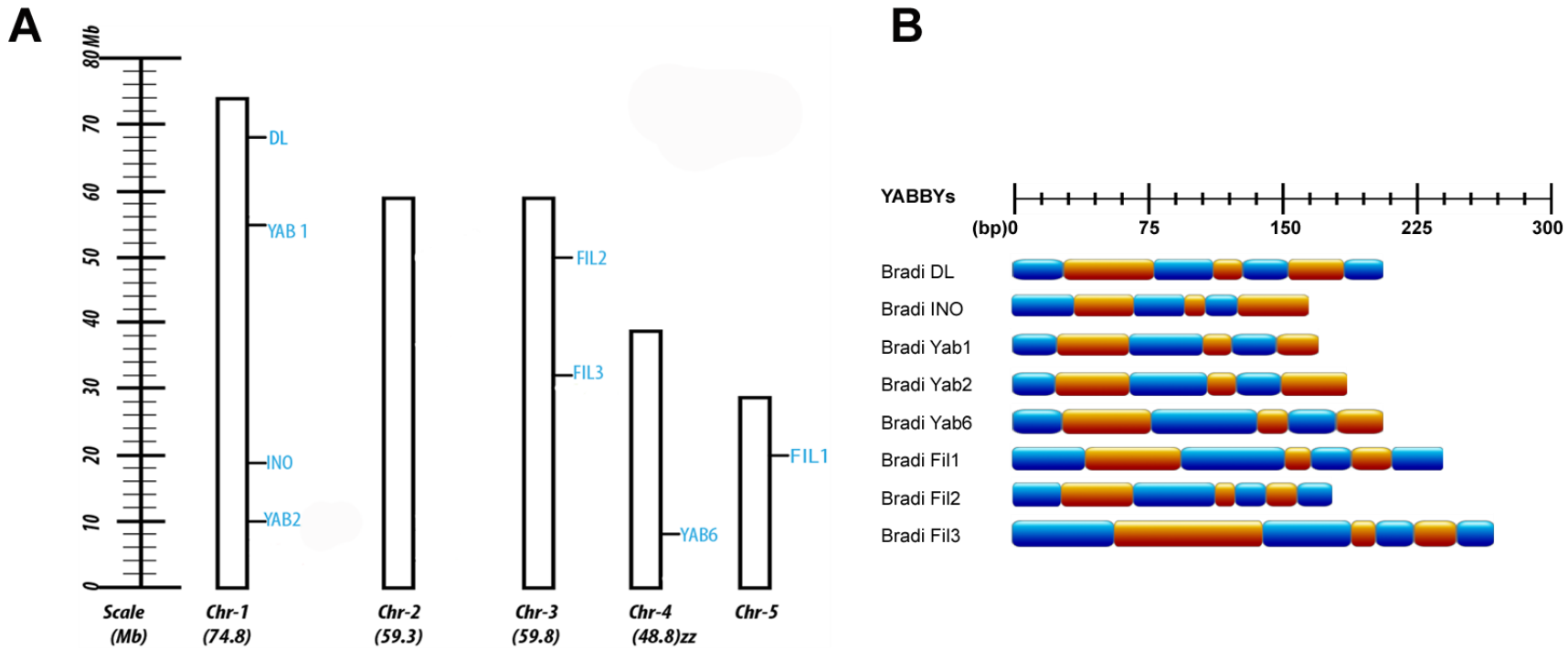


Figure 3.2: Chromosomal location and exon structure of *Brachypodium YABBY* genes. (A) The position of *YABBY* genes in the *Brachypodium* chromosomes. The number at the bottom of each chromosome bar indicates the chromosome length and the scale bar can be used for the position number of each gene. (B) The number of exons that constitute each YABBY CDS. The different colour indicates different exon and the length can be predicted from the scale.

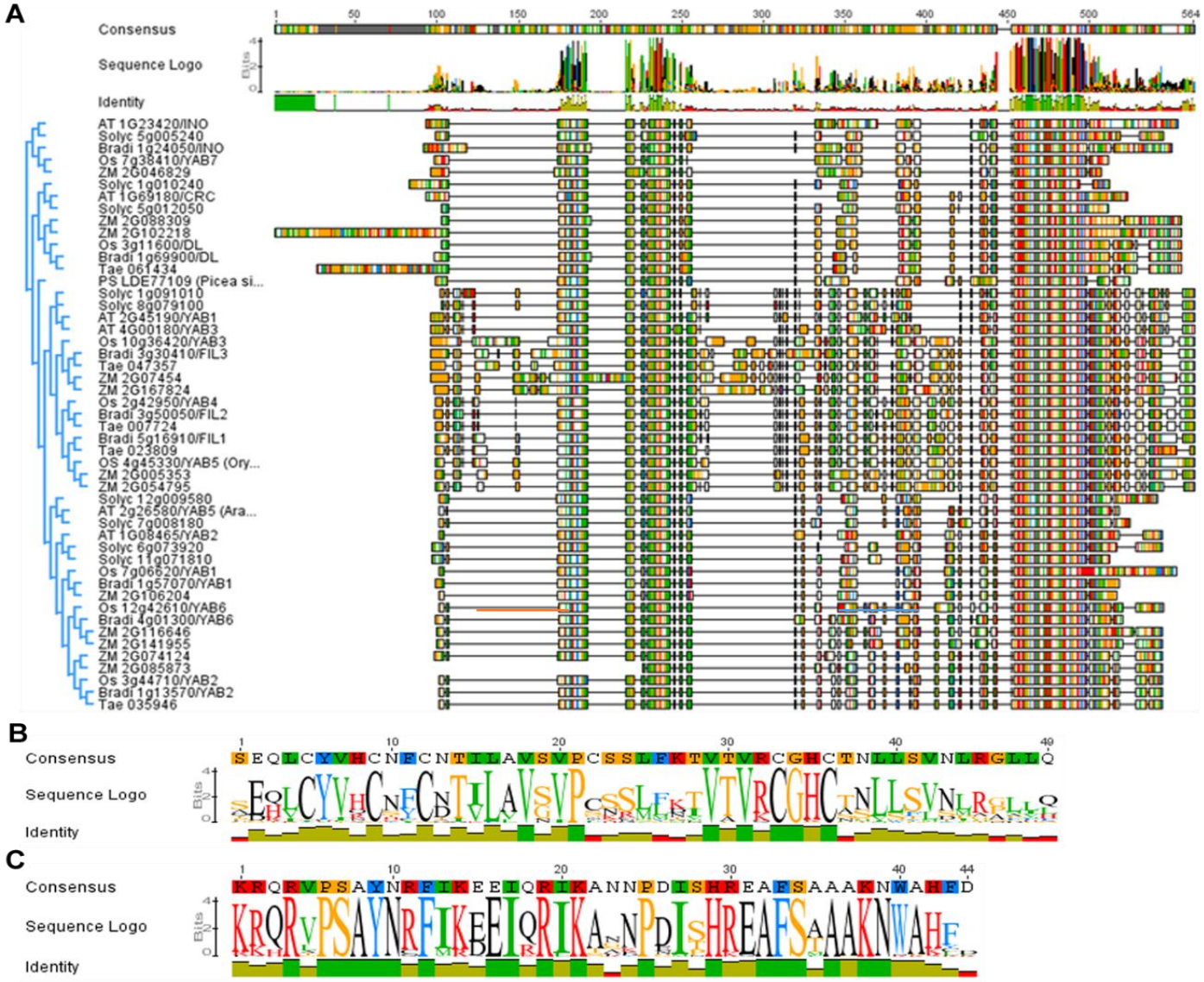


Figure 3.3: YABBY family protein alignment and conserved domains. (A) Sequence alignment of YABBY proteins from various monocot and dicot species. (B) Sequence logo of Zinc finger domain. (C) Homeobox/YABBY domain sequence logo. Identity bar-Green; residue at the position is the same across all sequences, yellow ; less than complete identity, red ;very low identity for the given position

3.4.2 Phylogenetic analyses

All YABBY homologs from *Arabidopsis*, rice and maize were used as queries in BLAST searches against other species, resulting in 488 sequences which were subsequently subjected to Bayesian analysis to infer phylogeny. Average pairwise amino acid identity was checked in order to estimate a reliable tree. Average p distance <0.8 was accepted and subsequently used for Bayesian Interference (BI) analysis.

Based on gene distributions and phylogeny it was possible to identify duplications and losses of *YABBY* genes during plant evolution. *Brachypodium* has eight orthologues, the same number as rice, but other grass species such as wheat, maize, barley, sorghum differ in numbers (**Table 3.5**).

Table 3.5: *YABBY* gene distribution in available sequenced genomes of grasses. The genes were divided according to their taxonomic tribes, and the orthologue assignment was based on their position in the phylogenetic analysis.

				FIL			YABBY2		
	DL	INO	FIL1	FIL2	FIL3	YAB1	YAB2	YAB6	Total
Oryzae									
Rice	1	1	1	1	1	1	1	1	8
Panicoideae									
Maize	2	1	2	1	2	1	2	1	12
Sorghum	1	1	1	1	1	1	1	1	8
Panicum virgatum	3	2	2	2	1	1	2	2	15
Panicum hallii	1	1	1	1	1	1	1	1	8
Setaria italica	1	1	1	1	1	1	1	1	8
Brachypodieae									
Brachypodium distachyon	1	1	1	1	1	1	1	1	8
Brachypodium stacei	1	1	1	1	1	1	1	1	8
Triticeae									
Wheat	3	1	3	3	3	3	3	3	22
Barley	1	-	-	2	1	-	1	1	6

The resulting tree suggests that the *YABBY* family is highly supported (>0.75) to be grouped into four clades (**Figure 3.4**). Paraphyletic clades were formed between *YABBY5* and *YABBY2*, which is consistent with previous findings (Bartholmes et al., 2012). However, in some species such as spruce (gymnosperm) and banana (monocot), the *YABBY* homologs cannot be separated into four clades. The spruce genes were clustered together forming a sister clade to the eudicot *YABBY5*. There were no *INO* orthologues identified in banana and most of *Musa YABBYs* clustered together within the *DL/CRC* clade. Only two *YABBY*-like genes were found in green algae in two different strains of *Micromonas pusilla*. Both of these genes have a homeobox/YABBY-like domain, but

only one of them contains both the homeobox/*YABBY* and the zinc finger domains characteristic of *YABBY* genes (**Figure 3.5** - asterisk). In addition, the algae genes do not fit into any of the five clades, suggesting that the subfamilies of *YABBY* genes evolved in the land plants after the green algae.

A grass-only phylogenetic tree was generated and was used to investigate further where *Brachypodium YABBY* genes are positioned in relation to the other grass species. As expected, all *Brachypodium YABBYs* (**Figure 3.6** - blue stars) are more closely related to wheat, compared to maize and rice. The *YABBY2* clade indicates that *YAB2* and *YAB6* evolved later than *YAB1* in the grasses lineage, with a highly supported posterior probability (0.88). The *FIL* clade shows the highest confidence level (1) and also suggests that *FIL1* and *FIL2* are more closely related compared to *FIL3*.

The use of outgroup sequences is crucial in evaluating a reliable ancestral direction for in-group tests in phylogenetic analyses. The outgroup sequences must have diverged before the in-group sequences diverged from each other, but not so much that homology is not detectable (Hall, 2011). The *YABBY* phylogeny generated in previous studies used as outgroup available gymnosperm sequence data, resulting in a very low confidence between the position of *YAB5* and *YAB2* in the trees (Bartholmes et al., 2012). From this study, using algae as an outgroup resulting in higher posterior probability confidence, especially the sisters' clade of YABBY5 and YABBY2. However, further validations and more species need to be included to propose algae usage as an outgroup for YABBY phylogeny.

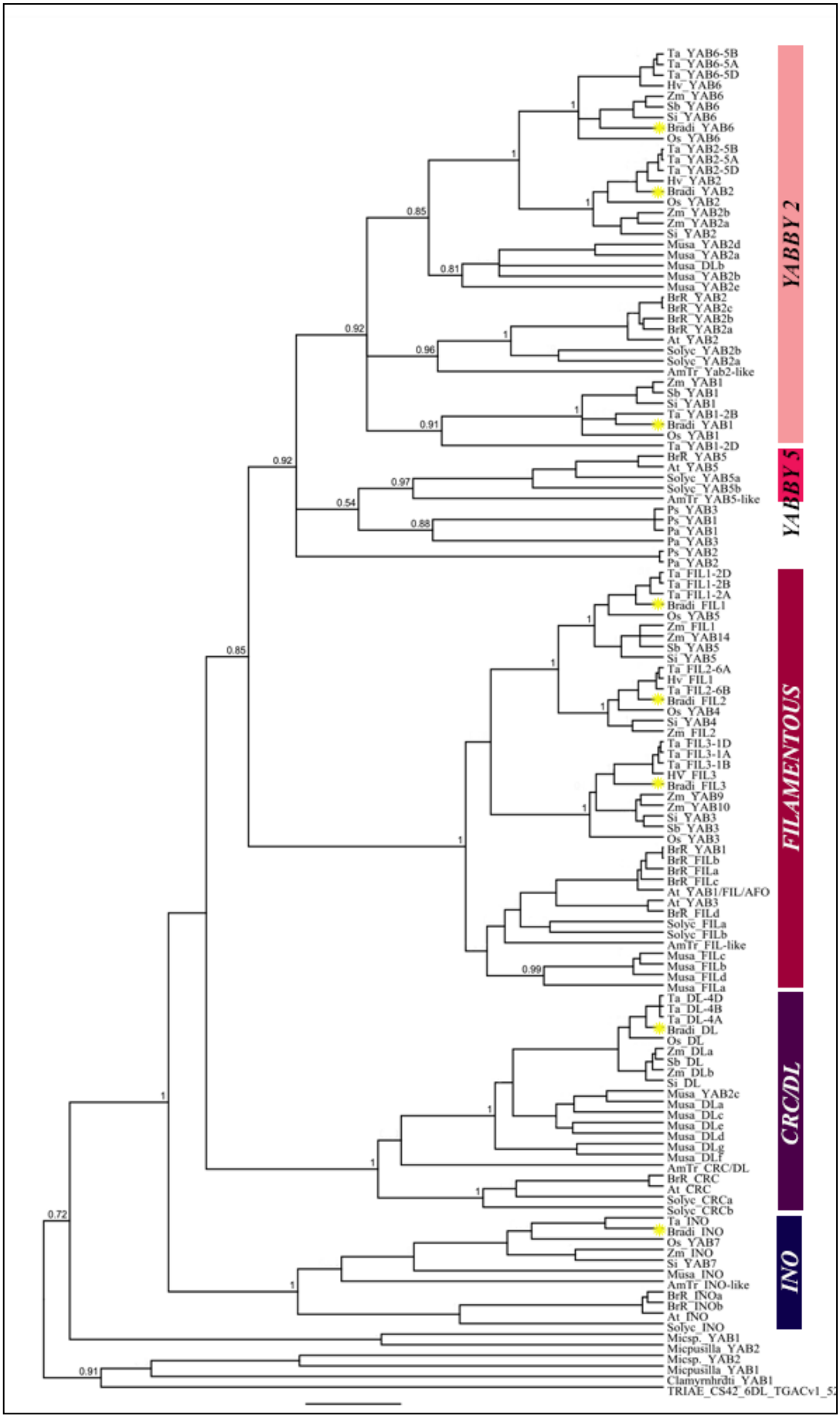

Ta_YAB6-5B
Ta_YAB6-5A
Ta_YAB6-5D
Hv_YAB6
Zm_YAB6
Sb_YAB6
Si_YAB6
Bradi_YAB6
Os_YAB6
Ta_YAB2-5B
Ta_YAB2-5A
Ta_YAB2-5D
Hv_YAB2
Bradi_YAB2
Os_YAB2
Zm_YAB2b
Zm_YAB2a
Si_YAB2
Musa_YAB2d
Musa_YAB2a
Musa_DLb
Musa_YAB2b
Musa_YAB2e
BrR_YAB2
BrR_YAB2c
BrR_YAB2b
BrR_YAB2a
At_YAB2
Solyc_YAB2b
Solyc_YAB2a
AmTr_Yab2-like
Zm_YAB1
Sb_YAB1
Si_YAB1
Ta_YAB1-2B
Bradi_YAB1
Os_YAB1
Ta_YAB1-2D
BrR_YAB5
At_YAB5
Solyc_YAB5a
Solyc_YAB5b
AmTr_YAB5-like
Ps_YAB3
Ps_YAB1
Pa_YAB1
Pa_YAB3
Ps_YAB2
Pa_YAB2
Ta_FIL1-2D
Ta_FIL1-2B
Ta_FIL1-2A
Bradi_FIL1
Os_YAB5
Zm_FIL1
Zm_YAB14
Sb_YAB5
Si_YAB5
Ta_FIL2-6A
Hv_FIL1
Ta_FIL2-6B
Bradi_FIL2
Os_YAB4
Si_YAB4
Zm_FIL2
Ta_FIL3-1D
Ta_FIL3-1A
Ta_FIL3-1B
HV_FIL3
Bradi_FIL3
Zm_YAB9
Zm_YAB10
Si_YAB3
Sb_YAB3
Os_YAB3
BrR_YAB1
BrR_FILb
BrR_FILa
BrR_FILc
At_YAB1/FIL/AFO
At_YAB3
BrR_FILd
Solyc_FILa
Solyc_FILb
AmTr_FIL-like
Musa_FILc
Musa_FILb
Musa_FILd
Musa_FILa
Ta_DL-4D
Ta_DL-4B
Ta_DL-4A
Bradi_DL
Os_DL
Zm_DLa
Sb_DL
Zm_DLb
Si_DL
Musa_YAB2c
Musa_DLa
Musa_DLc
Musa_DLe
Musa_DLd
Musa_DLg
Musa_DLf
AmTr_CRC/DL
BrR_CRC
At_CRC
Solyc_CRCa
Solyc_CRCb
Ta_INO
Bradi_INO
Os_YAB7
Zm_INO
Si_YAB7
Musa_INO
AmTr_INO-like
BrR_INOa
BrR_INOb
At_INO
Solyc_INO
Micsp._YAB1
Micpusilla_YAB2
Micsp._YAB2
Micpusilla_YAB1
Clamyrnhrdti_YAB1
TRIAE_CS42_6DL_TGACv1_52
YABBY 2
YABBY 5
FILAMENTOUS
CRC/DL
INO

Figure 3.4: The phylogeny of YABBY genes in plant lineages. Bayesian interference analysis of the YABBY gene family evolution across plant lineages. Yellow stars indicate the *Brachypodium* YABBYs. (Tree ID and all geneIDs attached in **Appendix 3.1**).

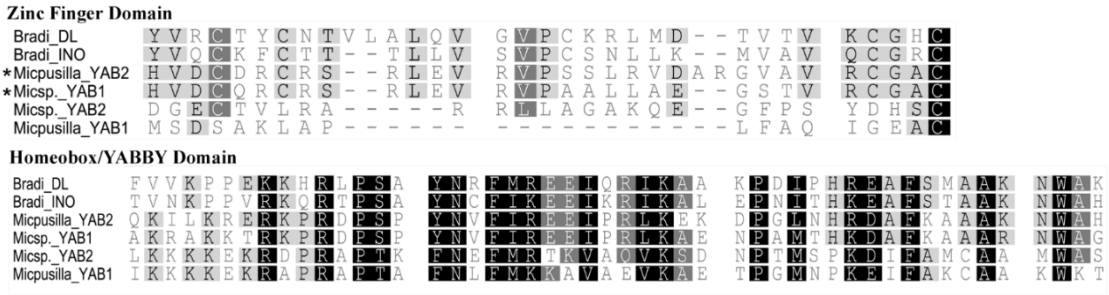


Figure 3.5: Alignment of YABBY-like proteins in algae. *Brachypodium* INO and DL proteins were used as reference. Asterisk- YABBY-like proteins in algae.

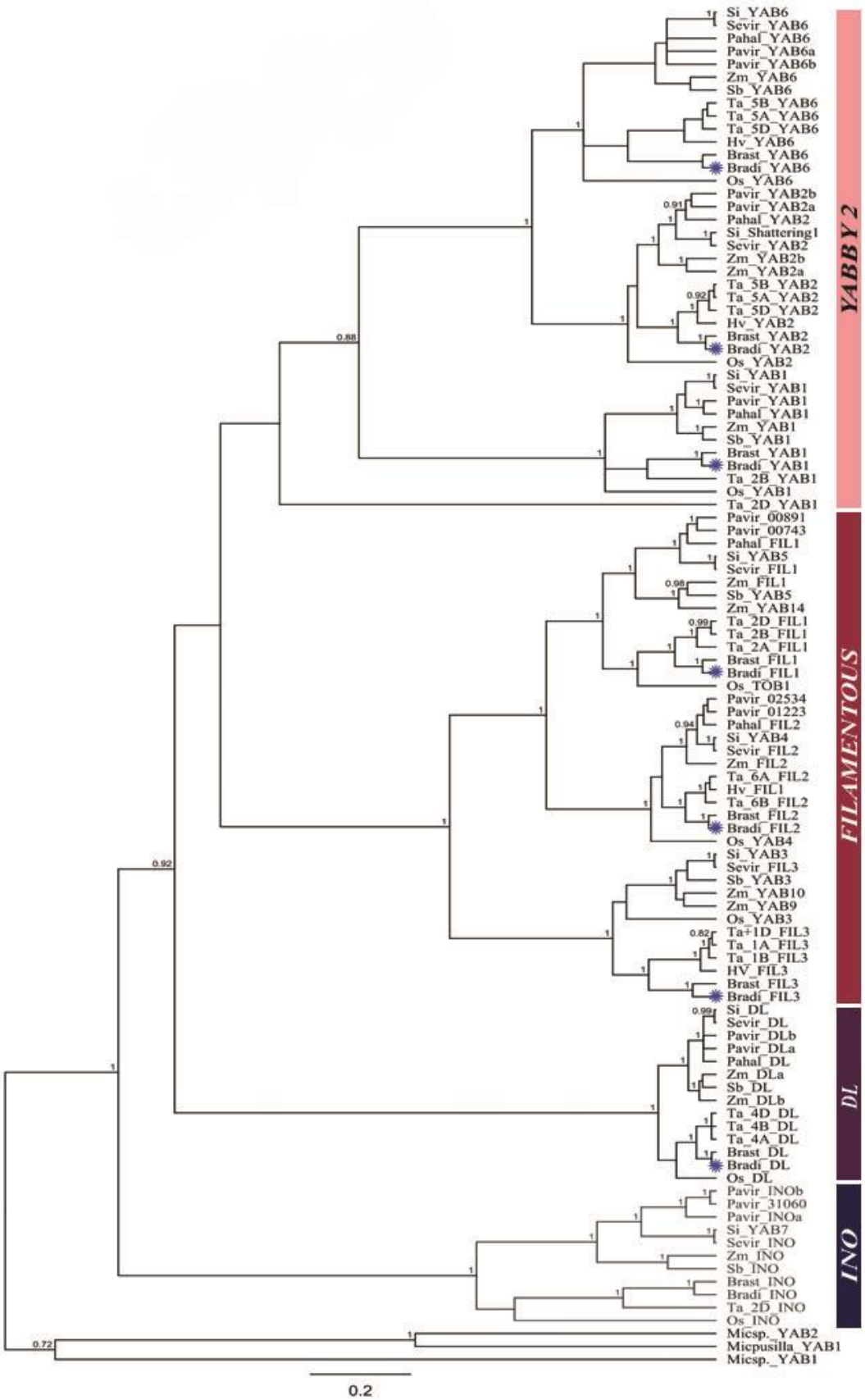


Figure 3.6: Evolution of *YABBY* genes in grasses. Bayesian interference analysis of the *YABBY* gene family evolution within the Poaceae. Blue stars indicate the *Brachypodium YABBYs*.

3.4.3 Temporal and spatial pattern of *YABBY* genes

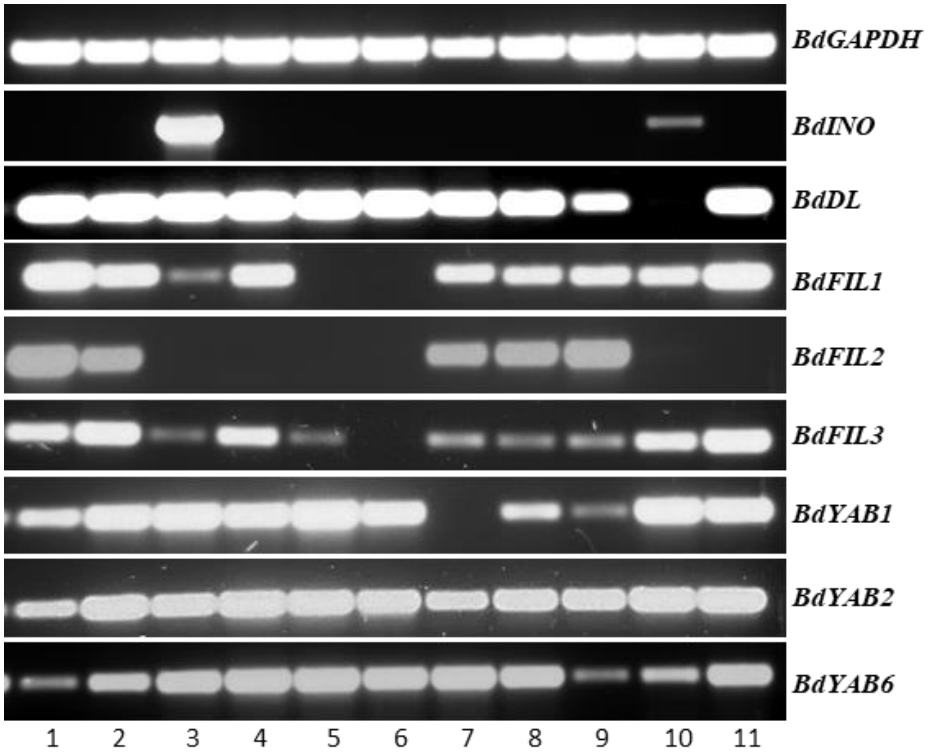
The organ expression pattern of *YABBY* genes in *Brachypodium*, was determined using RT-PCR with *YABBY*-specific primers and transcriptome. RT-PCR expression pattern **(Figure 3.7)** was similar to transcriptome expression profile **(Figure 3.8)**, validates the RNA-seq results. *YABBY* genes were found to be expressed across plant development, except for *BdINO* which seems to be restricted to pre-anthesis ovary and palea tissues as shown in **(Figure 3.7).** The expression in the palea is interesting because detailed analyses of the *Arabidopsis INO* have shown that it is exclusively expressed in the outer integuments and that its expression is conserved along plant evolution. However, previous work in our lab found that *PsINO* was expressed in the petal tissue of *Papaver somniferum* (a basal eudicot) (unpublished data). Expression of *PsINO* in the petal could corroborate expression of the gene in the non-reproductive tissues of the flower i.e., petal and palea. The *PsINO* expression in petal and *BdINO* in palea indicate that and ancestral function in non-reproductive tissues might have been lost in Arabidopsis.

The expression of *BdDL* is very low in the young leaf and palea but increases during the vegetative to early reproductive stage, coinciding with the initiation and maintenance of carpel primodia. Its expression decreases during the caryopsis development and maturation, suggesting a significant role of *BdDL* during the early stages of caryopsis development as has been reported in rice (**Figure 3.7**).

FIL clade genes, *BdFIL1*, *BdFIL2* and *BdFIL3* are very highly expressed during the inflorescence and emerging spikelet stages, comprising young organ primordia (**Figure 3.7**). Likewise, during similar plant developmental stages in wheat, *FILAMENTOUS* orthologues appear to be highly expressed in spikelets and seed coat layers (**Figure 3.11 and Table 3.6**). However, the expression pattern varies between the three FIL paralogs during grain development in *Brachypodium*. The expression of *BdFIL1* decreases during grain filing and increases at maturation, germination and seedling stage. *BdFIL2* on the other hand, starts being expressed during the grain maturation and continues to be expressed at the germination stage, while the expression levels of *BdFIL3* fluctuate along grain development (**Figure 3.7**).

YAB2 clade genes (*BdYAB1*, *BdYAB2* and *BdYAB6*) are expressed in all tissues and at all stages. Based on our transcriptomic data (**Appendix 3.4**), *BdYAB6* has the highest level of gene expression among all *YABBY*2 subfamily genes, perhaps suggestive of a more significant contribution during grain development. So far, *YAB2* in sorghum has been found to be responsible for shattering effects during parallel domestication, while in eudicots, *YAB2* is involved in the regulation of tomato fruit size (Lin et al., 2012, Cong et al., 2008).

In addition, we are using this transcriptome as a resource to identify genes that co-expressed with *YABBY* genes to elucidate the potential Biological regulations and its networks. The expression pattern analysis will be further explained in **Chapter 3**, and the co-expressions and gene networks will be described in **Chapter 4** in this thesis.



No	Plant Stage
1	Inflorescense (15-21 DAG)
2	Emerging Spikelet (25 DAG)
3	Pre-anthesis Ovary (just before anthesis)
4	Young grain (YG; 1-3 DAA)
5	Mig-length grain (ML; 3-8 DAA)
6	Full Length Grain (FL; 8-15 DAA)
7	Mature Grain (MG; 15-20 DAA)
8	Embryo (MG-E; 15-20 DAA)
9	Germination (1 day after vernalization)
10	Palea (anthesis stage)
11	Lemma (anthesis stage)

Figure 3.7: Expression analysis of *Brachypodium YABBY* genes using RT-PCR. The number and table below correspond to the plant stages. GAPDH expression has been used as a control.

