Identification and characterization of transcription factors regulating aspects of grain quality in wheat and related species

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

By

Amal Alyamani

Department of Genetics and Genome Biology

University of Leicester

June 2019

Identification and characterization of transcription factors regulating aspects of grain quality in wheat and related species

By

Amal Alyamani

Abstract

Grain quality is influenced by the coordination of carbohydrate, protein and lipid metabolism during the development of grain. Therefore, the understanding of the molecular mechanisms regulating seed storage protein genes (SSPs) and starch metabolism genes expression and accumulation should provide insights into the genetic control of the grain quality. The aim of this study was to compare between the known and novel orthologous TFs in both families (TaSPA/ TaBLZ1; O2 family) and (RSR1/ Q; euAP2 clade) were identified that the master regulators of grain quality and characterized in detail. Comparison across Triticeae species showed the differences and similarities between them such as changes in sequence, exon arrangement, and expression (temporal and spatial), which may reflect the divergence in the function of these genes and therefore the different quality profiles in the grains of the species examined. Also, microRNA172 is considered to play a vital role in regulating the targets of euAP2 genes during the development of both the spikelet initiation, floral and grain development. The expression level was compared between microRNA172 and their target, and the result revealed that microRNA172 was negatively regulated by their target. In addition, I aimed to explore the function of TaSPA (bZIP) by using RNAi. Protein interactions were also investigated between different TFs, and the result showed that TFs are a part of the network of proteins which were required for the control of the regulation of these genes.Together, this work provides a comparative approach in terms of the molecular and structural levels to determine how variation in TFs regulating grain storage metabolism could explain the variation in quality profile between species. This could be pivotal for floral and grain development, highlighting potential targets for genetic manipulation to improve future wheat