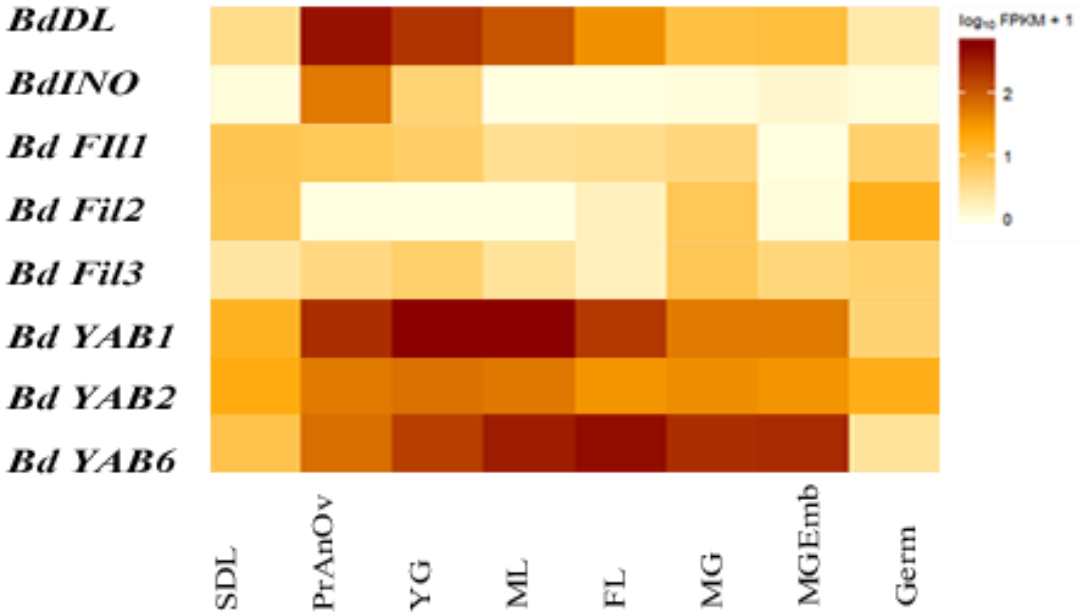


Figure 3.8: *YABBY* genes expression patterns across developmental stages included in the transcriptome (preanthesis ovary - PrAnOv; young, mid, full-length, mature grain – YG, ML, FL, MG; germination – Germ; seedling – SDL)

To allow a better understanding and interpretation of the *in-situ* expression analysis results, emerging spikelet (EmSpk) and pre-anthesis ovary (PrAnOv) paraffin sections were stained with toluidine blue, allowing a more detailed study of the tissue organisation in these stages **(Figure 3.9).** In addition, histone expression was used as a control for each mRNA *in-situ* hybridisation experiment.

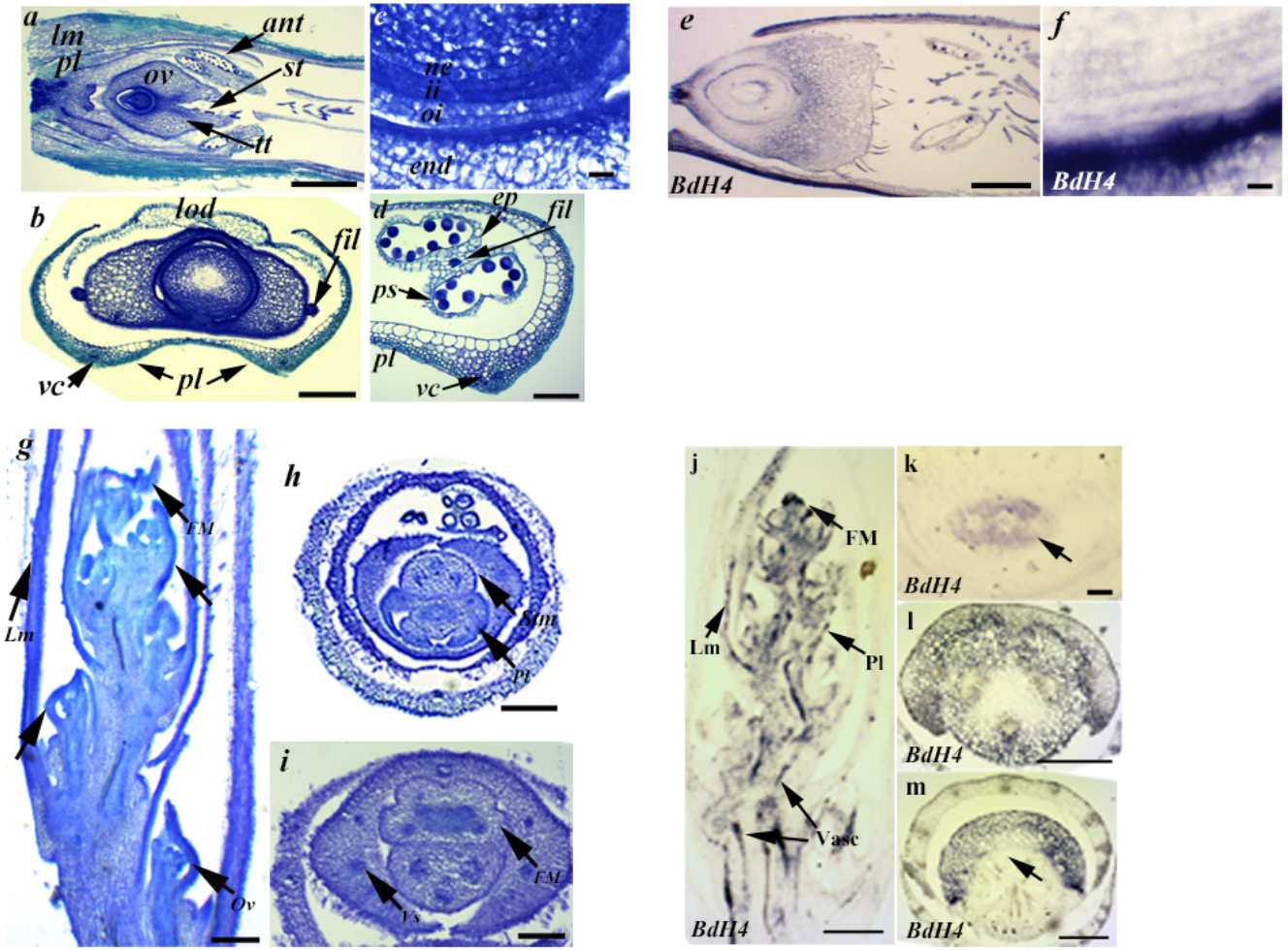


Figure 3.9: **Emerging spikelet and pre-anthesis ovary toluidine blue staining and *in-situ* hybridization histone expression. (a)** Floret at the pre-anthesis stage – longitudinal section **(b)** Floret at the pre-anthesis stage – cross section **(c)** Pre-anthesis ovary close up tissue layers **(d)** Pollen sac and palea with attached vasculature **(e)** Histone (BdH4) expression in pre-anthesis ovary longlitudinal section **(f)** PrAnOv caryopsis cell layers, **(g,h,i)** Emerging spikelet toluidine blue stained sections, **(j,k,lm)** Histone (BdH4) expression in longlitudinal and cross sections of emerging skpikelets. Lm-lemma, pl-palea, ov-ovary, st-stigma,lod-lodicule, fil-filament, vas-vasculature, ii-inner integument, oi-outer intergument, cc-cross cells, tc-tube cells, end-endosperm, fm- floral meristem. Scale 50 µm.

mRNA *in situ* hybridisation analysis was used to determine the spatial expression pattern of *BdDL* and *BdFIL1* in emerging spikelet tissues. *BdDL* and *BdFIL1* showed similar expression patterns, which were consistent with what has been reported for other grass species (**Figure 3.10**). *The BdDL* signal could be clearly detected in the carpel primordia (**Figure 3.10 g**), as expected from the expression pattern in wheat and rice in previous findings from Ishikawa et al. (2009).

The BdFIL signal could be detected in the floral primordia and in the presumptive palea/lemma region as shown in **Figure 3.10 d**. *FIL* paralogs were similarly expressed later during grain maturation and germination, indicating possible roles during these stages.

In addition, the RT-PCR analysis indicated that all *YABBY*s are expressed in the palea tissues, while mRNA *in-situ* hybridisation showed their expression to be restricted to the abaxial region of the palea (**Figure 3.10 h,i**).

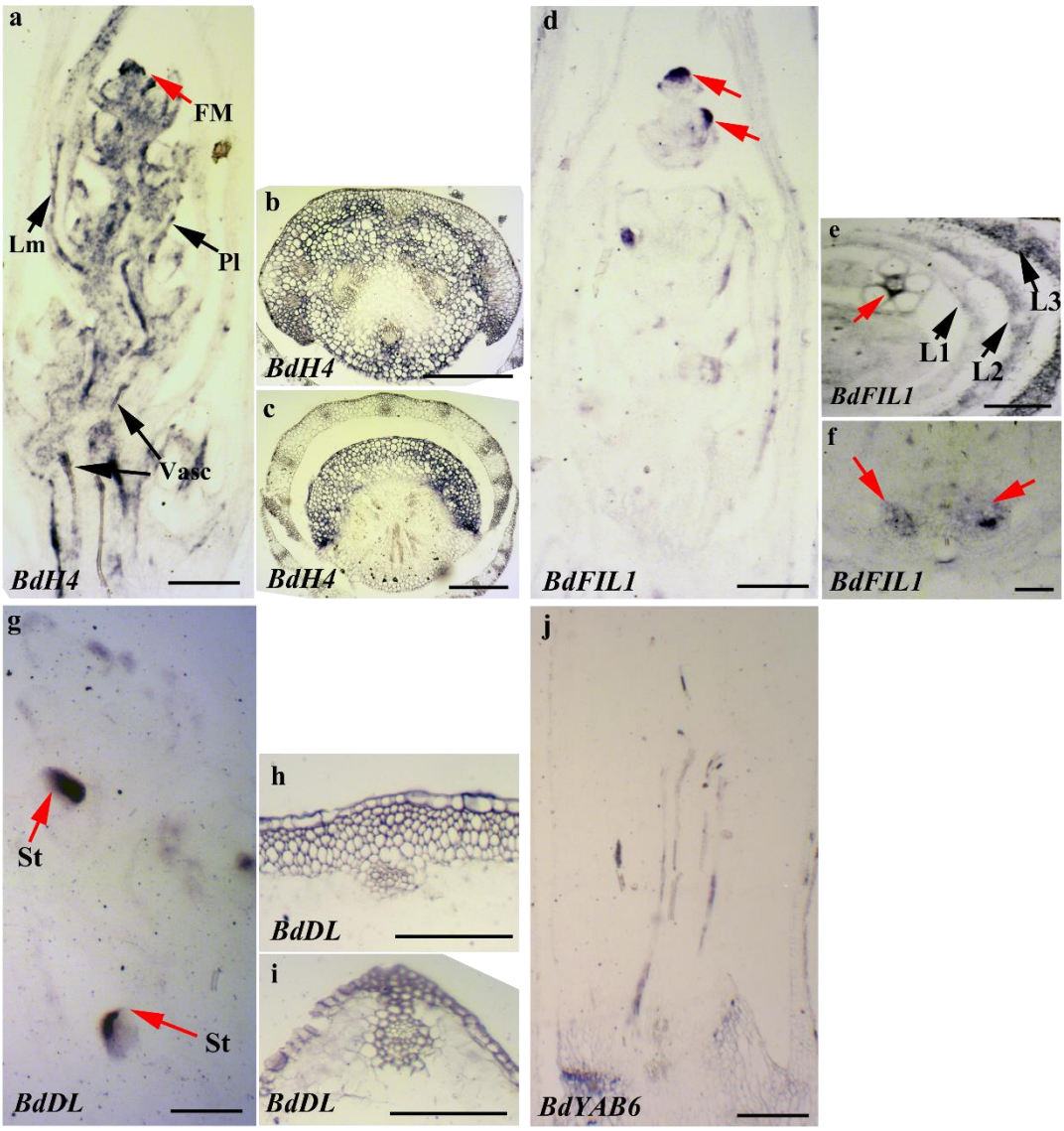


Figure 3.10: mRNA *in-situ* hybridation of *Brachypodium YABBY* genes at the emerging spikelet stage. (a,b,c) Histone (*BdH4*) expression **(d)** *BdFIL1* expression at the spikelet meristem (red arrow) **(e)** *BdFIL1* expression at the anther filament and lemma (black arrow) **(f)** *BdFIL1* expression in the vasculature region (red arrow) **(g)** *BdDL* expression in young carpel primordia (red arrow) **(h,i)** *BdDL* expression in the abaxial region of the palea **(j)** *BdYAB6* expression in vasculature veins. FM-floral meristem, Pl-palea, Lm-lemma, L1,L2,L3- younger to older layer of lemma, st- stigma. Scale, 50 µm.

3.4.4 Novel expression of *YABBY* genes in the grain

Our transcriptomic and RT-PCR expression analyses showed that the *Brachypodium YABBY* genes are expressed during various stages of grain development. We selected representative *YABBY* genes from all four clades to investigate the specific spatial tissue/cell expression (*BdDL*, *BdINO*, *BdFIL1*, and *BdYAB6*) using mRNA *in situ* hybridization (ISH). All these genes were expressed in the pericarp layers, and higher expression could be detected in the chlorenchyma layer, part of the inner pericarp (Figure 3.10). Similarly, transcriptomic experiments in wheat, available to access online, also detected some *YABBY* expression in the pericarp layer, supporting our mRNA ISH results (**Figure 3.12** and **Table 3.6**). In addition, published work in tomato detected *YABBY* expression in the pericarp layer of the fruit suggesting that pericarp expression might be conserved in monocots and eudicots (Pattison et al., 2015).

Further analyses of ovary/pericarp expression showed that *BdDL* antisense signal can be detected in the cells of the stigma, *BdFIL1* expression was detected in stigmatic papillae and the wing region (outer pericarp) in pre-anthesis grain (**Figure 3.11 j,k,l**) , and *BdYAB6* expression can be detected in surrounding vasculature region in grain cross-section(**Figure 3.11 m,n,o**). In addition, even though *BdYAB1* and *BdYAB6* belong to the same clade, the expression patterns show some differences: *BdYAB6* is only expressed in the abaxial domain of the palea, and surrounding the vasculature region, while *BdYAB*1 is expressed in the vasculature region of the palea (**Figure 3.11 p,q**).

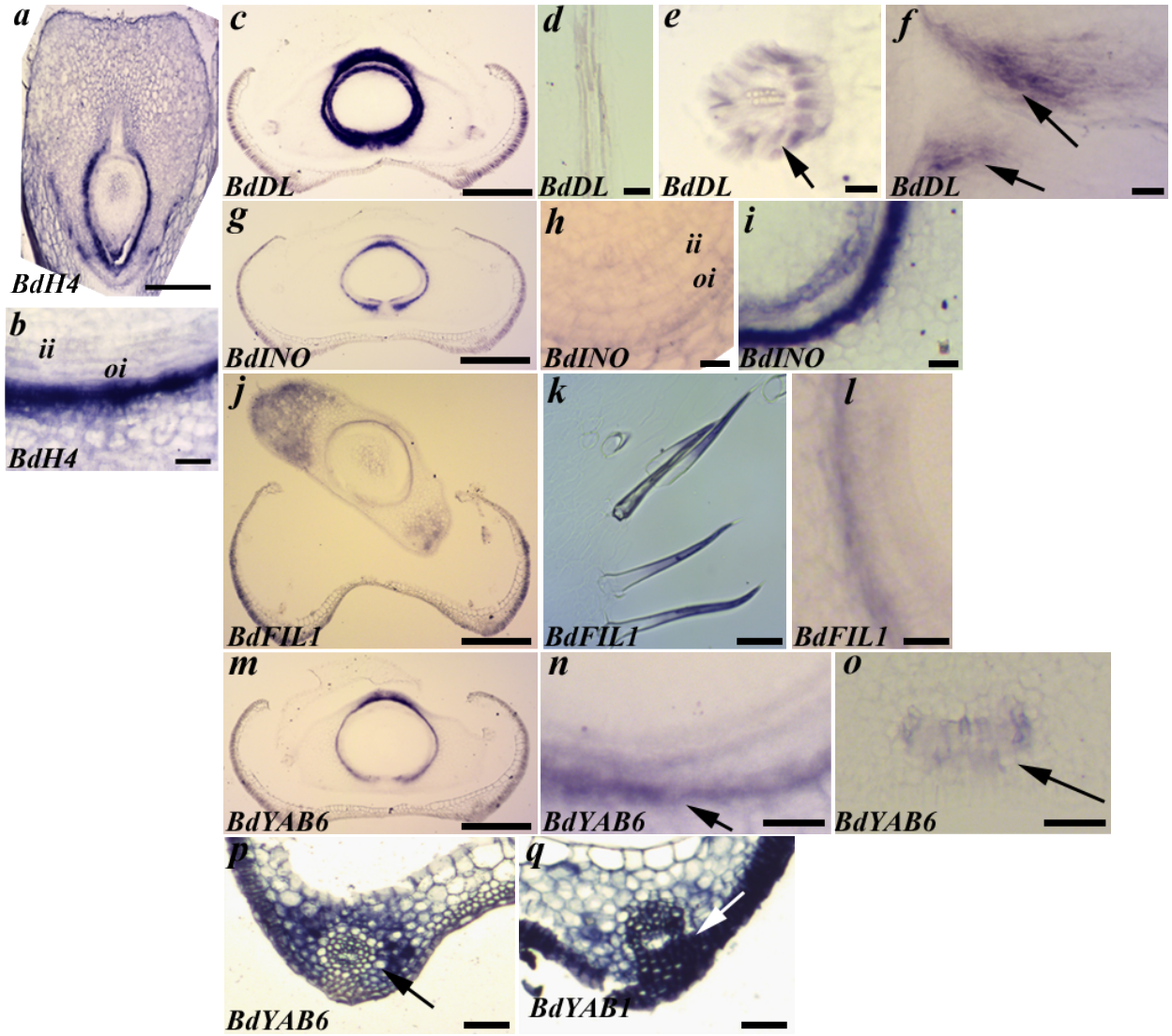


Figure 3.11: mRNA *in-situ* hybridization analysis of *YABBY* expression at the pre-anthesis stage. (a,b). *BdH4* expression as control, **(c)** *BdDL* expression in the pericarp layers **(d,e)** *BdDL* expression in filament walls (arrow) **(f)** *BdDL* expression in the transmitting tract (arrow) **(g,h,i)** *BdINO* expression in the pericarp **(j)** *BdFIL* expression in the endosperm and the pericarp tissues **(k)** *BdFIL* expression in the stigma hairs **(l)** *BdFIL* expression in the pericarp layer **(m,n)** *BdYAB6* expression in the pericarp layers **(o)** *BdYAB6* expression in the vasculature region (arrow) **(p)** *BdYAB6* expression in the palea and not in the vasculature (arrow), **(q)** *BdYAB1* expression in the palea and vasculature region (white arrow). ii-inner integument, oi-outer integument, scale 50 µm.

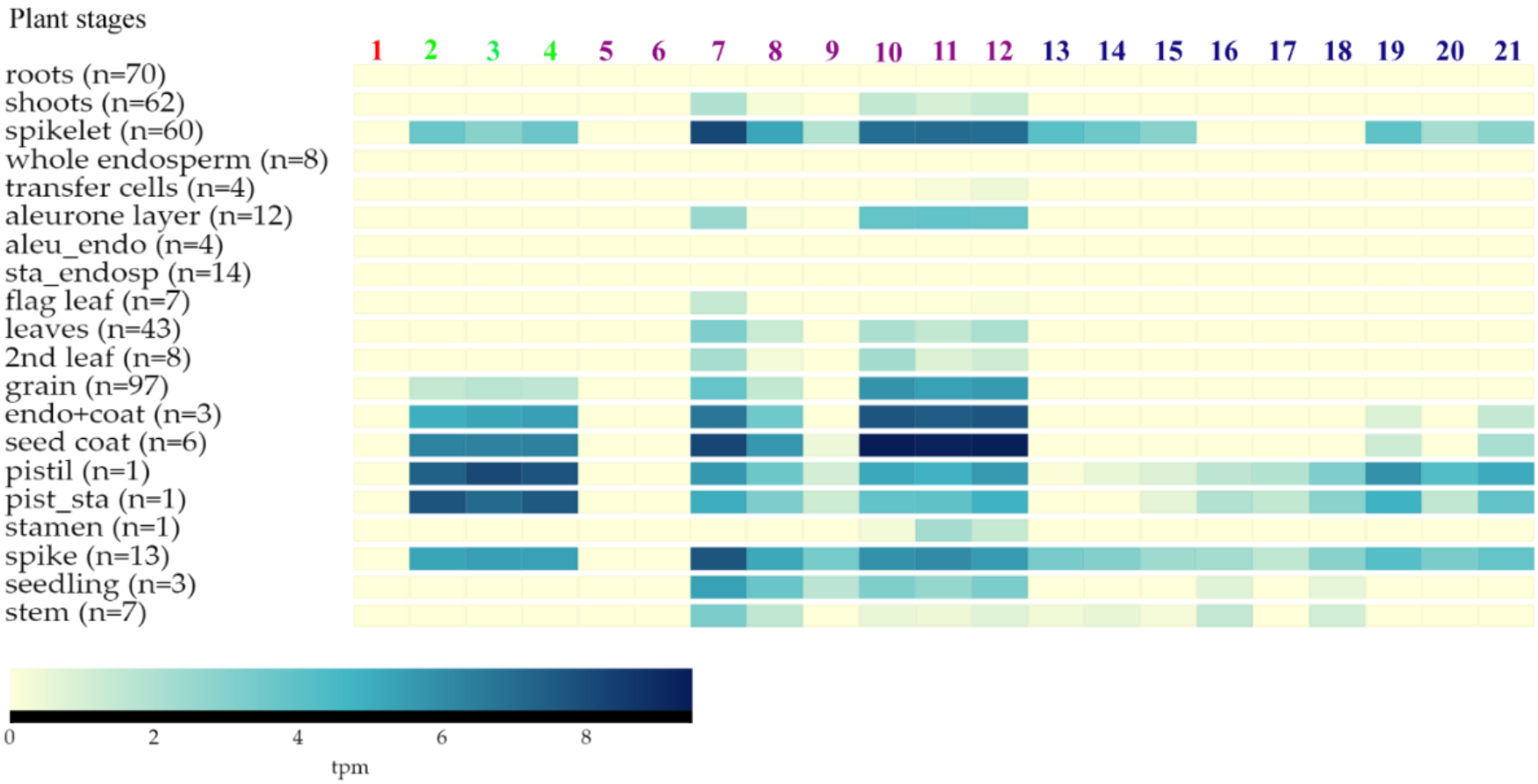


Figure 3.12: *YABBY* orthologues and expression profiling in wheat. Wheat Expression across plant stages extracted from the Wheat Expression Browser Website (http://www.wheat-expression.com). Corresponding numbers and gene IDs presented in the table next page.

Number	Gene ID	Synonym
INO		
1	TRIAE_CS42_2DS_TGACv1_178838_AA0601610	Ta_2D_INO
DL/CRC		
2	TRIAE_CS42_4AS_TGACv1_306676_AA1011910	Ta_4A_DL
3	TRIAE_CS42_4BL_TGACv1_321049_AA1054410	Ta_4B_DL
4	TRIAE_CS42_4DL_TGACv1_343031_AA1127860	Ta_4D_DL
YABBY2		
5	TRIAE_CS42_2BL_TGACv1_130161_AA0405330	Ta_2B_YAB1
6	TRIAE_CS42_2DL_TGACv1_158896_AA0528070	Ta_2D_YAB1
7	TRIAE_CS42_5AL_TGACv1_374236_AA1194610	Ta_5A_YAB2
8	TRIAE_CS42_5BL_TGACv1_404477_AA1301190	Ta_5B_YAB2
9	TRIAE_CS42_5DL_TGACv1_433350_AA1410470	Ta_5D_YAB2
10	TRIAE_CS42_5AS_TGACv1_393321_AA1271170	Ta_5A_YAB6
11	TRIAE_CS42_5BS_TGACv1_423385_AA1375620	Ta_5B_YAB6
12	TRIAE_CS42_5DS_TGACv1_457157_AA1483350	Ta_5D_YAB6
FILAMENTOUS FLOWER		
13	TRIAE_CS42_2AL_TGACv1_094021_AA0291200	Ta_2A_FIL1
14	TRIAE_CS42_2BL_TGACv1_130187_AA0405790	Ta_2B_FIL1
15	TRIAE_CS42_2DL_TGACv1_160252_AA0548640	Ta_2D_FIL1
16	TRIAE_CS42_6AL_TGACv1_471626_AA1511990	Ta_6A_FIL2
17	TRIAE_CS42_6BL_TGACv1_499794_AA1591850	Ta_6B_FIL2
18	TRIAE_CS42_6DL_TGACv1_528528_AA1714780	Ta_6D_FIL2
19	TRIAE_CS42_1AL_TGACv1_000156_AA0004990	Ta_1A_FIL3
20	TRIAE_CS42_1BL_TGACv1_033155_AA0137410	Ta_1B_FIL3
21	TRIAE_CS42_1DL_TGACv1_061067_AA0183920	Ta_1D_FIL3

Table 3. 1 : Corresponding numbers and gene IDs presented in Figure 3.10. The synonym IDs were used in the phylogenetic analyses in previous sections of this chapter.

3.5 Discussion

3.5.1 Evolution of *YABBY* genes in plants

Our findings show that sequence evolution of *YABBY* genes in the plant kingdom started before the emergence of the seed plant lineage (the gymnosperms). This contradicts recent findings by Finet et al. (2016) who claimed that no *YABBY*-like genes were present in algal genomes. We found *YABBY*-like genes in green algae in two different strains (from different habitats) of *Micromonas pusilla*. Both of the genes have a homeobox domain or *YABBY*-like domain as mentioned in the previous section (**Section 3.42**), but only one of them contains the zinc finger domain characteristic of *YABBY* genes. Perhaps none of these strains were considered in that analysis (Finet et al., 2016). Our current phylogenetic tree focusing on grasses indicates both gene duplication events and possible gene losses, giving rise to several paralogous genes in some species and the lack of orthologues in others. The presence of paralog genes in grasses lineage suggests that duplications occurred after the split of eudicots and monocots.

Less is known about the *FIL* and *YAB2* subfamilies in grasses, hence the functional relationship and the cause of allelic duplication is needs further investigations.

3.5.2 *INO* and *YABBY*1 might have been lost during the evolution of grasses

INO was maintained as a single copy in plant lineages except in the Triticeae. The INO single mutant in *Arabidopsis* results in sterile seeds due to failure of outer integument initiation (Villanueva et al., 1999). Our findings revealed that INO orthologues can be detected in wheat, but not in barley or rye (available transcriptome data in the lab). INO and YABBY1 do not seem to be expressed at all in wheat, at least in the developmental stages tested.

The recent publication of the rye genome (Bauer et al., 2017) would allow to further investigate if *INO* is present in rye. No functional report about *INO* has been published in maize, rice and wheat to date, hence it is hard to interpret the *INO* loss based only on expression analyses.

Duplicated genes in plant lineages can have several possible fates: gene loss, subfunctionalization, neofunctionalization and redundancy (Airoldi and Davies, 2012). The lack of expression of *INO* and *YAB1* in wheat and no detection in other members of

the Triticeae tribe might be due to post-duplication redundancy. Based on RT-PCR results in *Brachypodium*, which is a closely related species that evolved before the *Triticeae*, the expression profile for all *YABBY*2 subfamily genes is highly similar. In addition, the mRNA *in situ* hybridisation allowed the detection of *BdYAB1* and *BdYAB6* expression in the same tissues, leading to the hypothesis that these genes might have a redundant function in *Brachypodium*. However, the expression of the wheat *YAB1* orthologues was not detected in any of the plant stages/tissues analysed, in contrast with the expression of the other homeolog genes in the same clade. The *YAB1* low expression and then loss in the *Triticeae* might indicate that the duplication of this gene is not under selective pressure, hence likely to be lost during plant evolution. In addition, *YAB1* might not be highly involved in gene networks, hence it is not required to be maintained for gene balance (Edger and Pires, 2009).

3.5.3 YABBY expression is conserved in carpel and meristem formation

Expression of *BdDL* in the carpel primordia during carpel initiation appears to be similar to the expression of other *DL* orthologues in rice, wheat, maize and sorghum, suggesting functional conservation of *DL* in grasses (Ishikawa et al., 2009).

The expression pattern of *FIL* varies between species. Rice and wheat show a non-polar expression compared to *Arabidopsis*, where *FIL* is localized at the abaxial region of organ primordia (Tanaka et al., 2012, Bartholmes et al., 2012, Chen et al., 1999, Sawa et al., 1999). The expression pattern of *FIL* varies within monocots as well. Maize has been reported to show adaxialized expression of *FIL*, while no other case of polar *FIL* expression has been reported in monocots to date (Juarez et al., 2004). The non-polar expression of *BdFIL* in the carpel primordia appear to be similar with that of *OsFIL* in rice. A single *fil* mutant in rice failed to maintain carpel formation, and showed a homeotic transformation of carpels to stamens (Tanaka et al., 2012). Based on their expression patterns, *BdDL* and *BdFIL* would be expected to have a similar mutant phenotype to the rice *dl* and *fil* mutants. Further knockdown of the two genes in *Brachypodium* would enable the validation of this hypothesis.

The expression of all *BdYABBY*s in the palea and lemma might indicate their involvement in the organ's development. Palea/lemma has been suggested to be analogous to sepals in dicots (Schmidt and Ambrose, 1998), as they share a similar histology; fibrous sclerenchyma cells at the abaxial region (Prasad et al., 2005). The palea and lemma in

grasses are photosynthetic organs that supply the developing caryopsis with carbon and nitrogen, while also protect the grain from pathogens and herbivors (Abebe et al., 2004). The expression of *BdYABBYs* in the palea is consistent with a previous report on the expression of Os*YABBY1* in rice (Toriba et al., 2007).

3.5.4 YABBY expression in mature carpel is new

The YABBY gene expression in the developing carpel, especially post-anthesis, has not been studied previously in the grass caryopsis and so this study presents new insights. Previous expression analyses focused on the developing carpel primordia prior to anthesis (Ishikawa et al., 2009, Li et al., 2011, Ohmori et al., 2011, Yamaguchi et al., 2004). Based on our mRNA *in-situ* analyses, the expression of all representative *YABBY* genes can be clearly detected in the pericarp tissues. These genes might act redundantly during the development of the seed coat in *Brachypodium*. The relationship between seed coat and grain size should also be considered for grain improvement (Radchuk and Borisjuk, 2014). A comprehensive review paper in 2014 (Radchuk and Borisjuk, 2014) discussing the significant contribution of seed coat to endosperm filing process and final seed size. KLUH/CYTOCHROME 450 78A5, expressed highly in pericarp layers in Arabidopsis and the upregulation of this gene increases seed size (Adamski, et.al., 2009). In addition, two candidate genes (*DWARFT1* and *GIF1*) in rice were narrowed down from quantitative trait loci (QTL) was related to grain size were expressed highly in grain pericarp (Izawa et.al., 2010, Huang et.al, 2013). Hence, the expression of YABBYs in pericarp layer might shed a potential TFs responsible for seed development and/or enlargement.

3.5.5 YABBY6 is a novel gene that might be involved in grain development

Our preliminary results show that *YABBY*6 is potentially involved in the regulation of caryopsis development. Amongst all *YABBY*s, *YAB*6 has an interesting expression pattern with its transcripts being increased during grain development. The mRNA ISH signal detected in *Brachypodium* is consistent with the RNA-seq data available in wheat (WheatExpression browser). In *Brachypodium*, a strong signal of *BdYAB6* can be detected in the pericarp layer of mature pre-fertilization ovaries and young grains. At the young grain stage (12 DAA) in wheat, three *YAB*6 paralogs are expressed in the seed coat. The pericarp in grasses is important for starch synthesis and respiration during caryopsis development. The correlation of *YABBY*6 expression with genes involved in photosynthesis is interesting and something that will be further explored. In addition, GO terms of *YAB*6 from our transcriptomic analyses indicate a function in sequence

specific DNA-binding and protein heterodimerization. This is consistent with our yeast-2-hybrid results where YAB6 is the only protein interacting with other YABBY proteins in the same clade (BdYAB2 and BdYAB1) -refer to **Chapter 4**. Further analysis is required to reveal the function of *YABBY*6 during caryopsis development.

3.6 Conclusion

Duplications of *YABBY* genes in the *YABBY2* and *FILAMENTOUS* clades in grasses raised an interesting perspective. The duplicated genes might infer new functions or subfunctionalization. In addition, the novel expression of *YABBYs* in the seed coat might suggest new roles of *YABBY* genes. We also suggest *BdYAB6* as an interesting candidate for the regulation of grain development in grasses. Further functional analyses would help to address the uncertainty of several findings in this project.

Chapter 4 – YABBY functional analysis and gene networks

4.1 Introduction

The knowledge of structure and function of transcription factors (TFs) is important as they are key regulators of gene expression. Gene expression is primarily controlled at the level of initiation of transcription and is a complex process involving a network of interactions between different regulatory proteins and cis-regulatory elements present in the promoters of their target genes (Harbison et al., 2004).

In general, TFs consist of two distinct domains; a DNA binding domain that provides the ability to bind to specific DNA sequences (enhancers or promoter sequences) and a protein interaction domain that allows the formation of complexes with other proteins or enzymes. DNA sequences termed as *cis*-regulatory elements are usually found in the enhancer or promoter region of TF target genes. The ability to bind to such DNA sequences is a distinctive characteristic of TFs. Yeast-two-hybrid is a powerful tool that is used to investigate which proteins physically interact in order to gain an insight into the molecular networks involving given TFs. However, a closer look is needed to identify the sequence specific binding to the motifs, and how they interact to carry out functional roles in plants.

A considerable amount of research has been undertaken and published on gene function analyses. Loss-of-function mutants and gene silencing techniques are being used to get a better understanding of gene function, although functional redundancy between genes may often not allow the inference of function from single gene mutations. Other strategies are also being used, such as overexpression and/or ectopic expression of genes of interest. For example, a strong constitutive promoter, such as the 35S Cauliflower Mosaic Virus (CaMV35S) promoter, is often used to drive the expression of the gene of interest in order to study the effect that an excess number of transcripts and/or expression in other tissues would have in plant development and metabolism. In any case, the resulting plant phenotype needs careful evaluation and interpretation, especially when genes are ectopically expressed. Nevertheless, this technique is often used as a first step to predict the functions of investigated TFs.

In addition to the study of mutant phenotypes, knowledge on the type of DNA motifs that TFs bind to would be valuable in understanding transcriptional regulation. TFs bind to DNA in a sequence-specific manner in order to regulate gene expression. Discovery of motif binding sites usually starts with gene sets that show a similar expression pattern. It is hypothesised that genes with similar expression patterns would have similar biological functions. In theory, similar biological function genes would be regulated by the same TFs, and therefore their genomic region upstream their start codon would be expected to contain similar regulatory motif sequences (Zambelli et al., 2013).

The work presented in this chapter was primarily focussed on the functional analysis of the *Brachypodium YABBY* genes and the discovery of *YABBY* regulatory sequence motifs. This chapter extends the link between the transcriptomic approach and the current knowledge on the *YABBY* gene family presented in the previous chapters to:

i. Initiate a reverse genetics approach to infer functional roles of *YABBY* genes in *Brachypodium* using RNAi and *Arabidopsis* mutant complementation studies.
ii. Investigate the functional conservation of YABBY proteins, with a focus on BdDL.
iii. Investigate the protein interactions between the different YABBYs in *Brachypodium* and compare them to what is already known in *Arabidopsis* and other species.
iv. Identify potential DNA sequence motifs that are over-represented in genes with a known function in polarity determination.

4.2 Materials and methods

4.2.1 Plant Material

Brachypodium distachyon grains (Bd-21 accession) were germinated on moist filter paper after stratification at 4°C for 48h. Five-day old seedlings were transferred to 9cm square pots with a 2:2:1 multipurpose compost: vermiculite: sand mix and grown under controlled environment conditions with a 18h photoperiod at 20°-22°C and light intensity of 180-200 umol/m/s.

4.2.2 Gateway cloning (DL)

Two constructs were generated for plant transformation in order to study the function of *DL* in *Brachypodium*; an RNAi and an Overexpression construct. Gateway® technology was used to generate both constructs. This system is using the recombination properties of lambda bacteriophages to rapidity and efficiently clone and subclone DNA sequences into multiple vectors. It comprises of two main types of vectors:

i. Entry vectors, such as pDONR221 that are used to generate entry clones.

ii. Destination vectors, such as pIPKb027 and pIPKb002,that are used to generate the final expression clones.

The workflow used for preparing the constructs for plant transformation is shown in **Figure 4.1**.

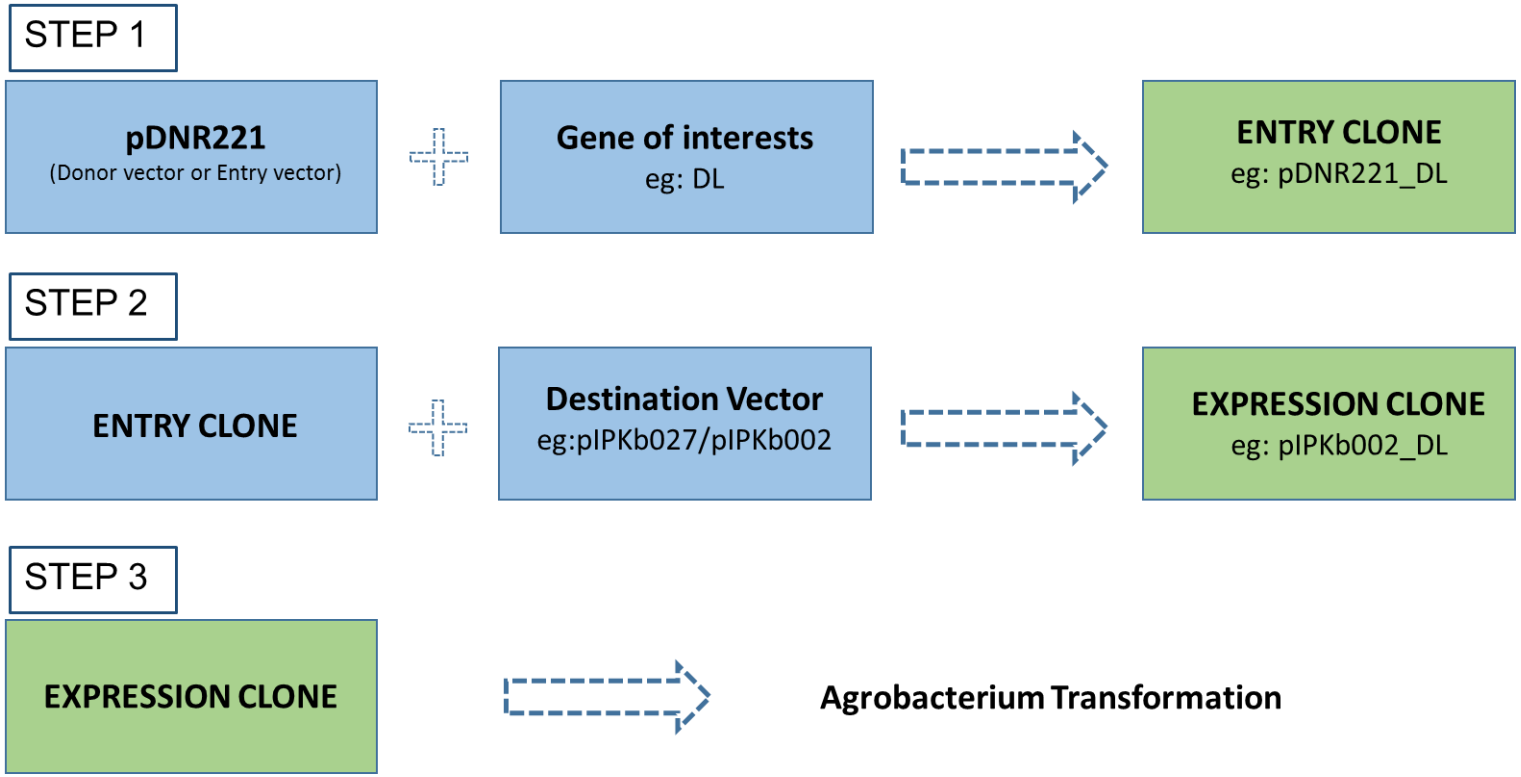


Figure 4.1. Gateway cloning workflow. Recombination of an entry vector (pDNR221) with the gene of interest results in the generation of an entry clone. The entry clones are subsequently recombined with a destination vector to generate the final expression clones. Expression clones were transformed into *Agrobacterium tumefaciens* (strain AGL1) through electroporation in order to be used for *Agrobacterium*-mediated plant transformation. Vector details are included in **Appendix 4.1**.

4.2.3 Entry Clones

4.2.3.1 Design of attB site primers

The RNAi primers were designed to amplify a specific region (*DL*-275 bp, *INO*-200 bp), avoiding the conserved domains. The Overexpression (OE) primers were designed to amplify the entire coding sequence (CDS) region (DL-618 bp,) and BdINO (500 bp). Forward primers were designed to include part of the attB1 recombination site, while reverse ones incorporated a part of the attB2 site. Details of the primers used and the amplified regions can be found in **Appendix 4.1**.

4.2.3.2 Amplification and purification of PCR products

PCR was used to amplify the DNA fragments using a high fidelity Taq polymerase (Q5, NEB). A good quality cDNA was used as template in a 20 µl PCR reaction, together with gene-specific primers containing an attB attachment. Then the PCR product was subsequently used as a template with attB adapter primers in a 50 µl PCR reaction. All primers sequences can be found in Appendix 4.2. The reactions and PCR program details are shown in **Table 4.1**.

Table 4.1: Two-step PCR workflow summary for thermocycling PCR 1 and PCR 2. PCR reaction and thermocycling condition for both PCRs.

2 STEPS PCR reactions (Q5 High-Fidelity DNA Polymerase)		
1st Step	2nd step	
20 µl	50 µl	SOLUTION
7.8	21.5	H_2O
4	10	5x Q5 Reaction Buffer
4	10	GC Enhancer
0.8	2	Forward primer with *attB* site attachment
0.8	2	Reverse primer with *attB* site attachment
2	3	cDNA template / PCR products
0.4	1	dNTPs (10 µM)
0.2	0.5	Q5 High-Fidelity DNA Polymerase

PCR Thermocycling Conditions for both reactions		
STEP	TEMPERATURE (∘C)	TIME (secs)
Initial denaturation	98	00:30
20 µl PCR (20 cycles) 50 µl PCR (30 cycles)	98	00:10
	55	00:20
	72	00:30
Final Extension	72	02:00
Hold	10	∞

Ten µl of the PCR products were run on a 1% agarose (Molecular Grade, Bioline) gel, stained with 0.5 µg/mL EtBr. Single bands with the correct expected size were directly purified from the remaining PCR reaction while in the cases where multiple bands occurred on the gel, they were purified after the correct band size was excised from the gel. All purifications were performed using the PCR/Gel purification kit (Bioline), according to manufacturer's instruction.

4.2.3.3 BP recombination reaction

BP recombination reaction was performed to transfer an *att*B-PCR product (gene of interest) into an *att*P–containing entry vector to create an entry clone. In this case, pDONOR221 vector (Invitrogen) was used. The BP recombination was performed following manufacturer's recommendations as shown in **Table 4.2**.

Table 4.2: BP recombination reaction

BP reaction mixture	Volume
*Vector ~500ng/reaction	1.0 µl
*Inserts DNA fragment ~500ng/reaction	2.0 µl
BP clonease	1.0 µl
TE Buffer	1.0 µl
	=5 µl total

*the concentrations may vary and the mixture were left overnight at room temperature (RT).

4.2.3.4 Escherichia coli (E. coli) transformation

The recombination products (entry clones) were transformed into DH5α cells (*E.coli*) following a standard heat-shock transformation protocol.

A tube containing 50 µl of DH5α competent cells was thawed on ice. 2.5 µl of the BP recombination products were pipetted into the cells and mixed gently. The tube was then left on ice for 30 min and subjected to a heat-shock for 40 se cat 42 °C before immediately transferred back on ice for 2 min. 700 µl of SOC medium was added into the mixture and left in a 37 °C shaker for an hour before the tube was centrifuged at 3,000 x g for 1 min. The supernatant was pipetted out and 100 µl of LB medium were used to resuspend the bacteria pellet. The transformation mixture was then plated on LB agar plates containing kanamycin (50 mg/L) and left for regeneration in a 37 °C incubator overnight.

The next day, PCR screening was performed for individual colonies on the incubated plates, using M13_F and M13_R general primers (Appendix 4.2). Colonies that showed the correct size of amplicon were selected for overnight cultures. Five ml of LB broth were added to a 50 ml universal tube, with 5 µl of Kanamycin to a final concentration of 50 mg/L and a very small amount of the selected colony from the plate. The cultures were then left in a 37 °C shaker at 250 ppm overnight.

4.2.3.5 Plasmid miniprep and sequencing

Plasmid isolation from the overnight cultures was performed using the ISOLATE II Plasmid Mini Kit (Bioline) following the manufacturer's instructions.

The isolated plasmids were quantified using NanoDrop 2000/2000c (Thermo Scientific) and were sequenced by GATC-Biotech. Sequenced plasmids were checked and only those showing 100% similarity to the inserted *Brachypodium* DNA sequence were selected to be taken forward.

4.2.4 Expression Clones

4.2.4.1 LR recombination reaction

LR Recombination was used to generate the expression clones. The entry clones were recombined with an *att*R-containing expression vector to create an expression clone. For this project, two different expression vectors were used; pIPKb027 for RNAi and pIPKb002 for overexpression (**Appendix 4.1**). Five µl of LR recombination reaction mixture was prepared following the table below.

Table 4.3: LR recombination reaction

LR reaction mixture	Volume
Entry Clone (25-75 ng/µl)	*1.0 µl
Destination Vector (75 ng/ µl)	*1.5 µl
LR Clonase II	1.0 µl
TE Buffer	1.5 µl
	=5 µl total

*Volumes may vary depending on the entry clone and destination vector concentration. The mixture was left overnight at RT.

The assembled mixture was left overnight at RT. On the next day, 2 µl of Proteinase K solution were added and the mixture was incubated at 37°C for 10 min to stop the recombination reaction. 2.5 µl were then transformed into *E.coli* (DH5α) competent cells as described in **Section 4.2.3.4**. The transformed cells were plated on LB agar containing spectinomycin (50 mg/L) and left for regeneration in a 37 °C incubator overnight.

PCR screening was performed on the next day and the colonies containing the correct bp size product were selected for overnight cultures. 5 ml of LB broth with 5 µl of spectinomycin (final concentration of 50 mg/L) were added to a 50 ml universal tube and mixed with a selected colony. Cultures were then left in a 37 °C shaker at 250 rpm overnight. Plasmids were isolated from selected overnight cultures and sequenced as described in **Section 4.2.3.5**.

4.2.5 *Agrobacterium* Transformation by Electroporation

Three µl of a 10-fold diluted expression clone were added to thawed *Agrobacterium tumefacians* AGL1 electrocompetent cells and mixed gently. The mixture was transferred into a pre-chilled, dried electroporation cuvette and fitted to a Gene Pulser (BioRad) set at 25 µF capacitance and 2.5 kV voltage until a beeping sound occurred. One ml of LB

broth was pipetted immediately into the cuvette and the mixture was transferred into a microcentrifuge tube. The tube was left in a 28 °C shaker for 2 hours and then centrifuged at 3,000 x g for 30 sec before the supernatant was pipetted out and 250 µl of fresh LB broth were added. The mixture was finally plated on LB agar media containing rifampicin (50 mg/L), carbenicillin (100 mg/L) and spectinomycin (50 mg/L). The plates were left in a 30 °C incubator for 2-3 days for colony regeneration. Colony PCR was performed and colonies with the correct size of insert were selected for plant transformation.

Agrobacterium bacterial stabs were send to Dr M. Moscou's lab at The Sainsbury Laboratory, Norwich UK, where *Agrobacterium*-mediated transformation of *Brachypodium* was performed.

4.2.6 *Arabidopsis* Plant transformation and genotyping (CRC)

Arabidopsis transformation experiments were performed in order to investigate the protein function conservation across plant species.

4.2.6.1 Plant material

Arabidopsis mutant lines were grown separately in a growth cabinet (23 – 25 ˚C, with 24 hours photoperiod) until flowering. The first bolts were clipped to encourage the proliferation of secondary bolts. Plants with clusters of immature flowers were selected and already developed siliques were clipped off before dipping.

4.2.6.2 Floral dipping / transformation

Arabidopsis floral dipping transformation was performed according to Clough and Bemt (1998). An initial 5 mL LB *Agrobacterium* culture with antibiotics; rifampicin (50 mg/L), carbenicillin (100 mg/L) and spectinomycin (50 mg/L) was grown overnight in a 28˚C shaker incubator. 400 µL of the overnight culture were then pipetted into 400 mL fresh LB media with the same antibiotic selection and were incubated overnight using the same conditions. On the third day, the bacteria cells were pelleted using centrifugation at 4600 rpm for 20 min, and subsequently re-suspended in 1 L of infiltration medium. The composition of the infiltration medium is shown in **Table 4.4**.

Table 4.4: Agrobacterium transformation infiltration medium

1 Litre of infiltration medium:	Weight/Volume
Half strength of MS salts	2.15 g
Full strength of Gamborg's B5 vitamins	3.16 g
MES	0.5 g
Sucrose	50.0g
1 mg/mL benzylaminopourine (6-BAP)	10 µL

After the pellets re-suspension, 400 µL Silwet L-77 were added and mixed well prior to dipping. The above ground parts of *Arabidopsis* plants were dipped with gentle agitation for 40 sec and then laid flat on trays lined with paper and covered with an incubator lid to keep a high humidity environment. The incubator lids were removed after two days, the plants pots were turned standing and were watered regularly thereafter. Watering stopped when the formed siliques reached maturity. Dry seeds were harvested and subjected to hygromycin screening.

4.2.6.3 Seed screening

The expression vectors used for plant transformation carried a hygromycin resistance selectable marker. Therefore, harvested seeds were screened for hygromycin resistance in order to identify true transformants. Seeds were initially sterilised using 10 % sodium hypochloride for 10 min, and then washed 3-4 times with sterile water and a final wash of 70 % ethanol, before being left to dry on sterile filter papers in a flow-hood cabinet. Sterile seeds were then plated on MS agar media containing 50 mg/L hygromycin.

The plated seeds were kept in the dark at 4 ˚C for two days for stratification, then the plates were transferred in the media room, 20 ˚C with 24 hours light. Seedlings with long hypocotyl and dark green cotyledon leaves were selected and transferred to soil.

Genomic DNA was extracted from young leaves and was subjected to PCR to confirm that plants carry the relevant T-DNA. A left T-DNA border (LB) primer was used in combination with two gene-specific primers in a multiplex PCR reaction to investigate if each plant was homozygous for the T-DNA insertion, heterozygous, or did not contain the insersion at all (WT). All primers are shown in **Appendix 4.2**.

4.2.7 Yeast-Two-Hybrid

4.2.7.1 PCR cloning

In order to clone the coding region of each gene of interest, primers were designed that amplify the sequence between the start and stop codon, with the attachment of specific restriction site sequences to facilitate their ligation into the relevant vector.

4.2.7.2 Selection of restriction enzymes

Prior to primer design, the compatibility of the restriction enzymes was checked to make sure that the restriction sites do not occur in the respective DNA fragment to be amplified and occur once on the plasmid's Multi-Cloning Site (MCS) (pGADT7 or pGBKT7). For this purpose, the NEBCUT online tool (http://nc2.neb.com/NEBcutter2/) was used and the compatible restriction site was then added to the 5' end of each primer. The primers used for each gene are shown in appendix 4.2.

4.2.7.3 High-fidelity PCR

PCR was performed to amplify the entire gene CDS using a high fidelity Taq polymerase (Q5, NEB). A good quality cDNA was used as template for a 50 µl PCR reaction. The reactions and PCR program details are shown in **Table 4.5**.

Table 4.5: PCR reaction and thermocycling conditions for Y2H cloning.

PCR reaction (50 µl)	Volume
H_2O	21.5 µl
5x Q5 Reaction Buffer	10.0 µl
GC Enhancer	10.0 µl
Forward primer with enzyme attachment (10 µM)	2.0 µl
Reverse primer with enzyme attachment (10 µM)	2.0 µl
cDNA template	3.0 µl
dNTPs (10 µM)	1.0 µl
Q5 High-Fidelity DNA Polymerase	0.5 µl

PCR Themocycling Conditions

Step		TEMPERATURE (∘C)	TIME (min)
Initial denaturation		98	00:30
Denaturation		98	00:10
Annealing	30 cycles	55	00:20
Extension		72	00:30
Final Extension		72	02:00
Hold		10	∞

Ten µl of the PCR products were run on a 1% agarose (Molecular Grade, Bioline) gel and stained with 0.05 µg/mL EtBr. Single bands with the expected size were directly purified using the PCR/Gel Purification kit (Bioline), while in the case where multiple bands appeared on the gel, they were purified after the correct band size was excised.

4.2.7.4 Double restriction enzyme digestions for DNA inserts and vectors

After the PCR purification, DNA quantification was performed using NanoDrop 2000/2000c (Thermo Scientific). Approximately 500 ng of DNA fragments and plasmids (pGADT7 or pGBKT7) were digested with selected restriction enzymes. Digestion reactions were setup following the recommended protocol from NEB (**Table 4.6**). The assembled reactions were incubated at 37 ˚C for 1 hour with NdeI, then the second enzyme was added (EcoRI HF) and incubation carried on for another 1.5 hours.

Table 4.6: Double restriction enzyme digestion reactions

For 25 µl reactions:

GENE/VECTOR	CONCENTRATION (ng/µl)	BUFFER (CutSmart) (µl)	RESTRICTION ENZYME (µl)		DNA (µl)	H_2O (µl)
			NdeI	EcoRI		
BdDL	40.5	2.5	0.5	0.5	12	9.5
INO	110	2.5	0.5	0.5	4	17.5
YAB6	106.3	2.5	0.5	0.5	5	16.5
FIL1	130	2.5	0.5	0.5	5	16.5
pGADT7	100	2.5	0.5	0.5	5	16.5
pGBDT7	100	2.5	0.5	0.5	5	16.5

4.2.7.5 T4 Ligation

The DNA fragment and vector ligations were performed following the recommendation of T4 DNA Ligase kit (New England BioLabs) and using a 1:3 vector to insert ratio. All genes (*Bradi1g69900/BdDL*, *Bradi1g24051/BdINO*, *Bradi4g01300/BdYAB6*, and *Bradi5g16910/BdFIL1*) were ligated with both pGADT7 and pGBKT7 vectors so they can be used as both bait and prey in the yeast-two-hybrid experiments. Ten µl reactions were prepared following the formula below and the assembled ligation reaction was left at RT overnight.

$$ng\ of\ DNA\ insert = \frac{ng\ of\ vector\ \times kb\ of\ DNA\ insert}{kb\ of\ vector} \times molar\ ratio\ \frac{3}{1}$$

4.2.7.6 E.coli transformation, plasmid miniprep and sequencing

2.5 µl of the ligation products were transformed into DH5α cells (*E.coli*) following the protocol described in **Section 4.2.3.4**. pGADT7 transformations were plated on LB containing ampicillin (100 mg/L) while pGBKT7 ones on LB plates containing kanamycin (50 mg/L). Plates were incubated overnight at 37˚C.

PCR screening was performed for selected colonies from the plates and those containing the correct bp size of fragment were selected for overnight cultures. Plasmids were

isolated from bacterial cultures grown in the presence of selective antibiotics (LB with 100mg/L ampicillin for pGADT7 and 50 mg/L kanamycin for pGBDT7) and sequenced as described in **section 4.2.3.5**.

4.2.7.7 Yeast Transformation

Yeast Strain

Two yeast strains, Y187 and Y2HGold (Clontech), were plated on YPDA medium and incubated at 30 ˚C for 2-3 days. Fresh colonies from these plates were used for transformation.

Yeast Transformation

Transformation of yeast cells with the recombinant plasmids was performed using the Quick and Easy Yeast Transformation Mix (Clontech), according to manufacturer's instructions. Firstly, a yeast colony was scraped from a plate and was resuspended in 350 µl of sterile water in a 1.5 ml microcentrifuge tube. Then the mixture was centrifuged at 3,000 x g for 3 min to pellet the cells. In another tube, the transformation mixture was assembled as described in **Table 4.7**.

Table 4.7: Yeast transformation mixture

Reagent	Vol. Per Reaction
Quick and Easy Yeast Transformation Mix	46.5 µl
Denatured Yeastmaker Carrier DNA	2.5 µl
*Incubated at 95 °C for 5 minutes and transfer back on ice to prevent reannealing prior to use	
Transforming DNA (150 -200 ng)	1 µl

50 µl of the transformation mixture were used to resuspend the cell pellet and the suspension was then incubated at 45˚C for 65-70 min. The transformed cells suspension was diluted 20 – fold with sterile water and 100 µl were plated onto selective media (Appendix 4.3 B). The plates were incubated at 30 ˚C for 2 -3 days.

4.2.7.8 Autoactivation Experiments

Firstly, the bait was subjected to autoactivation experiments to detect any autonomous activation to the reporter genes in the absence of the prey protein. A yeast colony of pGBKT7 with each gene of interest was resuspend in 100 µl sterile H_2O and plated on SD/-Trp plates containing X-α-Gal (40 µg/ml). Plates were incubated at 30 ˚C for 2- 3

days. The development of blue-coloured colonies after 3 days would indicate autoactivation. Proteins that appeared to be autoactivated were excluded from the interaction experiments.

Two-Hybrid Screening Using Yeast Mating

In every protein interaction experiment appropriate positive and negative controls were included, according to Clontech's recommendations. A colony of different bait (pGBKT7) and prey (pGADT7) proteins were mixed and resuspended in 500 µl of YPDA broth. The assembled mixtures were left in a 30 ˚C shaker for 20 hours and then centrifuged at 3,000 x g for 3 min. 200 µl of YPDA was added to resuspend the yeast pellet and 100 µl were plated on double dropout (DDO) plates (SD/-Trp/-Leu) and on DDO plates supplemented with 40 µg/ml X-α-Gal. The presence of blue colonies would indicate a positive interaction between the two proteins. To confirm that this interaction is real, positive colonies from DDO plates resuspended in 100 µl of sterile H_2O and plated on quadruple drop out plates (QDO: SD/-Ade/-His/-leu/-Trp with X-α-Gal). For yeast growth, all plates were incubated at 30˚C for 2- 3 days.

4.2.8 Promoter Analysis in DL genes

4.2.8.1 Gene set identification

The top 500 *BdDL* co-expressed genes were extracted from the available *Brachypodium* transcriptome (refer to **Chapter 2**). The gene list (**Data S4.1**) was subjected to GO term association and enrichment analysis. Genes associated with polarity were selected. Data for enriched GO terms list for each co-expressed gene are included in **Data S4.2**.

4.2.8.2 Expression analysis

RT-PCR was used to confirm the expression profiles of the selected genes at the developmental stages that *DL* was previously found to be expressed (**Chapter 3**, **section 3.4.3**). Based on the expression analysis, a total of 5 genes were selected and were used in the promoter enrichment analysis in order to identify candidate regulatory motifs.

4.2.8.3 Motif enrichment analysis

A region of 2.5 Kb upstream the ATG was extracted from Phytozome 12 Biomart (https://phytozome.jgi.doe.gov/biomart/) for all selected genes and RSAT (http://www.rsat.eu/) was used to identify sequence motifs that were enriched in these

regions. The identified motifs were then submitted to genome ontology analysis to identify the biological processes they are more likely to be involved in.

4.3 Results and discussion

4.3.1 Generated clones

YABBY genes were cloned into different vectors for various types of experiments. The 3' UTR regions of *BdDL* and *BdINO* were used to generate the RNAi constructs, while the full coding region for both were cloned into the overexpression (OE) vectors. These constructs were used to generate *Brachypodium* transgenic lines. To the date of this thesis submission, only *BdDL* RNAi transformants had been generated and analysed.

For Yeast-two-hybrid experiments, all YABBYs (except of the BdFIL3 protein) were cloned into pray and bait vectors (pGADT7 and pGBKT7, respectively) in order to check the protein-protein interactions amongst *YABBYs* in *Brachypodium*. The recombinant plasmids were stored at -80 ˚C in order to be used for further experiments, i.e. to check the YABBY interactions with other TFs in a Y2H system.

Finally, several *YABBY* constructs were generated to be used in *Arabidopsis* mutant complementation experiments in order to shed light in the protein function conservation along plant evolution. A summary of the gene constructs generated during this project is shown in **Table 4.8**.

Table 4.8: Gene constructs generated in this project for functional analyses.

Experiment	construct	Plasmid/vector	Strain	Gene	Cloning site	Insert size(bp)	Promoter	Antibiotics/ AminoAcid resistance
RNAi								
Entry Vector	BdDL_pDNR221	pDNR221		*BdDL*	*att* B+*att* P=>*att* L	275	T7 promoter	Kanamycin (50 µg/ml)
Destination Vector	BdDL_pIPKb027	pIPKb027	AGL1	*BdDL*	*att* L+*att* R=>*att* B	275	maize ubiquitin	Spectinomycin (100 µg/ml)
Entry Vector	BdINO_pDNR221	pDNR221		*BdINO*	*att* B+*att* P=>*att* L	200	T7 promoter	Kanamycin (50 µg/ml)
Destination Vector	BdINO_pIPKb027	pIPKb027	AGL1	*BdINO*	*att* L+*att* R=>*att* B	200	maize ubiquitin	Spectinomycin (100 µg/ml)
OE								
Entry Vector	BdDL_pDNR221	pDNR221		*BdDL*	*att* B+*att* P=>*att* L	618	T7 promoter	Kanamycin (50 µg/ml)
Destination Vector	BdDL_pIPKb002	pIPKb002	AGL1	*BdDL*	*att* L+*att* R=>*att* B	618	maize ubiquitin	Spectinomycin (100 µg/ml)
Entry Vector	BdINO_pDNR221	pDNR221		*BdINO*	*att* B+*att* P=>*att* L	500	T7 promoter	Kanamycin (50 µg/ml)
Destination Vector	BdINO_pIPKb002	pIPKb002	AGL1	*BdINO*	*att* L+*att* R=>*att* B	500	maize ubiquitin	Spectinomycin (100 µg/ml)
Yeast-2-Hybrid								
Bait Clone	BdDL_pGADT7	pGADT7[1],	Y187[2]	*BdDL*	NdeI-EcoRI	618	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdDL_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdDL*	Ndel-EcoRI	618	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]
Bait Clone	BdINO_pGADT7	pGADT7[1],	Y187[2]	*BdINO*	NdeI-EcoRI	500	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdINO_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdINO*	Ndel-EcoRI	500	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]
Bait Clone	BdFIL1_pGADT7	pGADT7[1],	Y187[2]	*BdFIL1*	NdeI-EcoRI	798	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdFIL1_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdFIL1*	Ndel-EcoRI	798	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]
Bait Clone	BdFIL2_pGADT7	pGADT7[1],	Y187[2]	*BdFIL2*	NdeI-EcoRI	532	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdFIL2_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdFIL2*	Ndel-EcoRI	532	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]

Continue to next page…

Bait Clone	BdYAB1_pGADT7	pGADT7[1],	Y187[2]	*BdYAB1*	NdeI-EcoRI	513	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdYAB1_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdYAB1*	Ndel-EcoRI	513	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]
Bait Clone	BdYAB2_pGADT7	pGADT7[1],	Y187[2]	*BdYAB2*	NdeI-EcoRI	558	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdYAB2_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdYAB2*	Ndel-EcoRI	558	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]
Bait Clone	BdDL_pGADT7	pGADT7[1],	Y187[2]	*BdYAB6*	NdeI-EcoRI	516	GAL4 AD	Ampicilin[1] (100 µg/ml), Tryptophan[2*]
Prey Clone	BdDL_pGBDT7	pGBKT7[3],	Y2HGold[4]	*BdyAB6*	Ndel-EcoRI	516	GAL4 DNA-BD	Kanamycin[3] (50 µg/ml), Leucine[4*]
Complementation								
	BdDL_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdDL*	KpnI-BamHI	618	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)
	BdINO_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdINO*	KpnI-BamHI	500	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)
	BdFIL1_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdFIL1*	KpnI-BamHI	798	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)
	BdFIL2_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdFIL2*	KpnI-BamHI	532	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)
	BdYAB1_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdYAB1*	KpnI-BamHI	513	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)
	BdYAB2_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdYAB2*	KpnI-BamHI	558	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)
	BdYAB6_pCAMBIA1300	[a]pCAMBIA1300	[a,b,c]GV3101	*BdYAB3*	KpnI-BamHI	516	CaMV35S	[a]Kanamycin (50 µg/ml), [b]Gentamycin (100 µg/ml), [c]Rifampicilin (50 µg/ml)

[1,2,3,4; a,b,c] corresponding vector and its antibiotic resistant

* unknown concntration, the mix is provided with Yeast-2-hybrid kit

4.3.2 *BdDL* RNAi transgenic plants

Agrobacterium-mediated transformation was used to generate *Brachypodium DL* RNAi plants. The plant transformations were carried out at The Sainsbury Laboratory, Norwich, in collaboration with Dr Matthew Moscou and T1 seeds were shipped to Leicester University for further characterisation.

Eight independent putative transformant lines were generated (T0) and six of them were tested positive for the hygromycin selection marker gene. **Table 4.9** shows a summary of the transgenic lines generated for this project.

Table 4.9: Summary of T0 *BdDL* RNAi lines generated

ID	Accession	Backbone	callusID	Generation	T-DNA insert
BDT-00095	Bd21	pIPKb027	1	T1	Negative
BDT-00096	Bd21	pIPKb027	7	T1	Negative
BDT-00097	Bd21	pIPKb027	9	T1	Positive
BDT-00098	Bd21	pIPKb027	9	T1	Positive
BDT-00099	Bd21	pIPKb027	9	T1	Positive
BDT-00100	Bd21	pIPKb027	9	T1	Positive
BDT-00101	Bd21	pIPKb027	21	T1	Positive
BDT-00102	Bd21	pIPKb027	21	T1	Positive

Twenty plants from each line (T1 generation) were grown independently in a growth cabinet as described in **Section 3.3.1** along with wild-type control plants. Every plant was screened for the presence of the hygromycin resistance gene. **Figure 4.2a** shows a summary of the lines grown, and the number of hygromycin positive plants for each line.

However, the number of positive Hygromycin is not as expected. A single copy of gene would inherit a 1:2:1 segregation. So, theoretically 15 plants would test as positive hygromycin out of 20 plant tested. Also, the Agrobacterium transformations would give multiple copies of TDNA transferred into plant and incorporated at different parts in the genome. This will distort segregation ratios, but increase the positive hygromycin plants. Line 21-1 and 21-2 shows the segregation is distorted somehow. It is hard to speculate the cause of the segregation distortion and would need further investigation. Lacking of

information in T0 generation and limited numbers of independent lines and sample size (n=20) does not help either. However, if those ratios are real (and not just an artefact due to the small sample size) it could potentially indicate something interesting as mentioned above. And this still need further investigation with bigger sample size, bigger number of independent lines and information of T0 generation.

In addition, there is a potential that the callus selection during original selection is false positives or seed contamination somehow might affect the outcome of the transgenic plants. However, if the original selection is not false positives, having two independent line does not enough to assess the rNAi lines and to come out with robust analytical points. It is also not enough to access any background effects if they are any.

None of the lines exhibited any obvious phenotypic differences compared to the wild-type. *BdDL* expression analysis of randomly selected plants using RT-PCR did not show any reduction in the levels of *BdDL* expression (**Figure 4.2b**). For whatever reason, the RNAi might just not be working. The possible explanation for the lack of phenotype for these lines will be discussed further in the next sections.

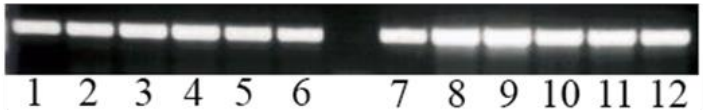
a

BdDL-RNAi genotyping summary

Lines	Sowed	+ve Hyg
21-1	20	5
7	20	0
9-4	20	12
WT	20	0
9-1	20	11
9-2	20	13
21-2	20	6
9-3	20	13
1	20	9

b

BdDL expression in RNAi +ve Hyg



BdDL	GAPDH
1 Tray3-A1	7 Tray3-A1
2 Tray3-C1	8 Tray3-C1
3 Tray5-A3	9 Tray5-A3
4 Tray6-B2	10 Tray6-B2
5 WT	11 WT
6 Tray7-A4	12 Tray7-A4

c

Primers	
control	
BdWE228_p1f	CTGC AGTC AATG ACCG CTGT
BdWE228_p1r	TGAC CTAT TGCA TCTC CCGC
Hygromycin	
Hpt_p2f	TGGC CGTG TGGA CCGC TATT
Hpt_p2r	GCTG TTGG GGAT GCCT CTGG

Figure 4.2: Summary of *BdDL* RNAi lines. (a) Genotyping of *BdDL* RNAi generated lines for the presence of the hygromycin selectable marker (+ve Hyg) **(b)** *BdDL* expression levels in transgenic lines using RT-PCR. GAPDH expression is used as a control **(c)** Primers used for genotyping the transgenic lines.

4.3.2.2 BdDL expected phenotype

Based on previous published work described in **section 1.6.4**, and on the expression pattern of *BdDL* as shown in **Section 3.4.3**, a carpel phenotype was expected to be observed in the *BdDL*-RNAi transgenic plants. Based on the conserved regulatory region between *Brachypodium* and rice, a drooping leaf phenotype was also expected in these lines, **Figure 4.3.** shows the phylogenetic footprinting of the *DL* locus between rice, *Brachypodium*, maize and sorghum, analysed by mVista (http://genome.lbl.gov). The rice gene structure is shown at the top panel and each line represents a pairwise alignment of *Brachypodium*, maize and sorghum *DL* with *OsDL*, respectively. The CNS region (pink) shows a 75% shared identity in each pairwise comparison. It has been previously suggested that the regulatory region responsible for the midrib formation in rice is located in the second intron of *DL* (Ohmori et al., 2011). **Figure 4.3** shows that the conserved regulatory regions in the second intron are similar in all tested plant species, and therefore a drooping leaf phenotype would also be expected to be observed in the transgenic *Brachypodium* plants.

The lack of phenotype in the transgenic lines needs to be further investigated. As a first step, RT-PCR was used to test if the *BdDL* expression had been reduced. The primers for the RT-PCR were designed to amplify a different region than the one selected for the cloning of the *BdDL*-RNAi construct, to allow a more accurate quantification of the endogenous transcripts. As **Figure 4.2b** shows, the *DL* expression in both transgenic and wild-type plants was similar, indicating that RNAi was not successful in these lines.

The BdDL_RNAi construct that was used for plant transformation is shown in **Figure 4.4**. The *BdDL* selected region was cloned in both orientations in order to allow the formation of a hairpin during transcription, driven by the maize ubiquitin 1 promoter (ZmUbi1). This construct was generated based on the primary transcript sequence of *BdDL*, expecting it will be enough to cause functional defects of *BdDL* in *Brachypodium*. However, the selected region appears to generate two different splice variants with a mismatch of 6 bp as shown in **Figure 4.5B**. It is possible that this difference in the transcript could decrease the efficiency of the RNA interference (Semizarov et al., 2003). The generation and use in plant transformation of a new construct using a different unique region would allow all *BdDL* transcripts to be targeted and shed more light in the function of *DL* in *Brachypodium*.

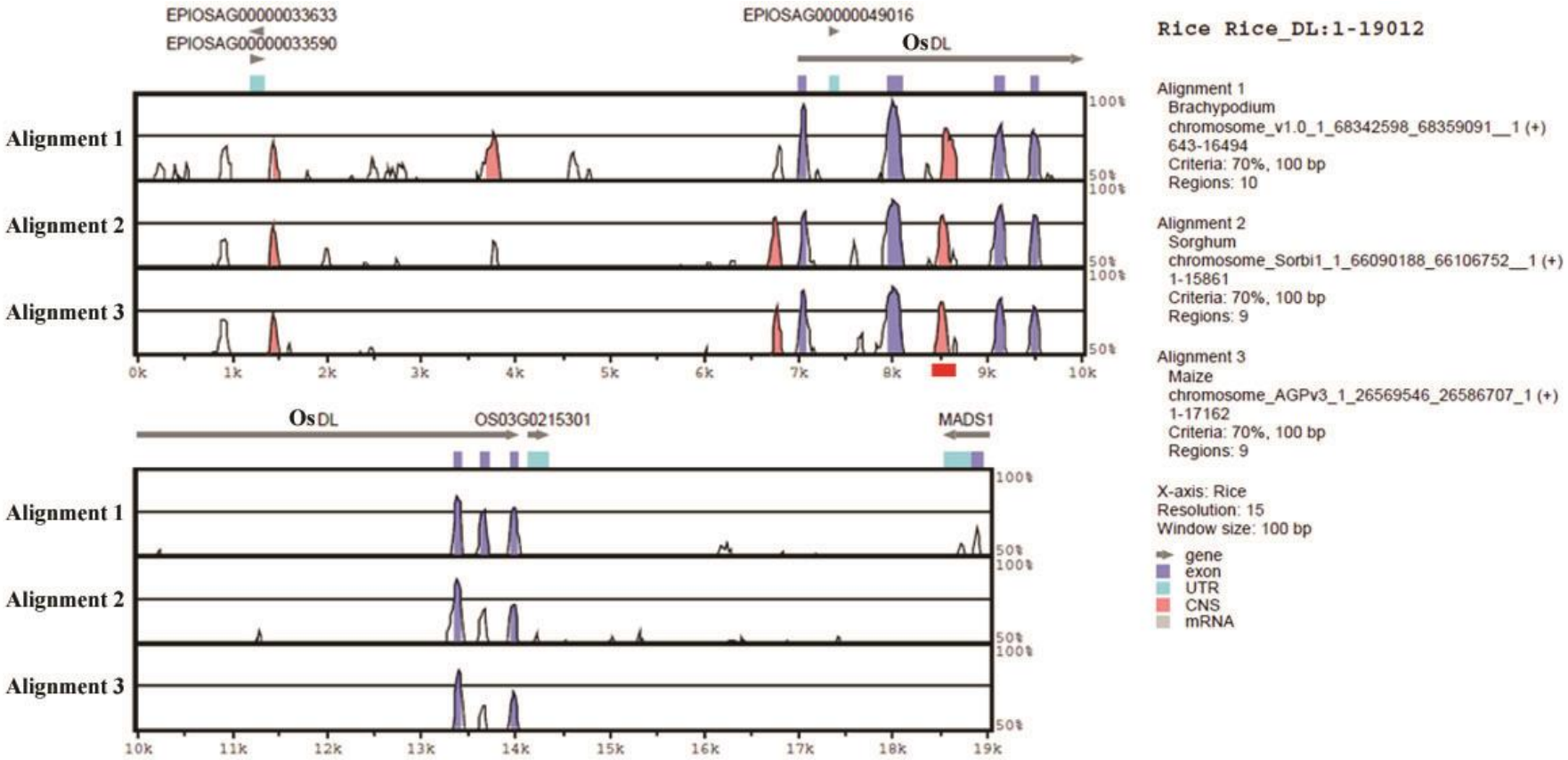


Figure 4.3: Phylogenetic footprinting of the *DROOPING LEAF* locus between rice, *Brachypodium*, maize and sorghum. Each alignment shows different pairwise comparison using *OsDL* as reference sequence.

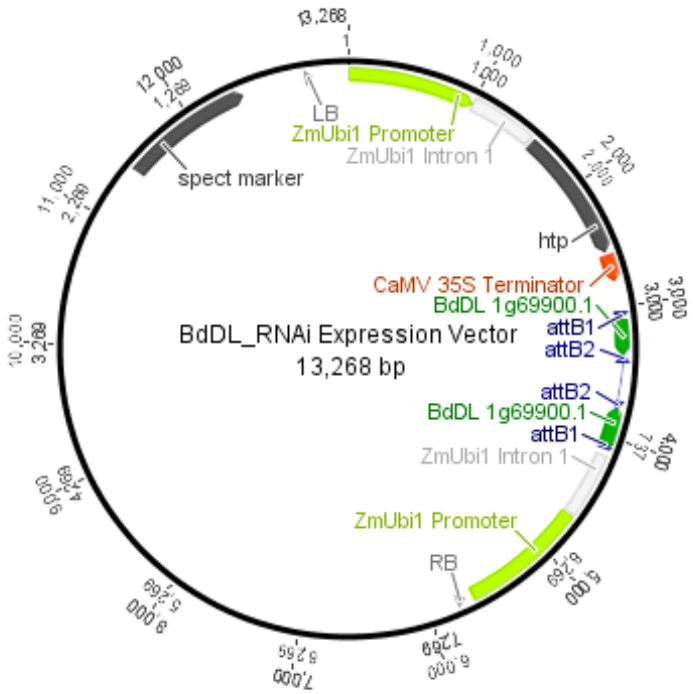


Figure 4.4: *BdDL* RNAi expression clone generated using GATEWAY® technology. DNA fragment (BdDL 1g69900.1) is cloned in two opposite orientations between the attB1 and attB2 recombination sites, resulting in the production of complimentary transcripts that can hybridise generating double stranded RNA.

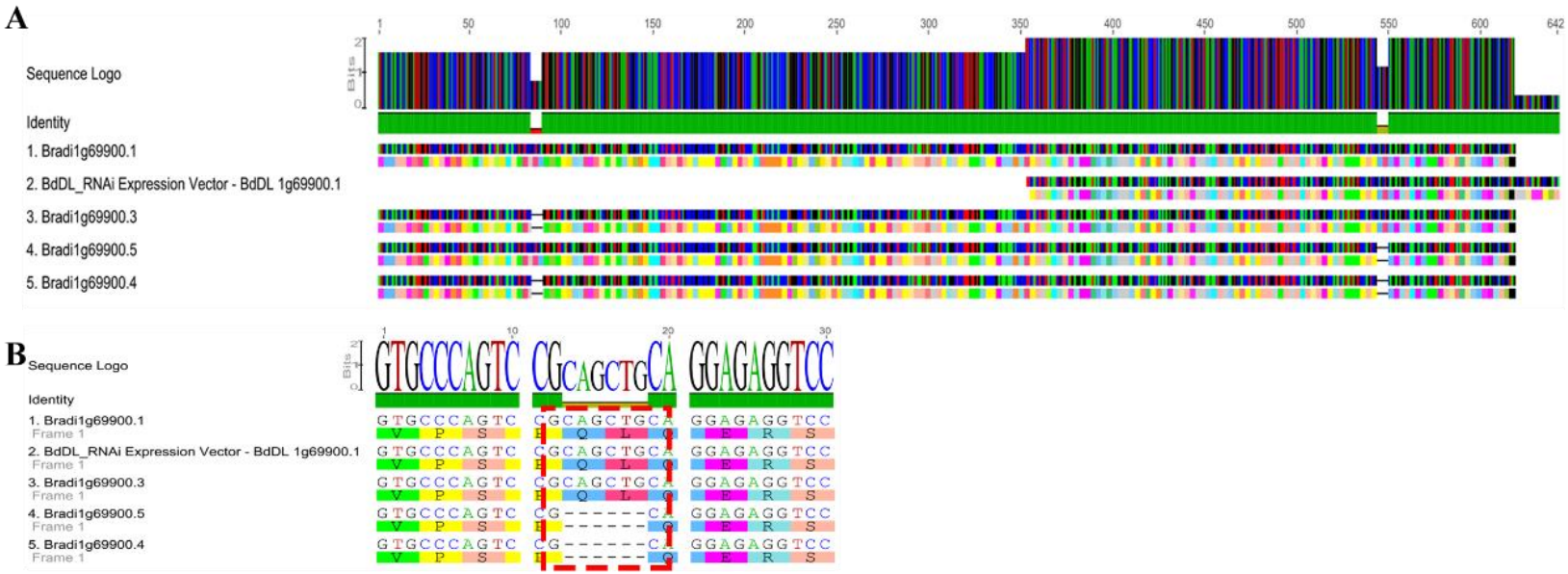


Figure 4.5: *BdDL* splice variants and DL_RNAi insert alignment. (A) The insert DNA at the end region of the transcripts, **(B)** alignment shows the spliced transcripts deletions at the cloned sites (dashed line red box).

4.3.2.3 BdDL transcripts analysis

Further investigations of *BdDL* transcript isoforms extracted from RNAseq metadata (**Chapter 2**) showed that their expression patterns slightly differ **Figure 4.6**. The primary *BdDL* transcript is not dominantly expressed in all stages while other transcripts are expressed at least as highly as the primary one in some occasions. In the same figure, only the seedling stage involves the primary and the third (Bradi1g69900.3) transcripts. During grain development, half of the transcripts are of the primary type, while the other half can be subdivided between the 3rd and 4th (Bradi1g69900.4) transcript. Finally, the 5th transcript variant is only detected at the germination stage.

The dynamics of isoform expression show that specific transcripts might be involved in different functions across plant development. In this analysis, the key function of *DL* has been reported to be the carpel formation, which would occurring at the emerging spikelet stage. Alternative splicing (AS) creates multiple transcripts from a single genes (Barbazuk 2008). Based on a review paper by Reddy (2007), several studies reported the potential of AS in biological regulations and its impact in important plant processes, such as photosynthesis, flowering, cereal grain quality and defense response. So, we need to take this into account when considering the region for DNA insert for RNAi knockdown constructs.

Hence the lack of a visible phenotype in the RNAi lines could be because of the efficiency of the selected region to target the respective *BdDL* transcripts. The isoforms for BdDL only existed in Brachypodium and an orthologue of DL in wheat. Otherwise in rice and Arabidopsis, the gene existed as a single transcript, showing the divergence of DL transcripts latter during evolution in grasses.

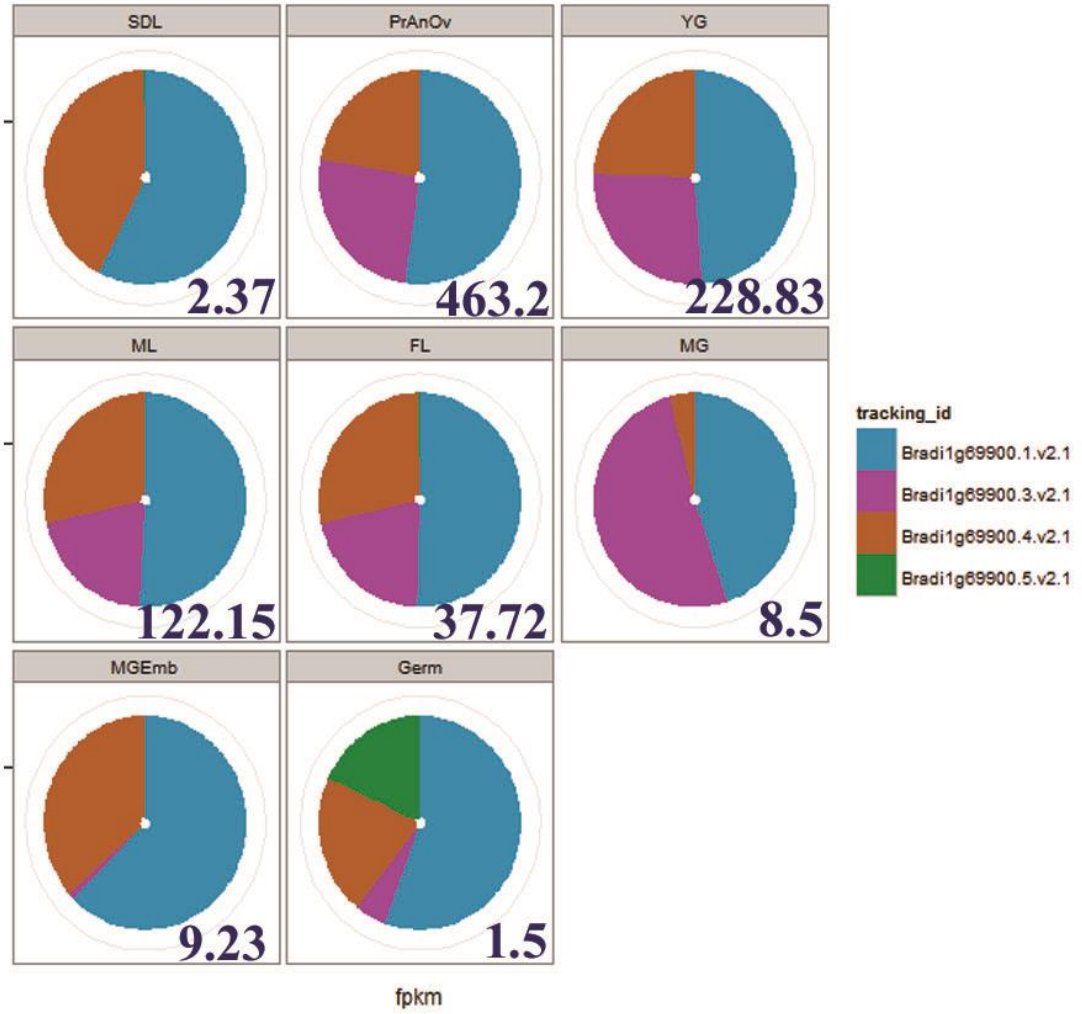


Figure 4.6: *BdDL* spliced transcript expression proportion for each developmental stage. The expression of the *BdD*L gene (in FPKM) is shown at the right bottom corner of each pie chart. SDL; seedling (16-21 DAS), PrAnOv; mature ovary (25-27 DAS), YG;young grain (1-4 DAA), ML;middle length grain (5-8 DAA), FL;full length grain (9-15 DAA), MG; mature grain (18-24 DAA), MGEmb;MG_{minus}Embryo (18-24 DAA), Germ; germinating seed (8 hours after vernalization)

4.3.3 Brachypodium FIL1 T-DNA insertion lines

This experiment was carried out to validate previous findings on *BdFIL1* reported by another PhD student in the lab, Philip Hands. The transgenic lines were supplied by the BrachyTag T-DNA mutagenesis project, initiated as part of the International *Brachypodium* Initiative (http://www/brachytag.org). The mutant lines were generated using an *Agrobacterium*-mediated transformation protocol and the binary vector pVec-GFP, developed by Alves et al. (2009).

Figure 4.7a below shows the location of the T-DNA insertion in *Brachypodium's FIL1* promoter region, which is located 156 bp upstream the start codon, along with the primers designed for genotyping (Hands, 2012).

The third generation of *FIL1* insertion line plants were grown independently under controlled conditions. The T-DNA insertion was confirmed by genotyping and *FIL1* gene expression was tested. Since a reduction in transcript levels was observed, the lack of an associated phenotype could be due to the functional redundancy between the three *BdFIL* genes. **Figure 4.7b** and **Table 4.10** show representative plant phenotypes along with their genotypes. There was a yellowing phenotype observed in some of the lines (**Figure 4.7b**), however, there was no correlation of this phenotype with the tested genotype, suggesting it could be caused by another insertion somewhere else in the genome.

The expression of *BdFIL2* and *BdFIL3*, homologues of *BdFIL1* was also investigated. Our data showed that *BdFIL2* expression was also reduced in the *BdFIL1* insertion lines. However, *BdFIL3*'s expression remained unaffected. (**Figure 4.7c**). Recent reports by Tanaka et al. (2017) showed that further knockdown of *TOB2* or *TOB3* in the *tob1* mutant (*FIL* orthologues in rice) affects spikelet development. These findings support the idea of functional redundancy between the *FIL* genes. Further work is needed to validate current results and generate transgenic lines with triple knockouts in order to better understand the function of *FIL* genes in *Brachypodium*.

FIL in other species is important for cell fate and maintenance of the meristem (Bartholmes et al., 2012, Chen et al., 1999, Vosnakis et al., 2012, Yamada et al., 2011). Its orthologue in rice, *tob1* resulting pleiotropic organ formation, where the palea was reduced or without palea, and floral meristem arrest (Tanaka et al., 2012). Since the functions of *FIL* is conserved in investigated species thus far, meristem arrest in floral organ will be expected.

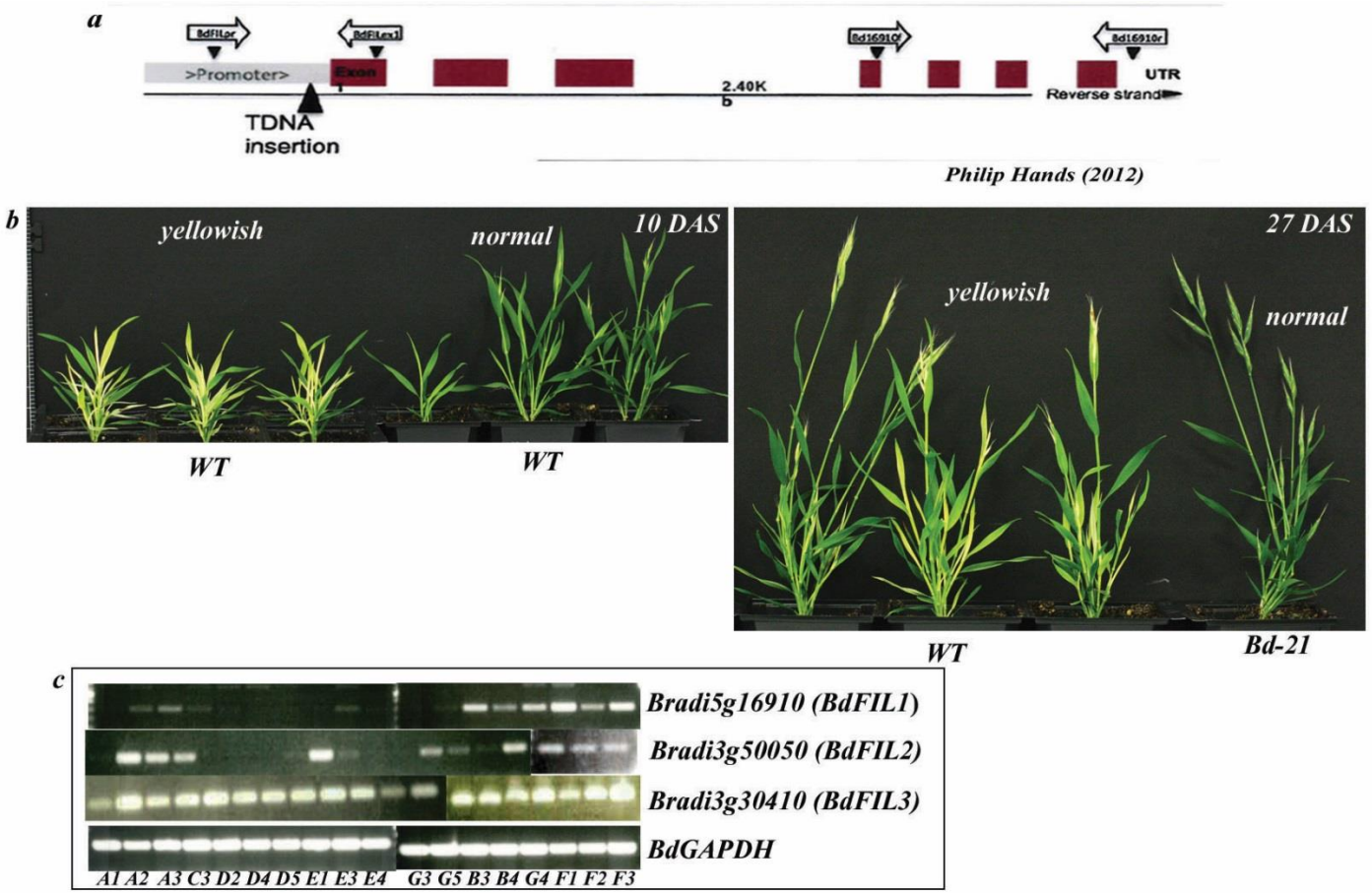


Figure 4.7: *BdFIL* T-DNA insertion lines. (a) Diagram showing the position of the T-DNA insertion in the promoter region of the *BdFIL* gene by Philip Hands (2012) **(b)** Representative yellowish versus normal phenotypes in the tested population, all of the plants are wild type (WT) in population **(c)** Expression analysis of *FIL* homologs in the tested population, *BdGAPDH* (control).

Table 4.10: Characterisation of *BdFIL* T-DNA insertion lines. The previous line name is based on previous systematic labelling in the lab, the new name is added for a shorter name, genotyped by multiplex PCR. The phenotype is based on visual observations. HOM-homozygous for the T-DNA insertion, HET-heterozygous for the T-DNA insertion, WT-wild-type (no insertion).

Previous line name	New name	Genotype	Phenotype
3-3	A1	HOM	Normal
3-4	A2	HOM	Normal
3-5	A3	HOM	Normal
3-2	B2	WT	Normal
3-3	B3	HET	Normal
3-4	B4	HET	Normal
3-5	B5	WT	Normal
3-8	C3	HOM	Normal
3-9	C4	HOM	Normal
315-13-5	D2	HOM	Normal
315-13-6	D3	HOM	Normal
315-13-7	D4	HOM	Normal
13-4	E1	HOM	Normal
13-5	E2	HOM	Normal
13-6	E3	HOM	Yellowish
13-7	E4	HOM	Normal

Previous line name	New name	Genotype	Phenotype
13-8	E5	HOM	Normal
13-9	E6	HOM	Normal
3-7	F1	WT	Normal
3-8	F2	WT	Normal
3-9	F3	WT	Normal
3-10	F4	WT	Normal
3-11	F5	WT	Normal
19-4	G1	WT	Yellowish
19-5	G2	WT	Yellowish
19-6	G3	HOM	Yellowish
19-7	G4	HET	Normal
19-8	G5	HOM	Yellowish
13-6	H1	WT	Normal
13-7	H3	WT	Normal
13-8	H4	WT	Normal
13-9	H5	WT	Normal

4.3.4 Complementation of the *Arabidopsis crc* mutant with *BdDL*

Protein functional conservation was investigated by introducing *Brachypodium* genes to their orthologue mutants in *Arabidopsis*. **Table 4.11** shows the available *Arabidopsis* mutants associated with each *YABBY* gene.

Table 4.11: Available *Arabidopsis* mutant lines. The NASC ID is the ID for purchasing seeds from the Nottingham *Arabidopsis* Stock Centre, Other name is the name of the mutant line, with its type of transformation/mutation and background genotype.

Gene	NASC ID	Other name	Type	Background	notes
INO	N3881	*Ino-1*	EMS	Ler	Essentially female sterile
	N547020	SALK-047502	T-DNA	Col-0	segregating
	N665032	SALK-063818C	T-DNA	Col-0	homozygous
FIL	N6394	*Fil1-1*	EMS	Ler	homozygous
CRC	N3814	*Crc-1*	EMS	Ler	Homozygous strong allele
	N3815	*Crc-2*	EMS	Ler	Homozygous weak allele
YAB2	N656112	SALK-003109c	T-DNA	Col-0	Homozygous knockout
	N656098	SALK-001887C	T-DNA	Col-0	Homozygous knockout
YAB3	N660999	SALK-086776C	T-DNA	Col-0	Homozygous knockout
YAB5	N668134	SALK-0451504C	T-DNA	Col-0	Homozygous knockout

Several attempts were made to introduce the *BdDL* gene into the *Arabidopsis crc* mutant. However, only 15 seedlings were recovered after the hygromycin screening. Surviving plantlets were transferred to compost for further observations. Below is the figure showing the wild type (WT) *Arabidopsis* silique and crc mutant with its carpel phenotype. General expectation of complementation experiments would be expected to form the wild type silique as shown in **Figure 4.8A** below but preliminary observation showing no phenotype complementation for positive hygromycin plants **Figure 4.8c**.

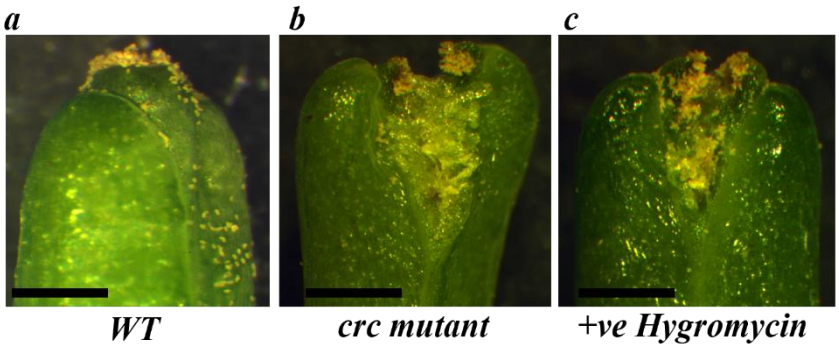


Figure 4.8: Phenotype observation of *crc* mutant. (a) Wild type silique, (b) *crc*-mutant silique (c) positive Hygromycin without complementation silique.

BdDL was expected to partially complement the *Arabidopsis crc* mutant phenotype. Previous findings in rice indicated that *OsDL* could partially compliment the *Arabidopsis* mutant phenotype (Fourquin et al., 2007). For futher work direction, PCR with *BdDL*-specific primers will have to be used to validate the presence of *BdDL* in the T1 *BdDL-crc* transgenic plants, followed by expression analysis using RT- PCR to confirm that *BdDL* is actually expressed. Then phenotype observation can be done with a good confidence that the BdDL is expressed, but it is complementing/not complementing *Arabidopsis* crc mutant phenotype. In addition, introducing BdDL gene into Arabidopsis *crc* heterozygous mutant plants would be another possible approach.

Molecular changes can be related to the amino acid sequence changes**. Figure 4.9** shows a sequence alignment of *Arabidopsis*, rice, *Brachypodium* and wheat DL/CRC proteins. We found two amino acid insertions (red circle) in the zinc finger domain of Brachypodium, rice and wheat, which is poaceae specific lineage. The function of the zinc finger domain is not specified unless broadly described. It is generally known to play a critical role in regulating signal transduction pathways, apoptosis and protein-protein interaction in eukaryotes. In addition, zinc finger domain proteins have been found to be key repressors of developmental and environmental responses (Ciftci-Yilmaz and Mittler, 2008, Englbrecht et al., 2004, Gupta et al., 2012). The additional amino acids might change the dynamics of the protein functions. Further mutant analysis is required to get an insight into the BdDL protein functional conservation.

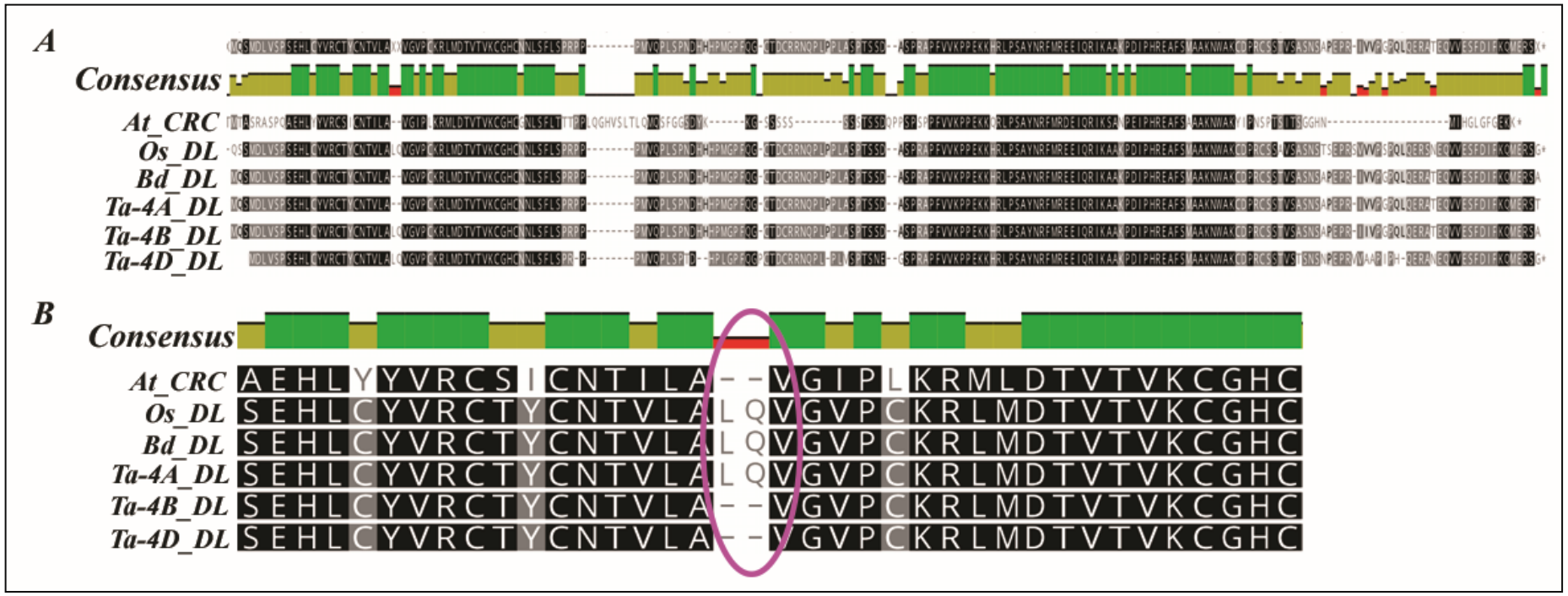


Figure 4.9: Sequence similarity and amino acid insertions in conserved domain between AtCRC, OsDL, BdDL and Ta* (wheat) proteins. **(a)** Zinc finger domain alignment between *Arabidopsis*, rice and *Brachypodium*. **(b)** Detail of zinc finger domain alignment. The violet circle indicates the insertion of two amino acids in the zinc finger domain of DL in *Brachypodium*, rice and wheat.

4.3.5 YABBY Protein-Protein Interactions

The Yeast-two-hybrid system (Clontech) was used to test the interactions between *YABBY* proteins in *Brachypodium*. In this assay, two proteins were expressed separately; the bait protein was expressed as a fusion to the Gal4 DNA-binding domain (DNA-BD), while prey proteins were expressed as fusions to the Gal4 activation domain (AD). When both of them are interact, the DNA-BD and AD will activate the transcription of four independent reporter genes (Matchmaker Gold Yeast Two-Hyrbrid System Manual). **Table 4.12** presents a summary of the YABBY protein interactions identified. Unlike in *Arabidopsis* (Stahle et al., 2009), YABBY proteins in *Brachypodium* did not interact with other YABBYs outside of their own clade. *YABBY* 2 clade in the grasses lineage interact with each other in its clade; YABBY6-YABBY1; YABBY6-YABBY6; and YABBY6-YABBY2. Interactions between proteins are pivotal for their functioning and controlling the biological process. In addition, an interaction between proteins observed in a Yeast-two-hybrid assay can only be biologically relevant when the two proteins are present in the same cells at the same time (de Folter et al., 2005). Hence, further expression localization of *YABBY1* and *YABBY2* using mRNA *in-situ* hybridization will give a clearer picture about these interactions. FIL1 and FIL2 were autoactivated and were therefore excluded from the interaction test with other proteins. It has been reported that FILAMENTOUS FLOWER in *Arabidopsis* and *Antirrhinum* also activated reporter genes in yeast (Bonaccorso et al., 2012).

Table 4.12: Protein-protein interactions between *Brachypodium* YABBY genes using Yeast 2 Hybrid

	AD						
BD	DL	INO	FIL1	FIL2	YAB1	YAB2	YAB6
DL	-	-	*	*	-	-	-
INO	-	-	*	*	-	-	-
FIL1	*	*	*	*	*	*	*
FIL2	*	*	*	*	*	*	*
YAB1	-	-	*	*	-	-	+
YAB2	-	-	*	*	-	-	+
YAB6	-	-	*	*	+	+	+

*autoactivated proteins and excluded from further analysis, - no interaction, + positive interaction

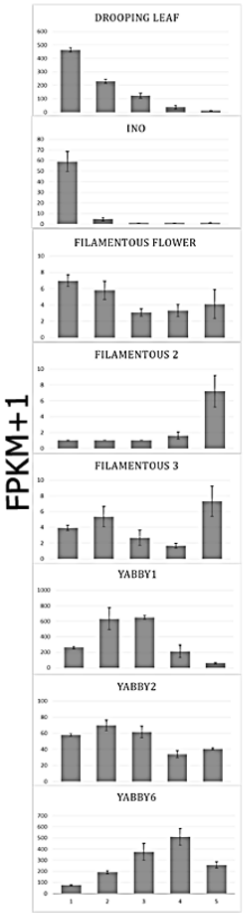
4.3.6 Motif discovery of Polarity Networks

4.3.6.1 GO related themes based on similar expression of YABBY genes

Similar expression patterns of genes i.e., gene co-expression, might indicate common involvement in particular biological functions (Doerks et al., 2002, Choi et al., 2005, Bergmann et al., 2004). Genes with similar functions usually contain similar DNA sequence motifs hence prediction of DNA sequence could lead to the discovery of the networks of desired biological processes (Pilpel et al., 2001).

Based on our transcriptomic data, the expression patterns of selected *Brachypodium YABBY* genes were used as a reference to identify the top 500 genes sharing a similar expression pattern across the plant developmental stages used in the transcriptome experiment. The enriched GO terms were subsequently analysed, with FDR <0.05 against the *Brachypodium* transcriptome GO terms. The resulting themes were related to plant development, as might be expected, except for *YABBY*6. *YABBY*6 was found to share expression patterns with many genes related to photosynthesis. *INO* is linked to genes related to exine and integument chemical pathway, while *FIL* is more related to mRNA and the regulation of flower development. *DL* on the other hand, is correlated with genes involved in polarity establishment. Since most of *YABBY*s were expressed in the pericarp tissue, GO terms involving fatty acid biosynthesis, nutrient reservoir activity, carbohydrate metabolic process and sterol biosynthetic process were expected (**Figure 4.10**).

A



B

Clade	Gene	Top Enriched GO terms (FDR < 0.05)
DL (DROOPING LEAF)	*DL*	• Fatty acid biosynthesis process • Cell wall organization • Abaxial/adaxial polarity
INO (INNER NO OUTER)	*INO*	• Cytoplasmic membrane-bounded vesicle • Extracellular region • Heme binding • Exine
FIL (FILAMENTOUS FLOWER)	*FIL1*	• Protein binding • Regulation of flower development • Mitochondrion
	FIL2	• Nutrient reservoir activity • Lipid particle • Monolayer-surrounded lipid storage body
	FIL3	• Protein auto phosphorylation • Serine family amino acid family • ATP binding • Kinase signalling pathway
YAB2 (YABBY2)	*YAB1*	• Cell wall organization • Carbohydrate metabolic process • Sterol biosynthetic process
	YAB2	• Sterol biosynthetic process
	YAB6	• Photosynthesis light harvesting • Plastid organization • Photosystem II assembly • Protein dephosphorylation

1- Pre-anthesis ovary; 2-young grain (YG: 1-3 DAA); 3-middle length grain (ML: 3-8 DAA); 4-full length grain (FL: 8-15 DAA); 5-mature grain (MG: 15-20 DAA)

Figure 4.10: GO term enrichment in *YABBY* co-expressed genes (a) Reference for similar expression pattern (n=500), **(b)** Top enriched GO terms for each gene set against the *Brachypodium* transcriptome.

To explore the networks involving polarity in the *DL* co-expressed genes, a subset gene list associated with polarity functions was created. RT-PCR was used to confirm that the expression was similar at the stage where the function of *DL* is apparent, which is mainly in the emerging spikelet stage. Out of the six genes tested, three of them were expressed at the emerging spikelet stage; *Bradi1g46450*, *Bradi5g02447* and *Bradi3g52320* (**Figure 4.11a**). A region of 2.5 Kb upstream the start codon of the selected genes was extracted and subjected to RSAT analysis. The RSAT analysis tool is publicly available online and it can be used to predict the transcription factor binding sites (TFBS) in promoters of co-regulated genes, without any prior knowledge of their binding specificity (Defrance et al., 2008). **Figure 4.11b** shows the feature map of the analysis, indicating the occurrence pattern in each upstream promoter region, which appear to be enriched with the motifs shown in **Figure 4.11c**. These motifs were submitted to GoMO to assign functional roles to DNA regulatory motifs (Bailey et al., 2006).

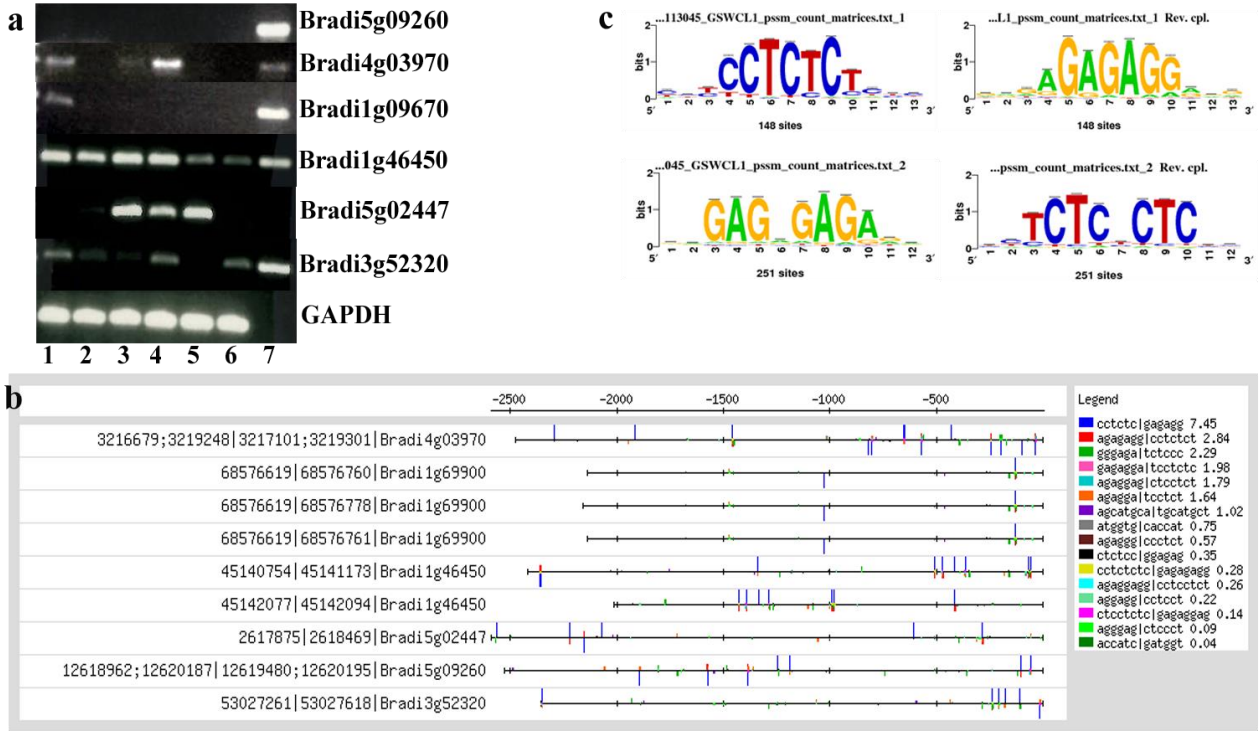
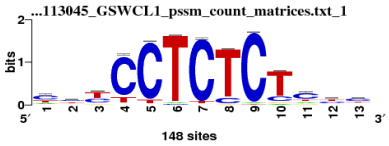


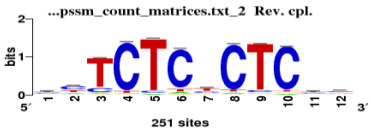
Figure 4.11: Motif discovery of polarity networks. (a) RT-PCR expression analyses for selected polarity-associated genes, 1-inflorescense; 2-emerging spikelet; 3-Pre anthesis ovary; 4-young grain (3-5 DAA); 5-palea; 6-lemma, 7-genomic DNA, **(b)** DNA motif enrichment in the 2.5 Kb region upstream of the start codon for selected gene (those expressed at the emerging spikelet stage), **(c)** enriched motifs.

DNA motifs retained from previous analysis was submitted to GoMo online software (bailey, et.al., 2009), which scans all promoters for provided nucleotide motifs to determine if any them are significantly associated with genes linked to one or more gene ontologies. The significant GO terms can then suggest the biological roles of these motifs. **Figure 4.12** and **Figure 4.13** show the motifs and their associated genome ontology terms, where 100 % implies the most specific form. The first motif is 100 % associated to polarity specification of adaxial-abaxial axis, primary shoot apical meristem and vegetative phase change, while the second motif is related to polarity specification, abaxial cell fate specification and leaf vascular tissue pattern formation. Based on these results further analysis would be required to validate the interaction between selected TFs and their predicted gene targets. Yeast 1 Hybrid experiments were initiated by inserting three tandem repeats of the motifs into a plasmid following manufacturer's recommendations (data not shown). This experiment has the potential to identify polarity networks in plants, as well as to discover novel genes that are involved in polarity complexes.



	GO term	score	*p* -value	Specificity	GO name	
	GO:0003700	8.21E-09	2.65E-07	~83%	MF	transcription factor activity
	GO:0005634	1.32E-05	2.65E-07	~2%	CC	nucleus
	GO:0045449	2.40E-04	2.65E-07	~6%	BP	regulation of transcription
	GO:0003677	5.44E-04	2.65E-07	~3%	MF	DNA binding
	GO:0005886	2.71E-03	2.65E-07	~1%	CC	plasma membrane
	GO:0005515	3.45E-03	2.65E-07	~0%	MF	protein binding
	GO:0006355	1.13E-02	1.33E-06	~8%	BP	regulation of transcription, DNA-dependent
	GO:0016563	1.49E-02	2.65E-06	~29%	MF	transcription activator activity
	GO:0009965	1.50E-02	2.92E-06	~52%	BP	leaf morphogenesis
	GO:0004674	1.56E-02	3.18E-06	~11%	MF	protein serine/threonine kinase activity
	GO:0006468	1.58E-02	3.18E-06	~18%	BP	protein amino acid phosphorylation
	GO:0007623	2.21E-02	1.06E-05	~6%	BP	circadian rhythm
	GO:0003777	2.23E-02	1.09E-05	80%	MF	microtubule motor activity
	GO:0009507	2.27E-02	1.14E-05	20%	CC	chloroplast
	GO:0010051	2.28E-02	1.14E-05	~55%	BP	xylem and phloem pattern formation
	GO:0016567	2.35E-02	1.38E-05	50%	BP	protein ubiquitination
*	GO:0009944	2.84E-02	3.93E-05	100%	BP	polarity specification of adaxial/abaxial axis
	GO:0007169	3.36E-02	8.96E-05	~12%	BP	transmembrane receptor protein tyrosine kinase signaling
	GO:0009744	3.52E-02	1.12E-04	~67%	BP	response to sucrose stimulus
	GO:0004842	3.56E-02	1.15E-04	~67%	MF	ubiquitin-protein ligase activity
	GO:0015297	3.79E-02	1.49E-04	~8%	MF	antiporter activity
	GO:0035196	3.81E-02	1.52E-04	~95%	BP	production of miRNAs involved in gene silencing by miRNA
*	GO:0010267	3.92E-02	1.69E-04	100%	BP	production of ta-siRNAs involved in RNA interference
*	GO:0010072	4.07E-02	2.01E-04	100%	BP	primary shoot apical meristem specification
*	GO:0010050	4.15E-02	2.14E-04	100%	BP	vegetative phase change
	GO:0048366	4.26E-02	2.37E-04	40%	BP	leaf development

Figure 4.12: Identified promoter sequence motif with its associated genome ontology classifications. 100 % specificity indicated the most specific ontology classification. *the most specific ontology.



	GO term	score	*p* -value	Specificit	GO name	
	GO:0003700	2.47E-09	2.65E-07	~83%	MF	transcription factor activity
	GO:0005634	6.58E-05	2.65E-07	~2%	CC	nucleus
	GO:0003677	8.08E-05	2.65E-07	~3%	MF	DNA binding
	GO:0045449	5.38E-04	2.65E-07	~6%	BP	regulation of transcription
	GO:0005515	1.40E-03	2.65E-07	~0%	MF	protein binding
	GO:0004674	1.69E-03	2.65E-07	~11%	MF	protein serine/threonine kinase activity
	GO:0006355	2.17E-03	2.65E-07	~8%	BP	regulation of transcription, DNA-dependent
	GO:0005886	3.81E-03	2.65E-07	~1%	CC	plasma membrane
	GO:0006468	4.81E-03	2.65E-07	~18%	BP	protein amino acid phosphorylation
	GO:0016563	8.93E-03	5.30E-07	~29%	MF	transcription activator activity
	GO:0048366	1.17E-02	7.96E-07	40%	BP	leaf development
	GO:0007169	1.23E-02	1.06E-06	~12%	BP	tyrosine kinase signaling pathway
	GO:0016301	1.28E-02	1.06E-06	~1%	MF	kinase activity
	GO:0048481	1.67E-02	4.77E-06	~86%	BP	ovule development
	GO:0009507	1.71E-02	6.10E-06	20%	CC	chloroplast
	GO:0004842	2.04E-02	1.43E-05	~67%	MF	ubiquitin-protein ligase activity
	GO:0045941	2.29E-02	2.20E-05	~32%	BP	positive regulation of transcription
	GO:0004672	2.29E-02	2.23E-05	~5%	MF	protein kinase activity
*	GO:0009944	2.55E-02	3.21E-05	100%	BP	polarity specification of adaxial/abaxial axis
	GO:0006511	2.66E-02	4.00E-05	~54%	BP	ubiquitin-dependent protein catabolic process
	GO:0010154	2.75E-02	4.69E-05	~19%	BP	fruit development
	GO:0010158	2.76E-02	4.75E-05	100%	BP	abaxial cell fate specification
	GO:0007623	2.77E-02	4.77E-05	~6%	BP	circadian rhythm
	GO:0009535	3.21E-02	8.41E-05	~66%	CC	chloroplast thylakoid membrane
*	GO:0005524	3.33E-02	1.02E-04	100%	MF	ATP binding
	GO:0003723	3.36E-02	1.07E-04	~2%	MF	RNA binding
*	GO:0010050	3.70E-02	1.46E-04	100%	BP	vegetative phase change
	GO:0007275	4.39E-02	2.97E-04	~0%	BP	multicellular organismal development
	GO:0010103	4.44E-02	3.11E-04	~74%	BP	stomatal complex morphogenesis
	GO:0000151	4.48E-02	3.24E-04	~22%	CC	ubiquitin ligase complex
	GO:0016564	4.54E-02	3.43E-04	~33%	MF	transcription repressor activity
*	GO:0010305	4.54E-02	3.43E-04	100%	BP	leaf vascular tissue pattern formation
	GO:0003777	4.68E-02	3.91E-04	80%	MF	microtubule motor activity
*	GO:0010152	4.72E-02	4.06E-04	100%	BP	pollen maturation

Figure 4.13: Identified promoter sequence motif with its associated genome ontology classifications. 100 % specificity indicated the most specific ontology classification. *the most specific ontology.

4.4 Conclusions and Recommendations

DROOPING LEAF RNAi lines did not show a reduction in the expression of *DL*. There was no phenotype observed and no expression reductions on selected plants. Therefore, more lines would need to be generated using RNAi and CRISPR-Cas9 could be a potential alternative approach to knock down and explore the functions of *DL* in *Brachypodium*. Genome editing through CRISPR technology allows the introduction of point mutations to the genes of interest, resulting in the introduction of a stop codon and the termination of protein translation.

Based on preliminary observation, introduction of *BdDL* to the *Arabidopsis crc* mutants did not complement the *crc* phenotype. Further expression analyses should be undertaken to investigate if *BdDL* was expressed in the transgenic plants. In addition, an insertion of two amino acids in the zinc finger domain of BdDL was identified, in comparison to *Arabidopsis* and rice. This insertion could have an effect on protein function. Thus, fusing the *BdDL* CDS to *AtCRC* promoter, and *BdDL* mutant deletion- at the insertion region fused with *AtCRC* promoter could shed more light in the protein functional implications of this two-amino acid insertion, in an evolutionary context.

Two motifs associated with polarity ontology were identified to be enriched in the promoter regions of *DL* co-expressed genes. Using a Yeast-1-hybrid system, potential proteins that interact with these motifs could be identified, adding additional knowledge to the polarity networks in plants.

Chapter 5- Conclusions and perspectives

The work in this thesis could be subdivided into two parts: the transcriptome with its global view of TFs expression profiles in addition to a focus on a particular family of transcription factors, the *YABBY* genes. The latter investigates the evolution, expression patterns and functional conservation/ diversification of *YABBY* genes in grasses using *Brachypodium* as a representative of the temperate grasses.

As mentioned above, the first part of this work involved establishing a valuable resource for grass research in the form of a comprehensive transcriptome of *Brachypodium distachyon* grain development- incorporating the two fundamental transition points of fertilization and germination. This resource is especially valuable to underpin detailed understanding of the variation and distinctiveness of grain biology in the species e.g, low starch content, thick cell walls, persistent nucellar tissues and absence of endosperm transfer cells. This research focused on the regulation of the developmental transitions and distinctive biological processes that are activated and/or repressed during grain development (**Chapter 1**), specifically with an interest in TF expression profile patterning.

The second part was focussing on diversification/conservation of key transcription factors (at sequence, expression and functional level). This explains variations in developmental and evolutionary aspects in diverse species. To achieve this, model system provide candidate TFs to perform analysis in other species. In term of our interest in TFs, we focus on the *YABBY* genes - a small family of genes with some members having key roles in fruit development. Initially it was shown that DL played key roles in grass carpel development (Yamaguchi et al., 2004). There is now evidence showing the family's importance in other aspects of crop development, regulating shattering effects in sorghum and fruit size regulation in tomato (Cong et al., 2008, Rodriguez et al., 2011, Lin et al., 2012) Using the reverse genetics approach, sequences were compared by identifying orthologues, gene family members and gene distributions across plant evolution (**Chapter 3**). Next the gene expression were compared by constructing detailed spatio-temporal expression to infer functions (**Chapter 3**). Functional characterization were done by observing phenotype of mutant or transgenic lines and finally combining

transcriptomic and molecular approach to sketch the landscape of gene networks involving *YABBY* genes (**Chapter 4**).

We believe that the understanding and discoveries of important regulator during grain development would benefits for crop improvement. This chapter will be focussing on the further project development based on findings in this thesis.

5.1 The value of *Brachypodium* as a model system for grain development

Though *Brachypodium* has been proposed as a model for temperate cereals relatively recently, it has already proven valuable in studying grain development. The detailed knowledge and characterization that we now have in *Brachypodium* make it a powerful reference point for comparison to cultivated grain crop species. As described in the introduction of **Chapter 2**, there are several features worthy of comparison related to anatomy, morphology, biological and chemical processes in *Brachypodium* (Gubatz et al., 2007, Hands et al., 2012, Sreenivasulu et al., 2010, Larre et al., 2010, Pellny et al., 2012, Trafford et al., 2013, Wobus et al., 2005, Su, 2000) in relation to crops such as barley, wheat and rice. Notable differences in *Brachypodium* grain biology are its low starch content, its globulin-dominant storage protein composition and its thick cell walls throughout the endosperm.

5.2 Generating a transcriptome of grain development – its use in *Brachypodium* and beyond

Comparative transcriptomic analyses are particularly powerful in research when there is detailed characterization of morphological/ anatomical features in the species being compared.

As a part of this project, a transcriptome comprised of eight plant developmental stages (five grain stages) was generated and analysed as described in **Chapter 2**. With the existence of comprehensive transcriptomes in rice (Itoh et al., 2016, Xue et al., 2012), wheat (Wan et al., 2008), barley (Druka et al., 2006, Sreenivasulu et al., 2008), an obvious options for future directions is to perform comparative analyses between different species to reveal the genetics underpinning the common and unique features in grain development. So with the transcriptome and its ontology establishment form this project, we can maximize available resources to infer more knowledge to find the genetic bases

of these variations, characterize the differences and highlight any novel findings amongst compared species.

On a more local taxonomic/evolutionary scale, comparative transcriptomics between *Brachypodium distachyon* Bd21, the cultivar focussed on in this thesis, and other *Brachypodium distachyon* cultivars or ecotypes could also be informative. There are collections available with cultivars showing natural variation in features such as grain structure (Garvin et al., 2008, Opanowicsz et al., 2008, Catalan et al., 2014) and it is likely that these accessions will be used to generate transcriptomic data that could be used in comparative studies.

Moving slightly beyond the *Brachypodium* genus, another avenue discussed extensively in our lab during my PhD was the lack of resources for the *Bromus* genus. *Bromus* is a genus renowned for its complex taxonomy (Scholz, 2008, Smith, 1970) while occupying a key taxonomic position in the Poaceae sister to the Triticeae and member of the Pooids (Catalan et al., 1997) – essentially the bridge between *Brachypodium* and the core pooids. *Bromus* has an intriguing, if somewhat confusing, taxonomic background that still pervades contemporary literature on the genus. *Anisantha* was initially a separate genus containing *sterilis* and *diandrus* species (Stace, 2010), though now both *Bromus* and *Anisantha* species are designated within the *Bromus* genus (Saarela et al., 2007, Cope and Gray, 2009, Hubbard, 1954). The *Anisantha/Bromus* dichotomy corresponds to two contrasting grain morphologies, most obvious being the long awns and grains of *Anisantha* and also different internal grain anatomies that have been examined in our lab. These species are economically and ecologically significant in their status as invasive species and as weed contaminants of crops such as wheat (BASF, 2006, HGCA, 2014). Even low levels of *Bromus* in wheat can reduce yields by 8% (Freer, 2012).

Extensive programs exist to document, assess and analyse the impact of *Bromus* on its habitats (both natural and farm-based) (REEnet) and also to control its levels through crop rotation and chemical methods (BASF, 2006, Freer, 2012, HGCA, 2014). We find it striking and intriguing that between these extensive research programs the detailed analyses of grain structure and development, in addition to the molecular genetic basis of *Bromus* growth and development generally, is lacking. The genus is not even included in the OneKP project which is well under way to producing transcriptomes of 1000 taxa across the angiosperms research (Matasci et al., 2014). A future project could be to generate a grain transcriptome for key *Bromus* species and provide a resource to explore

the genetic basis of grain development in the understudied genus as well as providing a wealth of sequence data for phylogenetic analyses to help clarify its taxonomy.

Another option might be to re-analyse our raw data generated in **Chapter 2** with the view to including an assessment of the long non-coding RNAs that might be involved in grain development in *Brachypodium* (Kim et al., 2017).

5.3 The evolution of the *YABBY* gene family- new insights

Phylogenetics is key to understanding how genes are evolving, identifying duplications and losses and using this information to explore the functional significance of these pattern of evolution. In this thesis, we found that complete *YABBY* genes have existed since green algae, *M. pusilla*. However, this does not answer the question on "where the *YABBY* genes coming from?" or at least predicting the sequence origin of these genes. To answer this question, Ancestral Sequence Reconstruction (ASR) is the process analysing modern sequences within an evolutionary context to infer the ancestral sequences at particular nodes of a tree (Thornton, 2004). With this in mind, this approach is similar to phylogenetic reconstruction, where we cannot know exactly the ancestral proteins, but we can predict and explain them (Hall, 2011) with a reasonable degree of confidence. Then, the estimate sequence can be used to synthesize genes encoding the sequence. Physical protein would be obtained by expressing the genes, then subsequently determine its biochemical properties. These properties would help to provide insights and construct the landscapes about cellular and environment and functional divergence of the proteins that existed from the past (Randall et al., 2016).

5.4 Future directions for *YABBY* studies

I have found that *BdYABBY6* could be a potential novel candidate to focus on in in terms of grain development based on its detailed expression pattern (both transcriptome and mRNA ISH) and capability of forming homo- and hetero-dimers with gene products in the same clade; BdYABBY1 and BdYABBY2. To date this thesis is written, no attention has been given to *YABBY6* characterization in grasses. Hence this thesis has established the expression profile at tissue and cell level, its co-expressed genes and protein-protein interactions.

Gene duplications in the *YABBY2* clade in grasses raised an interesting question. Are the duplications caused by positive selection? Or it might be caused by functional

diversification? If yes, what are the functions and how do they interact in plant complexes?

To answer the outstanding questions raised above, there are several avenues of research that could be pursued. First, in addition to *YABBY6* expression profiling at tissue level, cellular level expression analyses during grain developmental stages would be useful to observe the expression localisation. This expression is not only restricted to *Brachypodium*, yet it can be expanded to rice, maize, and other cereals such as wheat and barley. The comprehensive expression profiling and localization would be useful to interpret YABBY6 expression and conservation in grasses.

Knocking down/out genes using CRISPR/RNAi lines would give an insight of the importance of this gene during grain development. Fortunately, tilling populations with homogenous mutant of *YABBY6* in wheat are currently being generated through collaborations with Cristobal Uauy from John Innes Centre, Norwich. If the tilling population is a success, RNA-Seq data can be generated to compare the genes that might be affected prior and post tilling. The RNA-Seq data would allow us to extend the understanding about *YABBY6* in plants, and finding another gene that affected when knocking down *YABBY6* would sketch potential genes involved in *YABBY6* networks.

In addition, based on Y2H experiments, YABBY6 protein capable to interact with itself and other YABBYs in its own clade. However, Y2H experiments in **Chapter 3** were only limited to YABBY proteins in *Brachypodium* hence the experiments can be expanded to screen entire cDNA library to fish out novel proteins that interact with YABBY6.

Likewise in polarity network pipeline as described in **Chapter 3**, the same workflow can be used to identify the photosynthesis networks involving *YABBY6*. This experiments will generate DNA motif sequences that enriched in the promoter region of co-expressed genes. Hence using Y1H approach screening against cDNA library might reveal another potential regulators and genes involved in photosynthesis complexes in plants.

Appendices

Appendix 3.1 – Gene orthologs and its ID in phylogenetic tree

Table A.1 : Gene orthologs and its ID in phylogenetic tree. Table clustered the gene based on *YABBY* genes subfamily *-INNER NO OUTER (INO), CRABS CLAW/DROOPING LEAF (CRC/DL), YABBY2, YABBY5 and FILAMENTOUS FLOWER (FIL)*. Gene ID from various database, genome assembly and online resource attached.

Clade	Tree_ID	LocusName/AcessionNumber	Species (assembly)	Common name	Source
Outgroup					
	Micpusilla_YAB2	MicpuC2.EuGene.0000040541	Micromonas pusilla CCMP1545	green algae	https://phytozome.jgi.doe.gov/
	Micsp._YAB1	AZW_EuGene.0200010417	Micromonas sp. RCC299	green algae	https://phytozome.jgi.doe.gov/
INNER NO OUTER (INO)					
	AmTr_INO-like	evm_27.model.AmTr_v1.0_scaffold00096.22	Amborella trichopoda v1.0	Amborella	https://phytozome.jgi.doe.gov/
	At_INO	AT1g23420	Arabidopsis thaliana (TAIR10)	Arabidopsis	https://phytozome.jgi.doe.gov/
Eudicots	Bradi_INO	Bradi1g24051	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_DL	Brast06G084400.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	BrR_INOa	Brara.G00980	Brassica rapa FPsc v1.3	Mustard	https://phytozome.jgi.doe.gov/
	BrR_INOb	Brara.I03229	Brassica rapa FPsc v1.4	Mustard	https://phytozome.jgi.doe.gov/
Monocots	Musa_INO	GSMUA_Achr5T23610_001	Musa acuminata v1	Banana	https://phytozome.jgi.doe.gov/
	Os_YAB7	LOC_Os07g38410.1	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_INO	Sobic.002G346800.v2.1	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
	Sevir_INO	Sevir.2G372300.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
Poaceae	Si_YAB7	Seita.2G361300	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Solyc_INO	Solyc05g005240	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
	Ta_INO	TRIAE_CS42_2DS_TGACv1_178838_AA0601610	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Zm_YAB2	GRMZM2G046829_T01	Zea mays (Ensembl-18)	Maize	
CRABS CLAW (CRC) / DROOPING LEAF (DL)					
	AmTr_CRC	evm_27.model.AmTr_v1.0_scaffold00047.126	Amborella trichopoda (v1.0)	Amborella	https://phytozome.jgi.doe.gov/
Eudicots	At_CRC	AT1g69180	Arabidopsis thaliana (TAIR10)	Arabidopsis	https://phytozome.jgi.doe.gov/
	Bradi_DL	Bradi1g69900	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_DL	Brast02G095400.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Musa_DLa	GSMUA_Achr1T04480_001	Musa acuminata v1	Banana	https://phytozome.jgi.doe.gov/
	Musa_DLc	GSMUA_Achr4T01430_001	Musa acuminata v2	Banana	https://phytozome.jgi.doe.gov/
	Musa_DLd	GSMUA_Achr1T27150_001	Musa acuminata v3	Banana	https://phytozome.jgi.doe.gov/
Monocots	Musa_DLe	GSMUA_Achr6T24550_001	Musa acuminata v4	Banana	https://phytozome.jgi.doe.gov/
	Musa_DLf	GSMUA_Achr3T25290_001	Musa acuminata v5	Banana	https://phytozome.jgi.doe.gov/
	Musa_DLg	GSMUA_Achr8T04340_001	Musa acuminata v6	Banana	https://phytozome.jgi.doe.gov/
	Musa_YAB2c	GSMUA_Achr5T08930_001	Musa acuminata v7	Banana	https://phytozome.jgi.doe.gov/
	Os_DL	LOC_Os03g11600	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_DL	Sobic.001G456000	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
	Sevir_DL	Sevir.9G494400.2	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
	Si_DL	Seita.9G490500.1	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Solyc_CRCa	Solyc_5g012050	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
Poaceae	Solyc_CRCb	Solyc_1g010240	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
	Ta_DL-4A	TRIAE_CS42_4AS_TGACv1_306676_AA1011910	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Ta_DL-4B	TRIAE_CS42_4BL_TGACv1_321049_AA1054410	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Ta_DL-4D	TRIAE_CS42_4DL_TGACv1_343031_AA1127860	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Zm_Dla	GRMZM2G088309_T01	Zea mays (Ensembl-18)	Maize	
	Zm_DLb	GRMZM2G102218_T01	Zea mays (Ensembl-18)	Maize	

FILAMENTOUS FLOWER					
	AmTr_FIL-like	evm_27.model.AmTr_v1.0_scaffold00085.9	Amborella trichopoda (v1.0)	Amborella	https://phytozome.jgi.doe.gov/
	At_YAB1/FIL/AFO	AT2g45190	Arabidopsis thaliana (TAIR10)	Arabidopsis	https://phytozome.jgi.doe.gov/
	At_YAB3	AT4g00180	Arabidopsis thaliana (TAIR10)	Arabidopsis	https://phytozome.jgi.doe.gov/
	BrR_FILa	Brara.D02729	Brassica rapa FPsc v1.3	Mustard	https://phytozome.jgi.doe.gov/
Eudicots	BrR_FILb	Brara.C02288	Brassica rapa FPsc v1.4	Mustard	https://phytozome.jgi.doe.gov/
	BrR_FILc	Brara.G01769	Brassica rapa FPsc v1.5	Mustard	https://phytozome.jgi.doe.gov/
	BrR_FILd	Brara.I00030	Brassica rapa FPsc v1.6	Mustard	https://phytozome.jgi.doe.gov/
	Solyc_FILa	Solyc_1g091010	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
	Solyc_FILb	Solyc_8g079100	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
	Musa_FILa	GSMUA_Achr7T07130_001	Musa acuminata v1	Banana	https://phytozome.jgi.doe.gov/
Monocots	Musa_FILb	GSMUA_Achr8T11580_001	Musa acuminata v2	Banana	https://phytozome.jgi.doe.gov/
	Musa_FILc	GSMUA_Achr11T03800_001	Musa acuminata v3	Banana	https://phytozome.jgi.doe.gov/
	Musa_FILd	GSMUA_Achr11T17520_001	Musa acuminata v4	Banana	https://phytozome.jgi.doe.gov/
	Bradi_FIL1	Bradi5g16910	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_FIL1	Brast09G156300.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Hv_FIL1	Hv_18466	INSDC-GCA_000326085.1 (2012)	Barley	http://plants.ensembl.org/
	Os_YAB5	LOC_Os04g45330.1	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_FIL1	Sobic.006G160800	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
	Sevir_FIL1	Sevir.7G191100.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
Poaceae (FIL1)	Si_FIL1	Seita.7G181400	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Ta_2A_FIL1	TRIAE_CS42_2AL_TGACv1_094021_AA0291200	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Ta_2B_FIL1	TRIAE_CS42_2BL_TGACv1_130187_AA0405790	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Ta_2D_FIL1	TRIAE_CS42_2DL_TGACv1_160252_AA0548640	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Zm_FIL1	GRMZM2G054795_T01	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
	Zm_YAB14	GRMZM2G005353_T01	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
	Bradi_FIL2	Bradi3g50050	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_FIL2	Brast04G123800.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Hv_FIL2	Hv_60897	INSDC-GCA_000326085.1 (2012)	Barley	http://plants.ensembl.org/
	Os_YAB4	LOC_Os02g42950	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_YAB4	Sobic.004G302700	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
Poaceae (FIL2)	Sevir_FIL2	Sevir.1G255800.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
	Si_YAB4	Seita.1G251600	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Ta_6A_FIL2	TRIAE_CS42_6AL_TGACv1_471626_AA1511990	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Ta_6B_FIL2	TRIAE_CS42_6BL_TGACv1_499794_AA1591850	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Ta_6D_FIL2	TRIAE_CS42_6DL_TGACv1_528528_AA1714780	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Zm_FIL2	GRMZM2G529859_T02	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
	Bradi_FIL3	Bradi3g30410	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_FIL3	Brast03G169300.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Hv_FIL3	HV_70653	INSDC-GCA_000326085.1 (2012)	Barley	http://plants.ensembl.org/
	Os_YAB3	LOC_Os10g36420.1	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_YAB3	Sobic.001G199200.1	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
	Sevir_FIL3	Sevir.9G196700.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
poaceae (FIL3)	Si_YAB3	Seita.9G197300.1	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Ta_1A_FIL3	TRIAE_CS42_1AL_TGACv1_000156_AA0004990	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Ta_1B_FIL3	TRIAE_CS42_1BL_TGACv1_033155_AA0137410	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Ta_1D_FIL3	TRIAE_CS42_1DL_TGACv1_061067_AA0183920	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Zm_YAB10	GRMZM2G167824_T01	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
	Zm_YAB9	GRMZM2G074543_T03	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/

YABBY2					
Eudicots	AmTr_Yab2-like	evm_27.model.AmTr_v1.0_scaffold00004.182	Amborella trichopoda (v1.0)	Amborella	https://phytozome.jgi.doe.gov/
	At_YAB2	AT1g08465	Arabidopsis thaliana (TAIR10)	Arabidopsis	https://phytozome.jgi.doe.gov/
	BrR_YAB2	Brara.F00527	Brassica rapa FPsc v1.3	Mustard	https://phytozome.jgi.doe.gov/
	BrR_YAB2a	Brara.I05306	Brassica rapa FPsc v1.4	Mustard	https://phytozome.jgi.doe.gov/
	BrR_YAB2b	Brara_030728	Brassica rapa FPsc v1.5	Mustard	https://phytozome.jgi.doe.gov/
	BrR_YAB2c	Brara.H02872	Brassica rapa FPsc v1.6	Mustard	https://phytozome.jgi.doe.gov/
	Solyc_YAB2a	Solyc_6g073920	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
	Solyc_YAB2b	Solyc_11g071810	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
Monocots	Musa_YAB2a	GSMUA_Achr7T01330_001	Musa acuminata v2	Banana	https://phytozome.jgi.doe.gov/
	Musa_YAB2b	GSMUA_Achr6T31080_001	Musa acuminata v3	Banana	https://phytozome.jgi.doe.gov/
	Musa_YAB2c	GSMUA_Achr5T08930_001	Musa acuminata v4	Banana	https://phytozome.jgi.doe.gov/
	Musa_YAB2d	GSMUA_Achr4T22750_001	Musa acuminata v5	Banana	https://phytozome.jgi.doe.gov/
	Musa_DLb	GSMUA_Achr7T07830_001	Musa acuminata v6	Banana	https://phytozome.jgi.doe.gov/
Poaceae (YAB1)	Bradi_YAB1	Bradi1g57070	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_YAB1	Brast06G214100.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Os_YAB1	LOC_Os07g06620.2	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_YAB1	Sobic.002G039400.1	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
	Sevir_YAB1	Sevir.2G042800.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
	Si_YAB1	Seita.2G037900.1	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Ta_2B_YAB1	TRIAE_CS42_2BL_TGACv1_130161_AA0405330	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Zm_YAB1	GRMZM2G106204_T02	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
Poaceae (YAB2)	Bradi_YAB2	Bradi1g13570	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_YAB2	Brast02G248100.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Hv_YAB2	Hv_70676	INSDC-GCA_000326085.1 (2012)	Barley	http://plants.ensembl.org/
	Os_YAB2	LOC_Os03g44710.1	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sevir_YAB2	Sevir.9G153200.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
	Si_YAB2	Seita.9G154300.1	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Ta_5A_YAB2	TRIAE_CS42_5AL_TGACv1_374236_AA1194610	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Ta_5B_YAB2	TRIAE_CS42_5BL_TGACv1_404477_AA1301190	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Ta_5D_YAB2	TRIAE_CS42_5DL_TGACv1_433350_AA1410470	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Zm_YAB2a	GRMZM2G074124_T05	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
	Zm_YAB2b	GRMZM2G085873_T01	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
Poaceae (YAB6)	Bradi_YAB6	Bradi4g01300	Brachypodium distachyon (v3.1)	Purple false brome	https://phytozome.jgi.doe.gov/
	Brast_YAB6	Brast10G011700.1	Brachypodium stacei (v1.1)	B.stacei	https://phytozome.jgi.doe.gov/
	Hv_YAB6	Hv_66260	INSDC-GCA_000326085.1 (2012)	Barley	http://plants.ensembl.org/
	Os_YAB6	LOC_Os12g42610.1	Oryza sativa (v7_JGI)	Rice	https://phytozome.jgi.doe.gov/
	Sb_YAB6	Sobic.008G176300.1	Sorghum bicolor (v3.1)	Cereal grass	https://phytozome.jgi.doe.gov/
	Sevir_YAB6	Sevir.3G405000.1	Setaria viridis (v1.1)	Green foxtail	https://phytozome.jgi.doe.gov/
	Si_YAB6	Seita.3G388300.1	Setaria italica (v2.2)	Foxtail millet	https://phytozome.jgi.doe.gov/
	Ta_5A_YAB6	TRIAE_CS42_5AS_TGACv1_393321_AA1271170	Triticum aestivum (TGAC v1)	Wheat	https://www.ncbi.nlm.nih.gov/
	Ta_5B_YAB6	TRIAE_CS42_5BS_TGACv1_423385_AA1375620	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Ta_5D_YAB6	TRIAE_CS42_5DS_TGACv1_457157_AA1483350	Triticum aestivum (TGAC v1)	Wheat	https://phytozome.jgi.doe.gov/
	Zm_YAB6	GRMZM2G141955_T01	Zea mays (Ensembl-18)	Maize	https://phytozome.jgi.doe.gov/
YABBY5					
Eudicots	AmTr_YAB5-like	evm_27.model.AmTr_v1.0_scaffold00078.9	Amborella trichopoda (v1.0)	Amborella	https://phytozome.jgi.doe.gov/
	At_YAB5	AT2g26580	Arabidopsis thaliana (TAIR10)	Arabidopsis	https://phytozome.jgi.doe.gov/
	BrR_YAB5	Brara.C02454	Brassica rapa FPsc v1.3	Mustard	https://phytozome.jgi.doe.gov/
	Solyc_YAB5a	Solyc_7g008180	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
	Solyc_YAB5c	Solyc_12g009580	Solanum lycopersicum (iTAG2.3)	Tomato	https://phytozome.jgi.doe.gov/
Gymnosperm	Pa_YAB1	MA_10432332g0010	Picea abies 1.0	P.abies	http://congenie.org/citation
	Pa_YAB2	MA_407206g0010	Picea abies 1.1	P.abies	http://congenie.org/citation
	Pa_YAB3	MA_112273g0010	Picea abies 1.2	P.abies	http://congenie.org/citation

Appendix 3.2 – Plant materials used in RT-PCR and in-situ Hybridization

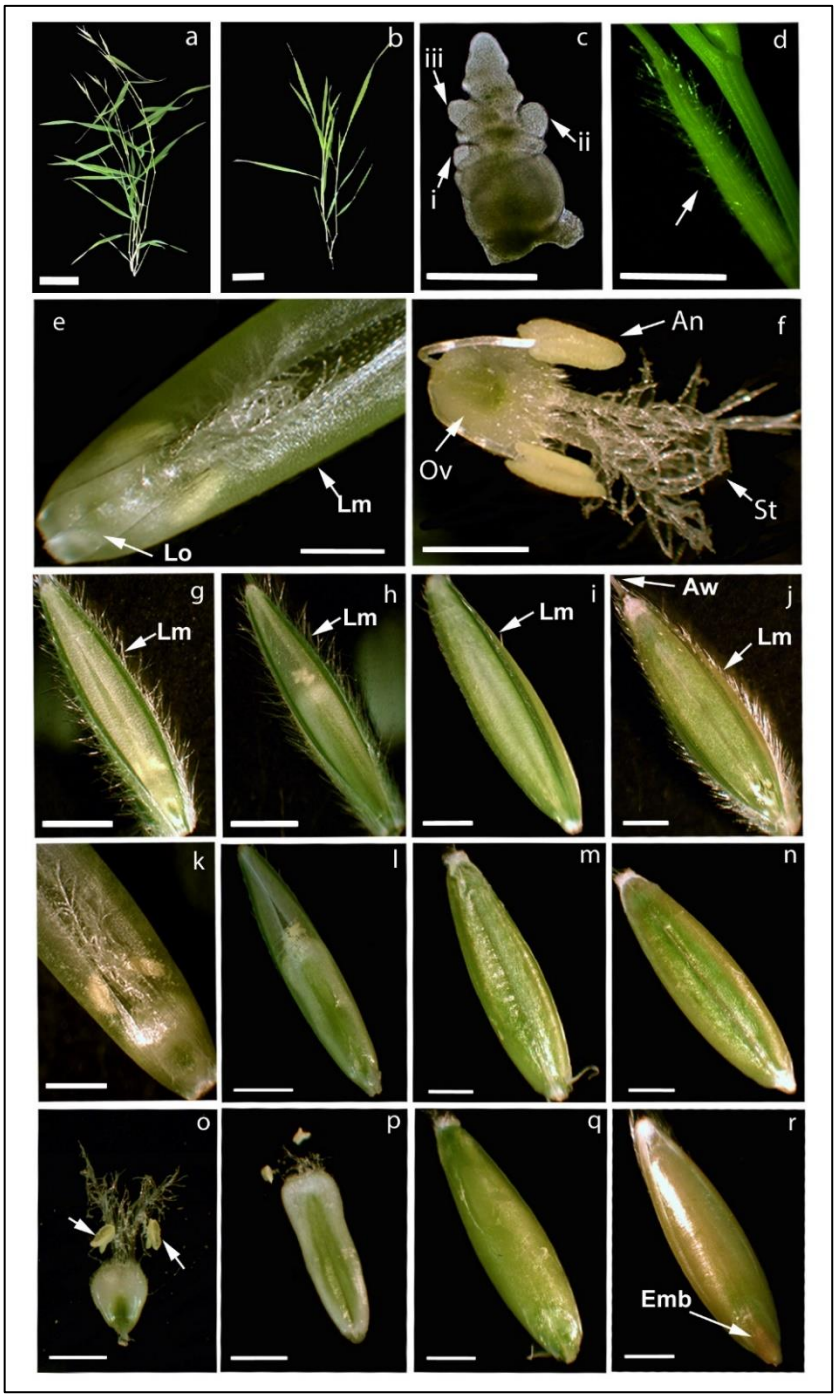


Figure A.1: ***Brachypodium*** **plant stages**. (a) Whole plant around 35 DAS. Scale; 3.0 cm (b)vegetative stage of plant, 16-21 DAS. Scale:3.0 cm (c)very young inflorescence, with three IM (arrow indicated), 18-21 DAS. Scale: 0.6 mm (d)emergent spikelet structure, 22-25 DAS. Scale,0.3 cm (e, f) pre-anthesis stage, 25-27 DAS. Scale,0.3 cm (g,k,o)young grain stage, 1-4 days after anthesis (DAA) Scale,0.4 cm. (h,l,p) middle length grain, 5-8 DAA. Scale,0.4 cm (I,m,q) full length grain, 9-15 DAA. Scale,0.4 cm (j,n,r) mature grain 18-24 DAA. Scale,0.4 cm. An;anther, Emb;embryo, Lm;lemma, Lo;lodicules, Ov;ovary, St;stamen

Appendix 3.3 – Primers used in YABBY evolution and expression profiling chapter (Chapter 3)

Gene name	Bd ortholog
DL	Bradi1g69900
INO	Bradi1g24050
FIL-1	Bradi5g16910
FIL-2	Bradi3g50050
FIL-3	Bradi3g30410
YAB1	Bradi1g57070
YAB2	Bradi1g13570
YAB6	Bradi4g01300

Primers used for experiments in Chapter 3

Primer	sequence (5'-3')
RT-PCR	
BdDL_F	TTCCATCTGCTTACAATCGC
BdDL_R	CATTCATGATGCCGGGATCA
BdINO_F	GGC GTC CTC TCC GTC AGC GT
BdINO_R	GCT CTC CGC AGC CTC TCC CT
BdFIL1_F	CGGCCATACAACCGATTTATC
BdFIL1_R	AGCTAGAATGCACCTCAAAG
BdFIL2_F	ATCCGCGTACAACCGCTTCA
BdFIL2_R	GACATTGGCAAAGAACTTG
BdFIL3_F	TCGGCATACAACCGCTTCAT
BdFIL3_R	GAAGCTCGGCGATGATGTAG
BdYAB1_F	CGATGGATATACCTTGCGAT
BdYAB1_R	AAAATTGTCCTAGGCACGTA
BdYAB2_F	CACTAGCCTGTTGTCGGTTA
BdYAB2_R	AATGTTTGGAAAATGCGCCC
BdYAB6_F	ATAAAAGCAAACAACCCCGAC
BdYAB6_R	AGCATGCAATTTCTCGTCTT
BdGAPDH_F	ACATCATTCCTAGCAGCACTGG
BdGAPDH_R	TGACGAAATGGTCGTTCAGAGC

Appendix 3.4 – YABBY expression FPKM value

Table A.2: *YABBY* genes expression value from transcriptome, at 7 plant stages pre-anthesis ovaries, young grains at 1-3 DAA, middle length grains at 3-8 DAA, full length grains at 8-15 DAA and mature grain at 15-20 DAA, seedling at 2 days after germination (DAG), germination at 5 ~12 hours after stratification.

	PrAnOv	YG	ML	FL	MG	GERM	SDL
INO	62.141	4.65555	0	0	0	0	0.251334
	47.0104	1.61715	0	0	0.212399	0	0
	64.5093	4.77385	0	0	0	0.233912	0
SUM	173.6607	11.04655	0	0	0.212399	0.233912	0.251334
STDEV	9.493466	1.789349	0	0	0.122629	0.135049	0.145108

	PrAnOv	YG	ML	FL	MG	GERM	SDL
YAB2	57.1531	64.6309	54.3232	37.7768	40.3783	15.4349	15.0597
	58.5737	65.5973	68.2976	28.8545	38.5924	18.7844	21.4439
	55.4882	76.2727	59.7337	32.8017	40.2791	17.593	21.8116
SUM	171.215	206.5009	182.3545	99.433	119.2498	51.8123	58.3152
STDEV	1.544361	6.460516	7.046249	4.471007	1.00368	1.697841	3.79652

	PrAnOv	YG	ML	FL	MG	GERM	SDL
DL	444.938	212.433	144.048	50.4467	11.5073	0.752297	1.87548
	478.075	241.715	113.994	32.9382	8.14185	1.75799	3.13179
	466.477	232.345	108.407	26.7784	6.08945	0.89377	2.09612
SUM	1389.49	686.493	366.449	110.1633	25.7386	3.404057	7.10339
STDEV	16.81519	14.95393	19.16915	12.27924	2.735315	0.544412	0.670772

	PrAnOv	YG	ML	FL	MG	GERM	SDL
YAB1	246.335	471.895	654.98	305.36	70.7165	3.48501	12.7663
	261.857	729.324	672.227	151.809	53.266	4.56117	15.0484
	277.245	695.274	622.018	179.459	51.8036	3.08538	16.8199
SUM	785.437	1896.493	1949.225	636.628	175.7861	11.13156	44.6346
STDEV	15.45505	139.8376	25.5111	81.8469	10.52265	0.763302	2.032153

	PrAnOv	YG	ML	FL	MG	GERM	SDL
YAB6	79.9039	187.728	292.093	576.559	287.359	1.6351	6.47364
	73.1509	205.447	445.624	426.31	249.769	1.65063	7.30481
	72.1755	180.095	384.21	525.199	234.016	1.83328	8.45267
SUM	225.2303	573.27	1121.927	1528.068	771.144	5.11901	22.23112
STDEV	4.208772	13.00609	77.27547	76.36715	27.40633	0.11021	0.993729

	PrAnOv	YG	ML	FL	MG	GERM	SDL
FIL2	0.030013	0	0	0.124675	7.53571	18.2996	6.91475
	0	0	0	0.726104	7.16921	19.0831	6.27685
	0	0	0	1.00029	3.91069	11.6759	5.49597
SUM	0.030013	0	0	1.851069	18.61561	49.0586	18.68757
STDEV	0.017328	0	0	0.447883	1.995539	4.069273	0.71059

	PrAnOv	YG	ML	FL	MG	GERM	SDL
FIL1	6.07689	6.06346	1.58008	1.77647	4.98205	4.47717	6.81538
	6.63563	4.44655	2.38883	1.94761	2.83129	4.65594	8.94723
	5.18069	3.87454	2.26453	3.15273	1.51427	3.16352	4.88128
SUM	17.89321	14.38455	6.23344	6.87681	9.32761	12.29663	20.64389
STDEV	0.733964	1.135265	0.435507	0.750077	1.750515	0.814959	2.033776

	PrAnOv	YG	ML	FL	MG	GERM	SDL
FIL3	3.18767	3.47728	2.7829	0.36035	8.46928	4.66284	1.67341
	3.08904	5.87097	1.06051	0.65787	5.80557	3.90643	1.15
	2.56559	3.78629	1.17253	0.968813	4.74262	3.8569	1.2905
SUM	8.8423	13.13454	5.01594	1.987033	19.01747	12.42617	4.11391
STDEV	0.334343	1.301994	0.963714	0.304256	1.919775	0.451691	0.270899

Appendix 4.1 – GATEWAY cloning vectors information

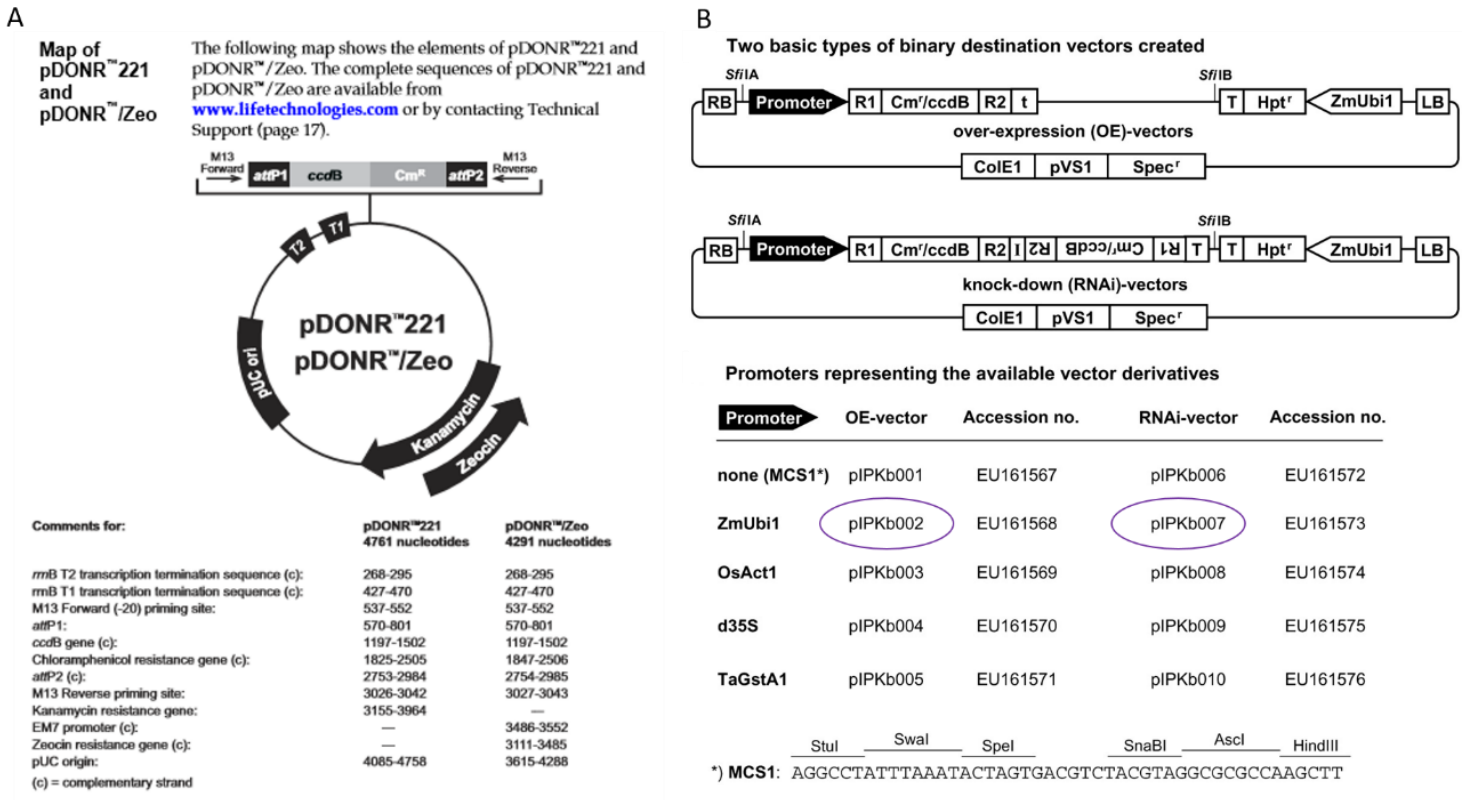


Figure A.2: Vectors information. (A) pDNR221 as Entry clone showing the map with attp1-attp2 cloning sites,Kanamycin antibiotic resistant plasmid. (B) pIPKb002 (OE-vector) and pIPKb007 (RNAi-vector) as destination vector, with ZmUni1 as promoter and Spectinomycin antibiotics resistant.

Appendix 4.2 – Primers used in Functional chapter (Chapter 4)

Primers used for experiments in Chapter 4

Primer	sequence (5'-3')
GATEWAY	
BdDL_RNAi-attB1	AAAAAGCAGGCTAC TTCCATCTGCTTACAATCGC
BdDL_RNAi-attB2	AGAAAGCTGGGTC TGATCCCGGCATCATGAATG
BdDL_OE-attB1	AAAAAGCAGGCTAC ATGCAGCAGAGCAGCATGGA
BdDL_OE-attB2	AGAAAGCTGGGTC CTAGCCGCTGCGCTCCATCT
BdINO_RNAi-attB1	AAAAAGCAGGCTAC GGCGTCCTCTCCGTCAGCGT
BdINO_RNAi-attB2	AGAAAGCTGGGTC GCTCTCCGCAGCCTCTCCCT
BdINO_OE-attB1	AAAAAGCAGGCTAC ATGTCGTCGTCGCTCGTCCA
BdINO_OE-attB2	AGAAAGCTGGGTCT CAGTTCCCCCTGTGCTGGA
Yeast-2-Hybrid	
BdDL-Ndel	TAACGACATATGATGCAGCAGAGCAGCATGGATT
BdDL_EcoRI	GAATTCCATATGCTAGCCGCTGCGCTCCATCTGC
BdINO_Ndel	TAACGACATATGATGTCGTCGTCGCTCGTCCA
BdINO_EcoRI	GAATTCCATATGTCAGTTCCCCCTGTGCTGGA
BdYAB6_Ndel	TAACGACATATGATGTCGTCGGCAGCCCAG ATCG
BdYAB6_EcoRI	GAATTCCATATGCTATTCGACGGCGAGCTTCTT
BdFIL1_Ndel	TAACGACATATGATGATGTCACCCGCGGAGTC
BdFIL1_EcoRI	GAATTCCATATGCTAAAATGGAGTGACGCCCA
BdYAB1 Ndel	TAACGACATATGATGTCGTCGCCGGTCCAGTTCA
BdYAB1 EcoR	GAATTCCATATGTTAATTGCCGCCACCACCACCC
Arabidopsis complementation	
Genotyping	
SALK_001887_LP	TGTTGCCATCCAGCTTTAATC
SALK_001887_RP	ACAAGGAAAGATTGTGCATCG
SALK_041504_LP	GTGTAGGTGAATGTCCCATGC
SALK_041504_RP	GTACGCAGAAGGTACTCGCTG
SALK_047502_LP	AAGCTCTGCCTTTCCTTTGTC
SALK_047502_RP	GAGTGAAAGCAAGGAGAAGGG
SALK_063818_LP	AACTTGCAAACCAAACCACAC
SALK_063818_RP	GGAGATGGAGAGGAATGAAGG
SALK_003109_LP	TTTCTCACAGTGTGCTGAAGC
SALK_003109_RP	TGTTGAGGGATAGCAAATTGG
SALK_086776_LP	CCCTAGATCGTAGATCCTAAGGC
SALK_086776_RP	CATCTACAAGAGATGCCTCGG

Cont to next page…

Cloning

BdDL_Kpn F	AGTTAG GGTACC ATGCAGCAGAGCAGCATGGAT
BdDL_BamHI R	AAGATA GGATCC CTAGCCGCTGCGCTCCATCTG
BdINO_Kpn F	AGAATG GGTACC ATGTCGTCGTCGCTCGTCCAC
BdINO_BamHI R	AAGATA GGATCC TCAGTTCCCCCTGTGCTGGAT
BdYAB1_Kpn F	AGTAAG GGTACC ATGTCGTCGCCGGTCCAGTT
BdYAB1_BamHI R	AAGATA GGATCC TTAATTGCCGCCACCACCAC
BdYAB2_Kpn F	AGGGTG GGTACC ATGTCGGCACAGATCGCGCC
BdYAB2_BamHI R	AAGATA GGATCC TCAGTAGAGACCTTGAACTT
BdYAB6_Kpn F	AGTAGG GGTACC ATGTCGTCGGCAGCCCAGAT
BdYAB6_BamHI R	AAGATA GGATCC CTATTCGACGGCGAGCTTCTT
BdFIL1_Kpn F	AGTGAG GGTACC ATGATGTCACCCGCGGAGTCGTC
BdFIL1_BamHI R	AAGATA GGATCC CTAAAATGGAGTGACGCCCATG
BdFIL2_Kpn F	AGTTAG GGTACC ATGTCTTCGTCGTCCTCCTCCT
BdFIL2_BamHI R	AAGATA GGATCC CTAAAATGGAGTGACGCCCATG
BdFIL3_Kpn F	AGTTAG GGTACC ATGTCGTCGTCCTCCTCCTC
BdFIL3_BamHI R	AAGATA GGATCC TCAATACGGAGCAACCCCCATGTT

Similar expression

Bradi5g09260_F	TTCAGTTGCTCGGTACTGGA
Bradi5g09260_R	TAGACTGGCACATTGTCATCAA
Bradi4g03970_F	TTAGTGGAAGTGGAACGCCAC
Bradi4g03970_R	GCAAGTTTAAATTATGATG
Bradi1g09670_F	TCAACCTGGCACTGGGCACTGGCA
Bradi1g09670_R	CGTAAATCCTCACTGGATT
Bradi1g46450_F	CAGGATCTGCAGAAGTGGTA
Bradi1g46450_R	TGTACAAAACACAATGGCATT
Bradi5g02447_F	ATAAGAGCCACATGGGAATA
Bradi5g02447_R	GTGAAAACATGCAGACTTTTGTA
Bradi3g52320_F	TCGATCGGTGGAGATCAATCGATT
Bradi3g52320_R	ATGCATCAAGCATGGATGGTCCAT

Appendix 4.3 – Yeast 2 Hybrid components and result interpretations

Table A.3: List of components needed in Y2H experiments

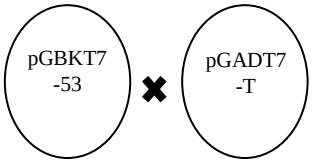
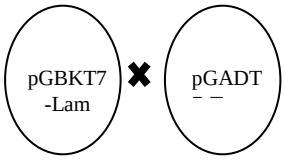
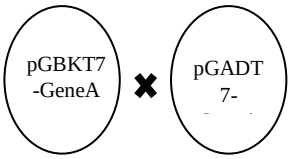
Matchmaker Vectors
pGBKT7 DNA-BD Cloning Vector (0.1 µg/µl)
pGADT7 AD Cloning Vector (0.1 µg/ µl)
pGBKT7-53 Control Vector (0.1 µg/ µl)
pGADKT7-T Control Vector (0.1 µg/ µl)
pGBDKT7-Lam Control Vector (0.1 µg/ µl)
Matchmaker Yeast Strains
Y2HGold Yeast Strains
Y187 Yeast Strain
Yeastmaker Yeast Transformation System
Quick & Easy Yeast Transformation Mix
Denatured Yeastmaker carrier DNA
Transforming DNA

Table A.4: Yeast Growth Media & Supplements

	General Media Preparation Instructions • The media has been prepared by dissolving pouch contents (supplied by Clontech) in 500ml ddH_2O, autoclaved for 15 minute at 121 °C, and allowed to cool prior usage. • X-α-Gal (Cat. No 630463) been dissolved at 20mg/ml in dimethylformamide (DMF) and kept in -20 °C until used.
YPDA Agar Plates	-Follow the general media preparations
SDO Agar Plates	**Single Dropout SD/ (-Trp) or (-Leu)** -Follow the general media preparations
SDO/X Agar Plates	**Single Dropout SD/ (-Trp) or (-Leu) containing 40µg/ml X-α-Gal** -Follow the general media preparations -1ml of X-α-Gal (20mg/ml working concentrations) been added into 500ml autoclaved, cooled media before use
DDO Agar Plates	**Double Dropout SD/-Leu/-trp** -Follow the general media preparations
DDO/X Agar Plates	**Double Dropout SD/-Leu/-trp containing 40µg/ml X-α-Gal** -Follow the general media preparations -1ml of X-α-Gal(20mg/ml working concentrations) been added into 500ml autoclaved, cooled media before use
QDO Agar Plates	**Quadraple Dropout SD/-Ade/-His/-leu/-Trp** - Follow the general media preparations
QDO/X Agar Plates	**Quadraple Dropout SD/-Ade/-His/-leu/-Trp containing 40µg/ml X-α-Gal**

	-1ml of X-α-Gal (20mg/ml working concentrations) been added into 500ml autoclaved, cooled media before use.

Table A.5: Yeast mating and plates screening results

Mating	Plated medium		
	SD/-Trp/-Leu	SD/-Trp/-Leu with X-α-Gal	SD/-Ade/-His/-leu/-Trp with X-α-Gal
<u>Positive control</u>  	**Expected result** Colonies present = Diploid yeast	**Expected result** Blue = Positive control is working	**Expected result** Blue = Positively interact
<u>Negative control</u>  	**Expected result** Colonies present = Diploid yeast	**Expected result** White= Negative control is working	**Expected result** No colonies= No interaction
<u>Gene of Interest</u>  	**Expected result** Colonies present = Diploid yeast No colonies = The two strains didn't mate	**Expected result** Blue = Positive interaction indication White= Negative interaction indication	**Expected result** Blue = Positively interact White= No interaction

References

ABEBE, T., MELMAIEE, K., BERG, V. & WISE, R. P. 2010. Drought response in the spikes of barley: gene expression in the lemma, palea, awn, and seed. *Funct Integr Genomics,* 10**,** 191-205.

ABEBE, T., SKADSEN, R. W. & KEAPPLER, H. 2004. Cloning and identification of highly expressed genes in Barley palea and lemma. *Crop Science,* 44**,** 9.

ABRAHAM, Z., IGLESIAS-FERNÁNDEZ, R., MARTÍNEZ, M., RUBIO-SOMOZA, I., DÍAZ, I., CARBONERO, P. & VICENTE-CARBAJOSA, J. 2016. A Developmental Switch of Gene Expression in the Barley Seed Mediated by HvVP1 (Viviparous-1) and HvGAMYB Interactions. *Plant Physiology,* 170**,** 2146-2158.

ADAMSKI, N. M., ANASTASIOU, E., ERIKSSON, S., O'NEILL, C. M., & LENHARD, M.2009. Local maternal control of seed size by *KLUH*/*CYP78A5*-dependent growth signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(47), 20115–20120.

AGALOU, A., PURWANTOMO, S., OVERNAS, E., JOHANNESSON, H., ZHU, X., ESTIATI, A., DE KAM, R. J., ENGSTROM, P., SLAMET-LOEDIN, I. H., ZHU, Z., WANG, M., XIONG, L., MEIJER, A. H. & OUWERKERK, P. B. 2008. A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. *Plant Mol Biol,* 66**,** 87-103.

AIROLDI, C. A. & DAVIES, B. 2012. Gene duplication and the evolution of plant MADS-box transcription factors. *J Genet Genomics,* 39**,** 157-65.

ALBANI, D., HAMMOND-KOSACK, M., SMITH, C., CONLAN, S., COLOT, V., HOLDSWORTH, M. & BEVAN, M. W. 1997. The Wheat Transcriptional Activator SPA: A Seed-Specific bZIP Protein That Recognizes the GCN4-like Motif in the Bifactorial Endosperm Box of Prolamin Genes. *The Plant Cell Online,* 9**,** 171-184.

ALLOCO, D. J., KOHANE, I. S. & BUTTE, A. J. 2004. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics,* 5**,** 10.

ALVES, S. C., WORLAND, B., THOLE, V., SNAPE, J. W., BEVAN, M. W. & VAIN, P. 2009. A protocol for Agrobacterium-mediated transformation of *Brachypodium distachyon* community standard line Bd21. *Nat Protoc,* 4**,** 638-49.

AUGEN, J. 2004. *Bioinformatics in the post-genomic era: Genome, transcriptome, proteome, and information-based medicine,* Addison-Wesley Professional.

BADIS, G., BERGER, M. F., PHILIPPAKIS, A. A., TALUKDER, S., GEHRKE, A. R., JAEGER, S. A., CHAN, E. T., METZLER, G., VEDENKO, A., CHEN, X., KUZNETSOV, H., WANG, C. F., COBURN, D., NEWBURGER, D. E., MORRIS, Q., HUGHES, T. R. & BULYK, M. L. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science,* 324**,** 1720-3.

BAILEY, T. L., WILLIAMS, N., MISLEH, C. & LI, W. W. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res,* 34**,** W369-73.

BARTHOLMES, C., HIDALGO, O. & GLEISSBERG, S. 2012. Evolution of the YABBY gene family with emphasis on the basal eudicot Eschscholzia californica (Papaveraceae). *Plant Biol (Stuttg),* 14**,** 11-23.

BASF 2006. Brome Guide.

BATES, J. C. 1943. varietal Differences in Anatomy of Cross-Section of Wheat Grain. *Chicago Journals,* 104**,** 5.

BAUER, E., SCHMUTZER, T., BARILAR, I., MASCHER, M., GUNDLACH, H., MARTIS, M. M., TWARDZIOK, S. O., HACKAUF, B., GORDILLO, A., WILDE, P., SCHMIDT, M., KORZUN, V., MAYER, K. F., SCHMID, K., SCHON, C. C. & SCHOLZ, U. 2017. Towards a whole-genome sequence for rye (Secale cereale L.). *Plant J,* 89**,** 853-869.

BAUM, D. A., YOON, H. S. & OLDHAM, R. L. 2005. Molecular evolution of the transcription factor LEAFY in Brassicaceae. *Mol Phylogenet Evol,* 37**,** 1-14.

BECRAFT, P. W. 2001. Cell fate specification in the cereal endosperm. *Seminars in Cell & Developmental Biology,* 12**,** 387-394.

BECHTEL, D. B. & WILSON, J. D. 2003. Amyloplast formation and starch granule development in hard red winter wheat. *Cereal Chemistry,* 80**,** 175.

BENFEY, P. N. & WEIGEL, D. 2001. Transcriptional Networks Controlling Plant Development. *Plant Physiology,* 125**,** 3.

BERGER, F., HAMAMURA, Y., INGOUFF, M. & HIGASHIYAMA, T. 2008. Double fertilization - caught in the act. *Trends Plant Sci,* 13**,** 437-43.

BEVAN, M. W., GARVIN, D. F. & VOGEL, J. P. 2010. *Brachypodium distachyon* genomics for sustainable food and fuel production. *Curr Opin Biotechnol,* 21**,** 211-7.

BOMMERT, P., SATOH-NAGASAWA, N., JACKSON, D. & HIRANO, H. Y. 2005. Genetics and evolution of inflorescence and flower development in grasses. *Plant Cell Physiol,* 46**,** 69-78.

BONACCORSO, O., LEE, J. E., PUAH, L., SCUTT, C. P. & GOLZ, J. F. 2012. FILAMENTOUS FLOWER Control Lateral Organ Development by Acting as Both an Activator and a Repressor. *BMC Plant Biology,* 12**,** 16.

BONACCORSO O., N. E. L., LIBBY PUAH, CHARLES P SCUTT, JOHN F GOLZ 2012. FILAMENTOUS FLOWER controls lateral organ development by acting as both an activator and a repressor. *BMC Plant Biology,* 12**,** 16.

BOWMAN, J. & SMYTH, D. R. 1999. CRABS CLAW, a gene that regulates carpel and nectary development in *Arabidopsis,* encodes a novel protein with zinc finger and helix-loop-helix domains. *Development,* 126**,** 10.

BOWMAN, J. L. 2000. The YABBY gene family and abaxial cell fate. *Curr Opin Plant Biol,* 3**,** 17-22.

BOWMAN, J. L., ESHED, Y. & BAUM, S. F. 2002. Establishment of polarity in angiosperm lateral organs. *Trends in Genetics,* 18**,** 8.

BRILLOUET, J. M. 1987. Investigation of the structure of a heteroxylan from the outer pericarp (Beeswing Bran) of wheat kernel. *Carbohydrate Research***,** 18.

BRKLJACIC, J., GROTEWOLD, E., SCHOLL, R., MOCKLER, T., GARVIN, D. F., VAIN, P., BRUTNELL, T., SIBOUT, R., BEVAN, M., BUDAK, H., CAICEDO, A. L., GAO, C., GU, Y., HAZEN, S. P., HOLT, B. F., 3RD, HONG, S. Y., JORDAN, M., MANZANEDA, A. J., MITCHELL-OLDS, T., MOCHIDA, K., MUR, L. A., PARK, C. M., SEDBROOK, J., WATT, M., ZHENG, S. J. & VOGEL, J. P. 2011. *Brachypodium* as a model for the grasses: today and the future. *Plant Physiol,* 157**,** 3-13.

BRUGGMANN, R., BHARTI, A., GUNDLACH, H., LAI , J., YOUNG , S., PONTAROLI, A., WEI, F., HABERER, G., FUKS , G. & DU , C. 2006. Uneven Chromosome contraction and Expansion in the Maize Genome. *Genome Research***,** 16:1241-1251.

BURTON, R. A. & FINCHER, G. B. 2014. Evolution and development of cell walls in cereal grains. *Front Plant Sci,* 5**,** 456.

CALEY, C. Y., DUFFUS, C. M. & JEFFCOAT, B. 1990. Photosynthesis in the pericarp of developing wheat grains. *Journal of Experimental Botany***,** 5.

CATALAN, P., CHALHOUB, B., CHOCHOIS, V., GARVIN, D. F., HASTEROK, R., MANZANEDA, A. J., MUR, L. A., PECCHIONI, N., RASMUSSEN, S. K., VOGEL, J. P. & VOXEUR, A. 2014. Update on the genomics and basic biology of *Brachypodium*: International *Brachypodium* Initiative (IBI). *Trends Plant Sci,* 19**,** 414-8.

CATALAN, P., KELLOGG, E. A. & OLMSTEAD, R. G. 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequences. *Mol Phylogenet Evol,* 8**,** 150-66.

CATALAN, P., MULLER, J., HASTEROK, R., JENKINS, G., MUR, L. A., LANGDON, T., BETEKHTIN, A., SIWINSKA, D., PIMENTEL, M. & LOPEZ-ALVAREZ, D. 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Ann Bot,* 109**,** 385-405.

CHEBLI, Y. & GEITMANN, A. 2017. Cellular growth in plants requires regulation of cell wall biochemistry. *Curr Opin Cell Biol,* 44**,** 28-35.

CHEN, Q., ATKINSON, A., OTSUGA, D., CHRSTENSEN, T., REYNOLDS, L. & DREWS, N. 1999. The *Arabidopsis FILAMENTOUS FLOWER* gene is required for flower formation. *Development,* 126**,** 12.

CHEN, X., LONG, H., GAO, P., DENG, G., PAN, Z., LIANG, J., TANG, Y., TASHI, N. & YU, M. 2014. Transcriptome assembly and analysis of Tibetan Hulless Barley (Hordeum vulgare L. var. nudum) developing grains, with emphasis on quality properties. *PLoS One,* 9**,** e98144.

CHRISTENSEN, U., ALONSO-SIMON, A., SCHELLER, H. V., WILLATS, W. G. & HARHOLT, J. 2010. Characterization of the primary cell walls of seedlings of *Brachypodium distachyon*--a potential model plant for temperate grasses. *Phytochemistry,* 71**,** 62-9.

CIFTCI-YILMAZ, S. & MITTLER, R. 2008. The zinc finger network of plants. *Cell Mol Life Sci,* 65**,** 1150-60.

CLOUGH, S. J. & BEMT, A. F. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal***,** 9.

COEN, E. S. & MEYEROWITZ, E. M. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature,* 353**,** 31-7.

CONG, B., BARRERO, L. S. & TANKSLEY, S. D. 2008. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat Genet,* 40**,** 800-4.

COPE, T. & GRAY, A. 2009. Grasses of the British Isles. Botanical Society of the British Isles Handbook No. 13. Botanical Society of the British Isles, London.

DE FOLTER, S., IMMINK, R. G., KIEFFER, M., PARENICOVA, L., HENZ, S. R., WEIGEL, D., BUSSCHER, M., KOOIKER, M., COLOMBO, L., KATER, M. M., DAVIES, B. & ANGENENT, G. C. 2005. Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *Plant Cell,* 17**,** 1424-33.

DEFRANCE, M., JANKY, R., SAND, O. & VAN HELDEN, J. 2008. Using RSAT oligo-analysis and dyad-analysis tools to discover regulatory signals in nucleic sequences. *Nat Protoc,* 3**,** 1589-603.

DRAPER, J., MUR, L. A. J., JENKINS, G., GHOSH-BISWAS, G. C., BABLAK, P., HASTEROK, R. & ROUTLEDGE, A. P. M. 2001a. *Brachypodium distachyon*.

A New Model System for Functional Genomics in Grasses. *Plant Physiology,* 127**,** 1539-1555.

DREA, S., LEADER, D. J., ARNOLD, B. C., SHAW, P., DOLAN, L. & DOONAN, J. H. 2005. Systematic spatial analysis of gene expression during wheat caryopsis development. *Plant Cell,* 17**,** 2172-85.

DRUKA, A., MUEHLBAUER, G., DRUKA, I., CALDO, R., BAUMANN, U., ROSTOKS, N., SCHREIBER, A., WISE, R., CLOSE, T., KLEINHOFS, A., GRANER, A., SCHULMAN, A., LANGRIDGE, P., SATO, K., HAYES, P., MCNICOL, J., MARSHALL, D. & WAUGH, R. 2006. An atlas of gene expression from seed to seed through barley development. *Funct Integr Genomics,* 6**,** 202-11.

DUMAS, C. & ROGOWSKY, P. 2008. Fertilization and early seed formation. *C R Biol,* 331**,** 715-25.

ECKARDT, N. A. 2010. YABBY genes and the development and origin of seed plant leaves. *Plant Cell,* 22**,** 2103.

EDGER, P. P. & PIRES, J. C. 2009. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res,* 17**,** 699-717.

EISEN, M. B., SPELLMAN, P. T., BROWN, P. O. & BOTSTEIN, D. 1999. Cluster analysis and display of genome-wide expression patterns. *PNAS,* 95**,** 8.

ENGLBRECHT, C. C., SCHOOF, H. & BOHM, S. 2004. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genomics,* 5**,** 39.

ESHED, Y., BAUM, S. F. & BOWMAN, J. 1999. Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *cell,* 99**,** 11.

EVERS, T. & MILLAR, S. 2002. Cereal Grain Structure and Development: Some Implications for Quality. *Journal of Cereal Science,* 36**,** 261-284.

FAUTEUX, F. & STROMVIK, M. V. 2009. Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. *BMC Plant Biol,* 9**,** 126.

FEBRER, M., GOICOECHEA, J. L., WRIGHT, J., MCKENZIE, N., SONG, X., LIN, J., COLLURA, K., WISSOTSKI, M., YU, Y., AMMIRAJU, J. S., WOLNY, E., IDZIAK, D., BETEKHTIN, A., KUDRNA, D., HASTEROK, R., WING, R. A. & BEVAN, M. W. 2010. An integrated physical, genetic and cytogenetic map of *Brachypodium distachyon*, a model system for grass research. *PLoS One,* 5**,** e13461.

FENG, L., ZHU, J., WANG, G., TANG, Y., CHEN, H., JIN, W., WANG, F., MEI, B., XU, Z. & SONG, R. 2009. Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize alpha-zein super gene family. *Plant Mol Biol,* 69**,** 649-59.

FINET, C., BERNE-DEDIEU, A., SCUTT, C. P. & MARLETAZ, F. 2013. Evolution of the ARF gene family in land plants: old domains, new tricks. *Mol Biol Evol,* 30**,** 45-56.

FINET, C., FLOYD, S. K., CONWAY, S. J., ZHONG, B., SCUTT, C. P. & BOWMAN, J. L. 2016. Evolution of the YABBY gene family in seed plants. *Evol Dev*.

FOURQUIN, C., VINAUGER-DOUARD, M., CHAMBRIER, P., BERNE-DEDIEU, A. & SCUTT, C. P. 2007. Functional conservation between CRABS CLAW orthologues from widely diverged angiosperms. *Ann Bot,* 100**,** 651-7.

FRANCO-ZORRILLA, J. M. & SOLANO, R. 2016. Identification of plant transcription factor target sequences. *Biochim Biophys Acta*.

FRANCIN-ALLAMI, M., LOLLIER, V., PAVLOVIC, M., SAN CLEMENTE, H., ROGNIAUX, H., JAMET, E., GUILLON, F. & LARRÉ, C. 2016. Understanding the Remodelling of Cell Walls during Brachypodium distachyon Grain Development through a Sub-Cellular Quantitative Proteomic Approach. *Proteomes,* 4**,** 21.

FREER, B. 2012. Beating Brome. *Farmers Weekly,* 21 September 2012.

FREEMAN, P. L. & PALMER, G. H. 1984. The structure of pericarp and testa of Barley. *J. Inst. Brew.,* 90**,** 7.

FUJIWARA, S., ODA, A., YOSHIDA, R., NIINUMA, K., MIYATA, K., TOMOZOE, Y., TAJIMA, T., NAKAGAWA, M., HAYASHI, K., COUPLAND, G. & MIZOGUCHI, T. 2008. Circadian Clock Proteins LHY and CCA1 Regulate SVP Protein Accumulation to Control Flowering in Arabidopsis. *Plant Cell,* 20**,** 2960-2971.

GARVIN, D. F. 2007. *Brachypodium*: a new monocot model plant system emerges. *Journal of the Science of Food and Agriculture,* 87**,** 1177-1179.

GARVIN, D. F., GU, Y.-Q., HASTEROK, R., HAZEN, S. P., JENKINS, G., MOCKLER, T. C., MUR, L. A. J. & VOGEL, J. P. 2008. Development of Genetic and Genomic Research Resources for , a New Model System for Grass Crop Research. *Crop Science,* 48**,** S-69.

GIFFORD, E. M. & FOSTER, A. S. 1989. *Morphology and evolution of vascular plants,* New York, W.H. Freeman and Company.

GIRIN, T., DAVID, L. C., CHARDIN, C., SIBOUT, R., KRAPP, A., FERRARIO-MERY, S. & DANIEL-VEDELE, F. 2014. *Brachypodium*: a promising hub between model species and cereals. *J Exp Bot,* 65**,** 5683-96.

GOMEZ, E., ROYO, J., GUO, Y., THOMPSON, R. & HUEROS, G. 2002. Establishment of cereal endosperm expression domains: identification and properties of a maize transfer cell-specific transcription factor, ZmMRP-1. *Plant Cell,* 14**,** 599-610.

GOMEZ, E., ROYO, J., MUNIZ, L. M., SELLAM, O., PAUL, W., GERENTES, D., BARRERO, C., LOPEZ, M., PEREZ, P. & HUEROS, G. 2009. The maize transcription factor myb-related protein-1 is a key regulator of the differentiation of transfer cells. *Plant Cell,* 21**,** 2022-35.

GONZALEZ-CALLE, V., IGLESIAS-FERNANDEZ, R., CARBONERO, P. & BARRERO-SICILIA, C. 2014. The BdGAMYB protein from Brachypodium distachyon interacts with BdDOF24 and regulates transcription of the BdCathB gene upon seed germination. *Planta,* 240**,** 539-52.

GOTO, K., KYOZUKA, J. & BOWMAN, J. 2001. Truning floral organs into leaves, leaves into floral organs. *Current Opinion in Genetics & Development,* 11**,** 8.

GRAEBER, K., LINKIES, A., MULLER, K., WUNCHOVA, A., ROTT, A. & LEUBNER-METZGER, G. 2010. Cross-species approaches to seed dormancy and germination: conservation and biodiversity of ABA-regulated mechanisms and the Brassicaceae DOG1 genes. *Plant Mol Biol,* 73**,** 67-87.

GU, Q., FERRÁNDIZ, C., YANOFSKY, M. F. & MARTIENSSEN, R. A. 1998. The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development,* 125**,** 9.

GUBATZ, S., DERCKSEN, V. J., BRUSS, C., WESCHKE, W. & WOBUS, U. 2007. Analysis of barley (Hordeum vulgare) grain development using three-dimensional digital models. *Plant J,* 52**,** 779-90.

GUBLER, F., MILLAR, A. A. & JACOBSEN, J. V. 2005. Dormancy release, ABA and pre-harvest sprouting. *Curr Opin Plant Biol,* 8**,** 183-7.

GUILLON, F., LARRE, C., PETIPAS, F., BERGER, A., MOUSSAWI, J., ROGNIAUX, H., SANTONI, A., SAULNIER, L., JAMME, F., MIQUEL, M., LEPINIEC, L. & DUBREUCQ, B. 2012. A comprehensive overview of grain development in *Brachypodium distachyon* variety Bd21. *J Exp Bot,* 63**,** 739-55.

GUPTA, S. K., RAI, A. K., KANWAR, S. S. & SHARMA, T. R. 2012. Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS One,* 7**,** e42578.

HALL, B. G. 2011. *Phylogenetic Trees Made Easy,* Sunderland, Massachusetts U.S.A, Sinauer Associates, Inc. Publishers.

HAMANT, O. & PAUTOT, V. 2010. Plant development: a TALE story. *C R Biol,* 333**,** 371-81.

HANDS, P. 2012. Comparative Grain Development in Temperate Grasses. PhD Thesis, University of Leicester.

HANDS, P. & DREA, S. 2012. A comparative view of grain development in *Brachypodium distachyon*. *Journal of Cereal Science,* 56**,** 2-8.

HANDS, P., KOURMPETLI, S., SHARPLES, D., HARRIS, R. G. & DREA, S. 2012. Analysis of grain characters in temperate grasses reveals sistinctive pattern of endosperm organization associated with grain shape. *Journal of Experimental Botany***,** 14.

HARBISON, C. T., GORDON, D. B., LEE, T. I., RINALDI, N. J., MACISAAC, K. D., DANFORD, T. W., HANNETT, N. M., TAGNE, J. B., REYNOLDS, D. B., YOO, J., JENNINGS, E. G., ZEITLINGER, J., POKHOLOK, D. K., KELLIS, M., ROLFE, P. A., TAKUSAGAWA, K. T., LANDER, E. S., GIFFORD, D. K., FRAENKEL, E. & YOUNG, R. A. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature,* 431**,** 99-104.

HAY, A. & TSIANTIS, M. 2010. KNOX genes: versatile regulators of plant development and diversity. *Development,* 137**,** 3153-3165.

HGCA 2014. Identification and control of brome grasses. *Information Sheet,* 31.

HUBBARD, C. E. 1954. *Grasses,* Pelican Books.

HUANG R., JIANG L., ZHENG J., WANG T., WANG H., HUANG Y. 2013. Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci.* 18 218–226 10.1016/j.tplants.2012.11.001

HUO, N., VOGEL, J. P., LAZO, G. R., YOU, F. M., MA, Y., MCMAHON, S., DVORAK, J., ANDERSON, O. D., LUO, M. C. & GU, Y. Q. 2009. Structural characterization of *Brachypodium* genome and its syntenic relationship with rice and wheat. *Plant Mol Biol,* 70**,** 47-61.

ISHIKAWA, M., OHMORI, Y., TANAKA, W., HIRABAYASHI, C., MURAI, K., OGIHARA, Y., YAMAGUCHI, T. & HIRANO, H. Y. 2009. The spatial expression patterns of DROOPING LEAF orthologs suggest a conserved function in grasses. *Genes Genetic Syst,* 84**,** 10.

ITOH, J., SATO, Y., SATO, Y., HIBARA, K., SHIMIZU-SATO, S., KOBAYASHI, H., TAKEHISA, H., SANGUINET, K. A., NAMIKI, N. & NAGAMURA, Y. 2016. Genome-wide analysis of spatiotemporal gene expression patterns during early embryogenesis in rice. *Development,* 143**,** 1217-27.

IZAWA Y., TAKAYANAGI Y., INABA N., ABE Y., MINAMI M., FUJISAWA Y. 2010. Function and expression pattern of the alpha subunit of the heterotrimeric G protein in rice. *Plant Cell Physiol.* 51 271–281 10.1093/pcp/pcp186

JIN, J., ZHANG, H., KONG, L., GAO, G. & LUO, J. 2014. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res,* 42**,** D1182-7.

JUAREZ, M. T., TWIGG, R. W. & TIMMERMANS, M. C. 2004. Specification of adaxial cell fate during maize leaf development. *Development,* 131**,** 4533-44.

JUNG, K., AN, G. & RONALD, P. 2008. Towards A Better Bowl of Rice: Assigning Function to Ten of Thousands of Rice Genes. *Nature Reviews Genetics,* 9**,** 91-101.

K.B, S., FOLEY, R. C. & SANCHEZ, L. O. 2002. Transcription factors in plant defense and stress response. *Current Opinion in Plant Biology,* 5**,** 7.

KAUFMANN, K., PAJORO, A. & ANGENENT, G. C. 2010. Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat Rev Genet,* 11**,** 830-42.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-9.

KELLOGG, E. A. 1998. Relationships of cereal crops and other grasses. *Proc Natl Acad Sci U S A,* 95**,** 2005-10.

KELLOGG, E. A. 2001. Evolutionary history of the grasses. *Plant Physiology,* 125**,** 8.

KELLOGG, E. A. & BUELL, C. R. 2009. Splendor in the grasses. *Plant Physiol,* 149**,** 1-3.

KIM, E. D., XIONG, Y., PYO, Y., KIM, D. H., KANG, B. H. & SUNG, S. 2017. Spatio-temporal analysis of coding and long noncoding transcripts during maize endosperm development. *Sci Rep,* 7**,** 3838.

LARRE, C., PENNINCK, S., BOUCHET, B., LOLLIER, V., TRANQUET, O., DENERY-PAPINI, S., GUILLON, F. & ROGNIAUX, H. 2010. *Brachypodium distachyon* grain: identification and subcellular localization of storage proteins. *J Exp Bot,* 61**,** 1771-83.

LAUDENCIA-CHINGCUANCO, D. L. & VENSEL, W. H. 2008. Globulins are the main seed storage proteins in *Brachypodium distachyon*. *Theor Appl Genet,* 117**,** 555-63.

LEE, J. Y., BAUM, S. F., ALVAREZ, J., PATEL, A., CHITWOOD, D. H. & BOWMAN, J. L. 2005. Activation of CRABS CLAW in the Nectaries and Carpels of *Arabidopsis*. *Plant Cell,* 17**,** 25-36.

LEPRINCE, O., PELLIZZARO, A., BERRIRI, S. & BUITINK, J. 2017. Late seed maturation: drying without dying. *J Exp Bot,* 68**,** 827-841.

LI, H., LIANG, W., YIN, C., ZHU, L. & ZHANG, D. 2011. Genetic interaction of OsMADS3, DROOPING LEAF, and OsMADS13 in specifying rice floral organ identities and meristem determinacy. *Plant Physiol,* 156**,** 263-74.

LI, L. F. & OLSEN, K. M. 2016. To Have and to Hold: Selection for Seed and Fruit Retention During Crop Domestication. *Curr Top Dev Biol,* 119**,** 63-109.

LICAUSI, F., OHME-TAKAGI, M. & PERATA, P. 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist,* 199**,** 639-649.

LIN, Z., LI, X., SHANNON, L. M., YEH, C. T., WANG, M. L., BAI, G., PENG, Z., LI, J., TRICK, H. N., CLEMENTE, T. E., DOEBLEY, J., SCHNABLE, P. S., TUINSTRA, M. R., TESSO, T. T., WHITE, F. & YU, J. 2012. Parallel domestication of the Shattering1 genes in cereals. *Nat Genet,* 44**,** 720-4.

LIU, H. L., XU, Y. Y., XU, Z. H. & CHONG, K. 2007. A rice YABBY gene, OsYABBY4, preferentially expresses in developing vascular tissue. *Dev Genes Evol,* 217**,** 629-37.

MATASCI, N., HUNG, L.-H., YAN, Z., CARPENTER, E. J., WICKETT, N. J., MIRARAB, S., NGUYEN, N., WARNOW, T., AYYAMPALAYAM, S. & BARKER, M. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience,* 3**,** 1-10.

MCABEE, J. M., KUZOFF, R. K. & GASSER, C. S. 2005. Mechanisms of derived unitegmy among Impatiens species. *Plant Cell,* 17**,** 1674-84.

MORRISON, L. N. 1976. The structure of the chlorophyll-containing cross cell and tube cells of the inner pericarp of wheat during grain development. *Botanical Gazette,* 137**,** 10.

MEISTER, R. J., KOTOW, L. M. & GASSER, C. S. 2002. SUPERMAN attenuates positive INNER NO OUTER autoregulation to maintain polar development of *Arabidopsis* ovule outer integumetns. *development,* 129**,** 9.

MOCHIDA, K. & SHINOZAKI, K. 2010. Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol,* 51**,** 497-523.

NAGASAWA, N. 2003. SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. *Development,* 130**,** 705-718.

NAKASE, M., AOKI, N., MATSUDA, T. & ADACHI, T. 1997. Characterization of a novel rice bZIP protein which binds to the α-globulin promoter. *Plant Molecular Biology,* 33**,** 513-522.

NAKAYAMA, H., YAMAGUCHI, T. & TSUKAYA, H. 2010. Expression patterns of AaDL, a CRABS CLAW ortholog in Asparagus asparagoides (Asparagaceae), demonstrate a stepwise evolution of CRC/DL subfamily of YABBY genes. *Am J Bot,* 97**,** 591-600.

NG, M. & YANOFSKY, M. F. 2001. Function and Evolution of the Plant MADS-BOX Gene Family. *Nature Reviews,* 2**,** 10.

NUIN, P. A., WANG, Z. & TILLIER, E. R. 2006. The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics,* 7**,** 471.

OHMORI, Y., TORIBA, T., NAKAMURA, H., ICHIKAWA, H. & HIRANO, H. Y. 2011. Temporal and spatial regulation of DROOPING LEAF gene expression that promotes midrib formation in rice. *Plant J,* 65**,** 77-86.

OLSEN, O. A. 2001. ENDOSPERM DEVELOPMENT: Cellularization and Cell Fate Specification. *Annu Rev Plant Physiol Plant Mol Biol,* 52**,** 233-267.

OLSEN, O. A., POTTER, R. H. & KALLA, R. 1992. Histo-differentiation and molecular biology of developing cereal endosperm. *Seed Science Research,* 2**,** 117-131.

OLSEN, P., LENK, I., JENSEN, C. S., PETERSEN, K., ANDERSEN, C. H., DIDION, T. & NIELSEN, K. K. 2006. Analysis of two heterologous flowering genes in *Brachypodium distachyon* demonstrates its potential as a grass model plant. *Plant Science,* 170**,** 1020-1025.

ONODERA, Y., SUZUKI, A., WU, C. Y., WASHIDA, H. & TAKAIWA, F. 2001. A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. *J Biol Chem,* 276**,** 14139-52.

ONATE, L., VICENTE-CARBAJOSA, J., LARA, P., DIAZ, I. & CARBONERO, P. 1999. Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. *J Biol Chem,* 274**,** 9175-82.

OPANOWICSZ, M., HANDS, P., BETTS, D., PARKER, M. L., TOOLE, G. A., CLARE MILLS, E. N., DOONAN, J. H. & DREA, S. 2011. Endosperm Development in *Brachypodium distachyon*. *Journal of Experimental Botany,* 62**,** 735-748.

OPANOWICZ, M., HANDS, P., BETTS, D., PARKER, M. L., TOOLE, G. A., MILLS, E. N., DOONAN, J. H. & DREA, S. 2011. Endosperm development in *Brachypodium distachyon*. *J Exp Bot,* 62**,** 735-48.

ORASHAKOVA, S., LANGE, M., LANGE, S., WEGE, S. & BECKER, A. 2009. The CRABS CLAW ortholog from California poppy (Eschscholzia californica, Papaveraceae), EcCRC, is involved in floral meristem termination, gynoecium differentiation and ovule initiation. *Plant J,* 58**,** 682-93.

OSTERGAARD, L. 2009. Don't 'leaf' now. The making of a fruit. *Curr Opin Plant Biol,* 12**,** 36-41.

PARKER, D., BECKMANN, M., ENOT, D. P., OVERY, D. P., RIOS, Z. C., GILBERT, M., TALBOT, N. & DRAPER, J. 2008. Rice blast infection of *Brachypodium distachyon* as a model system to study dynamic host/pathogen interactions. *Nat Protoc,* 3**,** 435-45.

PATTISON, R. J., CSUKASI, F., ZHENG, Y., FEI, Z., VAN DER KNAAP, E. & CATALA, C. 2015. Comprehensive Tissue-specific Transcriptome Analysis Reveals Distinct Regulatory Programs During Early Tomato Fruit Development. *Plant Physiol*.

PELLNY, T. K., LOVEGROVE, A., FREEMAN, J., TOSI, P., LOVE, C. G., KNOX, J. P., SHEWRY, P. R. & MITCHELL, R. A. 2012. Cell walls of developing wheat starchy endosperm: comparison of composition and RNA-Seq transcriptome. *Plant Physiol,* 158**,** 612-27.

PINOT, F. & BEISSON, F. 2011. Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *FEBS J,* 278**,** 195-205.

PINYOPICH, A., DITTA, G. S., SAVIDGE, B., LILHEGREN, S. J., BAUMANN, E., WISMAN, E. & YANOFSKY, M. F. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424**,** 4.

POGSON, B. J., GANGULY, D. & ALBRECHT-BORTH, V. 2015. Insights into chloroplast biogenesis and development. *Biochim Biophys Acta,* 1847**,** 1017-24.

PRASAD, K., PARAMESWARAN, S. & VIJAYRAGHAVAN, U. 2005. OsMADS1, a rice MADS-box factor, controls differentiation of specific cell types in the lemma and palea and is an early-acting regulator of inner floral organs. *Plant J,* 43**,** 915-28.

RADCHUK, V. & BORISJUK, L. 2014. Physical, metabolic and developmental functions of the seed coat. *Front Plant Sci,* 5**,** 510.

HONG, R.L., HAMAGUCHI, L., BUSCH, M.A., WEIGEL, D. 2003. Regulatory elements of the Floral Homeotic Gene AGAMOUS identified by phylogenetic footprinting and shadowing. The Plant Cell, 15, 1296-1309.

REINHARDT, D., FRENZ, M., MANDEL, T. & KUHLEMEIER, C. 2005. Microsurgical and laser ablation analysis of leaf positioning and dorsoventral patterning in tomato. *Development,* 132**,** 15-26.

RIECHMANN, J. L. 2002. Transcriptional regulation: a genomic overview. *Arabidopsis Book,* 1**,** e0085.

RIECHMANN, J. L., HEARD, J., MARTIN, G., REUBER, T. L., JIANG, C. Z., KEDDIE, J., ADAM, L., PINEDA, O., SAMAHA, R., RATCLIFFE, O. J., CREELMAN, R., PILGRIM, M., BROUN, P., ZHANG, J. Z., GHANDEHARI, D., SHERMAN, B. K. & YU, G. L. 2000. *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Arabidopsis: Research Articles,* 290**,** 6.

RITCHIE, S., SWANSON, S. J. & GILROY, S. 2007. Physiology of the aleurone layer and starchy endosperm during grain development and early seedling growth: new insights from cell and molecular biology. *Seed Science Research,* 10**,** 21.

RODRIGUEZ, G. R., MUNOS, S., ANDERSON, C., SIM, S. C., MICHEL, A., CAUSSE, M., GARDENER, B. B., FRANCIS, D. & VAN DER KNAAP, E. 2011. Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol,* 156**,** 275-85.

RUBIO-SOMOZA, I., MARTINEZ, M., ABRAHAM, Z., DIAZ, I. & CARBONERO, P. 2006. Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. *Plant J,* 47**,** 269-81.

SAARELA, J. M., PETERSON, P. M., KEANE, R. M., CAYOUETTE, J. & GRAHAM, S. W. 2007. Molecular phylogenetics of Bromus (Poaceae: Pooideae) based on chloroplast and nuclear DNA sequence data. *Aliso: A Journal of Systematic and Evolutionary Botany,* 23**,** 450-467.

SAWA, S., ITO, T., SHIMURA, Y. & OKADA, K. 1999. FILAMENTOUS FLOWER Controls the Formation and Development of *Arabidopsis* Inflorescences and Floral Meristem. *The Plant Cell* 11**,** 19.

SCHIEFELBEIN, J. 2003. Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. *Current Opinion in Plant Biology***,** 5.

SCHMIDT, R. J., BURR, F. A., AUKERMAN, M. J. & BURR, B. 1990. Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc Natl Acad Sci U S A,* 87**,** 46-50.

SCHMIDT, R. J. & AMBROSE, B. A. 1998. The blooming of grass flower development. *Current Opinion in Plant Biology,* 1**,** 8.

SCHOLZ, H. 2008. Some comments on the genus Bromus (Poaceae) and three new species. *Willdenowia - Annals of the Botanic Garden and Botanical Museum Berlin-Dahlem,* 38**,** 411-422.

SCHREIBER, A. W., SUTTON, T., CALDO, R. A., KALASHYAN, E., LOVELL, B., MAYO, G., MUEHLBAUER, G. J., DRUKA, A., WAUGH, R., WISE, R. P., LANGRIDGE, P. & BAUMANN, U. 2009. Comparative transcriptomics in the Triticeae. *BMC Genomics,* 10**,** 285.

SESSIONS, A. & YANOFSKY, M. F. 1999. Dorsoventral patterning in plants. *genes and Development,* 13**,** 5.

SHAN, H., ZAHN, L., GUINDON, S., WALL, P. K., KONG, H., MA, H., DEPAMPHILIS, C. W. & LEEBENS-MACK, J. 2009. Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol Biol Evol,* 26**,** 2229-44.

SHEWRY, P. R. & HALFORD, G. 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. *Journal of Experimental Botany,* 53**,** 12.

SIEGFRIED, K. L., ESHED, Y., BAUM, S. F., OTSUGA, D., DREWS, G. N. & BOWMAN, J. 1999. Members of YABBY gene family specify abaxial cell fate in *Arabidopsis*. *development,* 126**,** 12.

SINGH, K. B. 1998. Transcriptional regulation in Plants: The Importance of Combinatoral control. *Plant Physiology,* 118**,** 10.

SMITH, P. M. 1970. Taxonomy and nomenclature of the brome-grasses (Bromus L. s.l.). *Notes RBG Edinburgh,* 30**,** 361-375.

SREENIVASULU, N., BORISJUK, L., JUNKER, B. H., MOCK, H.-P., ROLLETSCHEK, H., SEIFFERT, U., WESCHKE, W. & WOBUS, U. 2010. Barley Grain Development. 281**,** 49-89.

SREENIVASULU, N., RADCHUK, V., STRICKERT, M., MIERSCH, O., WESCHKE, W. & WOBUS, U. 2006. Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA- regulated maturation in developing barley seeds. *Plant J,* 47**,** 310-27.

STACE, C. 2010. *New flora of the British Isles*, Cambridge University Press.

STAHLE, M. I., KUEHLICH, J., STARON, L., VON ARNIM, A. G. & GOLZ, J. F. 2009. YABBYs and the transcriptional corepressors LEUNIG and LEUNIG_HOMOLOG maintain leaf polarity and meristem activity in *Arabidopsis*. *Plant Cell,* 21**,** 3105-18.

STEINWAND, M. A., YOUNG, H. A., BRAGG, J. N., TOBIAS, C. M. & VOGEL, J. P. 2013. *Brachypodium* sylvaticum, a model for perennial grasses: transformation and inbred line development. *PLoS One,* 8**,** e75180.

SUBBURAJ, S., LUO, N., LU, X., LI, X., CAO, H., HU, Y., LI, J. & YAN, Y. 2016. Molecular characterization and evolutionary origins of farinin genes in Brachypodium distachyon L. *J Appl Genet,* 57**,** 287-303.

SU, J. C. 2000. Starch synthesis and grain filling in rice Starch synthesis and grain filling in rice. *Developments in crop science* 26, 107-124..

SUSSEX, I. M. 1955. Morphogenesis in Solanum tuberosum L.: Apical structure and developmental pattern of the juvenile shoot. *Phytomorphology* 5**,** 20.

TANACKOVIC, V., SVENSSON, J. T., JENSEN, S. L., BULEON, A. & BLENNOW, A. 2014. The deposition and characterization of starch in *Brachypodium distachyon*. *J Exp Bot,* 65**,** 5179-92.

TANAKA, W., TORIBA, T. & HIRANO, H. Y. 2017. Three TOB1-related YABBY genes are required to maintain proper function of the spikelet and branch meristems in rice. *New Phytol,* 215**,** 825-839.

TANAKA, W., TORIBA, T., OHMORI, Y., YOSHIDA, A., KAWAI, A., MAYAMA-TSUCHIDA, T., ICHIKAWA, H., MITSUDA, N., OHME-TAKAGI, M. & HIRANO, H. Y. 2012. The YABBY gene TONGARI-BOUSHI1 is involved in lateral organ development and maintenance of meristem organization in the rice spikelet. *Plant Cell,* 24**,** 80-95.

TIMMERMANS, M. C., JUAREZ, M. T. & PHELPS-DURR, T. L. 2004. A conserved microRNA signal specifies leaf polarity. *Cold Spring Harb Symp Quant Biol,* 69**,** 409-17.

TORIBA, T., HARADA, K., TAKAMURA, A., NAKAMURA, H., ICHIKAWA, H., SUZAKI, T. & HIRANO, H. Y. 2007. Molecular characterization the YABBY gene family in Oryza sativa and expression analysis of OsYABBY1. *Mol Genet Genomics,* 277**,** 457-68.

TRAFFORD, K., HALEUX, P., HENDERSON, M., PARKER, M., SHIRLEY, N. J., TUCKER, M. R., FINCHER, G. B. & BURTON, R. A. 2013. Grain development in *Brachypodium* and other grasses: possible interactions between cell expansion, starch deposition, and cell-wall synthesis. *J Exp Bot,* 64**,** 5033-47.

VICENTE-CARBAJOSA, J. & CARBONERO, P. 2004. Seed maturation: developing an intrusive phase to accomplish a quiescent state. *International Journal of Developmental Biology,* 49**,** 645-651.

VILLANUEVA, J. M., BROADHVEST, J., HAUSER, B. A., MEISTER, R. J., SCHNEITZ, K. & GASSER, C. S. 1999. INNER NO OUTER regulates abaxial-adaxial patterning in *Arabidopsis* ovules. *Genes and Development,* 13**,** 10.

VOGEL, J. 2008. Unique aspects of the grass cell wall. *Curr Opin Plant Biol,* 11**,** 301-7.
VOGEL, J. & BRAGG, J. 2009a. *Brachypodium distachyon*, a New Model for the Triticeae. *Genetics and Genomics of the triticeae.* Albany, USa: Springer Science+Business Media 427-449.
VOGEL, J. P., GU, Y., TWIGG, P., LAZO, G., LAUDENCIA-CHINGCUANCO, D., HAYDEN, D., DONZE, T., VIVIAN, L., STAMOVA, B. & COLEMAN-DERR, D. 2006. EST Sequencing and Phylogenetic Analysis of the model grass *Brachypodium Distachyon*. *Theory Applied Genetics,* 113**,** 186-195.
VOSNAKIS, N., MAIDEN, A., KOURMPETLI, S., HANDS, P., SHARPLES, D. & DREA, S. 2012. A FILAMENTOUS FLOWER orthologue plays a key role in leaf patterning in opium poppy. *Plant J,* 72**,** 662-73.
WAITES, R. & HUDSON, A. 1995. PHANTASTICA: a gene required for dorsoventrality of leaves in *Anthirrhinum majus*. *Development,* 121**,** 12.
WAITES, R., SELVADURAL, H. R. N., OLIVER, I. R. & HUDSON, A. 1998. The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in Antirrhinum. *Cell,* 93**,** 11.
WAN, Y., POOLE, R., HUTTLY, A., TOSCANO-UNDERWOOD, C., FEENEY, K., WELHAM, S., GOODING, M., MILLS, C., EDWARDS, K., SHEWRY, P. & MITCHELL, R. 2008. Transcriptome analysis of grain development in hexaploid wheat. *BMC Genomics,* 9**,** 121.
WANG, K., HAN, X., DONG , K., GAO, L., LI, H., MA, W., YAN, Y. & YE, X. 2011. Characterization of Seed Proteome in *Brachypodium distachyon*. *Journal of Cereal Science,* 52**,** 177-186.
WANG, W., HAO, Q., TIAN, F., LI, Q. & WANG, W. 2016. Cytokinin-Regulated Sucrose Metabolism in Stay-Green Wheat Phenotype. *PloS one,* 11**,** e0161351.
WANG, Q., REDDY, V. A., PANICKER, D., MAO, H. Z., KUMAR, N., RAJAN, C., VENKATESH, P. N., CHUA, N. H. & SAROJAM, R. 2016. Metabolic engineering of terpene biosynthesis in plants using a trichome-specific transcription factor MsYABBY5 from spearmint (Mentha spicata). *Plant Biotechnol J*.
WILSON, J. D., BECHTEL, D. B., TODD, T. C. & SEIB, P. A. 2006. Measurement of Wheat Starch Granule Size Distribution Using Image Analysis and Laser Diffraction Technology. *Cereal Chemistry Journal,* 83**,** 259-268.
WOBUS, U., SREENIVASULU, N., BORISJUK, L., ROLLETSCHEK, H., PANITZ, R., GUBATZ, S. & WESCHKE, W. 2005. Molecular Physiology and Genomics of Developing Barley Grains. *Research Developmental Plant Molecular Biology,* 2**,** 29.
XIONG, F., YU, X. R., ZHOU, L., WANG, F. & XIONG, A. S. 2013. Structural and physiological characterization during wheat pericarp development. *Plant Cell Rep,* 32**,** 1309-20.
XUE, L. J., ZHANG, J. J. & XUE, H. W. 2012. Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS One,* 7**,** e31081.
YAMADA, T., ITO, M. & KATO, M. 2003. Expression pattern of INNER NO OUTER homologue in Nymphaea (water lily family, Nymphaeaceae). *Dev Genes Evol,* 213**,** 510-3.
YAMADA, T., YOKOTA, S., HIRAYAMA, Y., IMAICHI, R., KATO, M. & GASSER, C. S. 2011. Ancestral expression patterns and evolutionary diversification of YABBY genes in angiosperms. *Plant J,* 67**,** 26-36.
YAMAGUCHI, T., NAGASAWA, N., KAWASAKI, S., MATSUOKA, M., NAGATO, Y. & HIRANO, H. Y. 2004. The YABBY gene DROOPING LEAF regulates

carpel specification and midrib development in Oryza sativa. *Plant Cell,* 16**,** 500-9.

YAMAGUCHI, T., NUKAZUKA, A. & TSUKAYA, H. 2012. Leaf adaxial-abaxial polarity specification and lamina outgrowth: evolution and development. *Plant Cell Physiol,* 53**,** 1180-94.

YAN, L., FU, D., LI, C., BLECHL , A., TRANQUILLI, G., BONAFEDE, M., A, S., VALARIK, M., YASUDA , S. & DUBCOVSKY, J. 2006. The Wheat and Barley Vernalization gene VRN3 is an Orthologue of FT. *Proc National Academy Science USA***,** 103:19581-19586.

YIN, L. L. & XUE, H. W. 2012. The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. *Plant Cell,* 24**,** 1049-65.

YU, S.-M., LEE, Y.-C., FANG, S.-C., CHAN, M.-T., HWA, S.-F. & LIU, L.-F. 1996. Sugars act as signal molecules and osmotica to regulate the expression of α-amylase genes and metabolic activities in germinating cereal grains. *Plant Molecular Biology,* 30**,** 1277-1289.

YU, X., ZHOU, L., ZHANG, J., YU, H., XIONG, F. & WANG, Z. 2015. Comparison of starch granule development and physicochemical properties of starches in wheat pericarp and endosperm. *J Sci Food Agric,* 95**,** 148-57.

ZAMBELLI, F., PESOLE, G. & PAVESI, G. 2013. Motif discovery and transcription factor binding sites before and after the next-generation sequencing era. *Brief Bioinform,* 14**,** 225-37.

ZHAO, W., SU, H. Y., SONG, J., ZHAO, X. Y. & ZHANG, X. S. 2006. Ectopic expression of TaYAB1, a member of YABBY gene family in wheat, causes the partial abaxialization of the adaxial epidermises of leaves and arrests the development of shoot apical meristem in *Arabidopsis*. *Plant Science,* 170**,** 364-371.

ZHOU, Z., WANG, L., LI, J., SONG, X. & YANG, C. 2009. Study on programmed cell death and dynamic changes of starch accumulation in pericarp cells of Triticum aestivum L. *Protoplasma,* 236**,** 49-58.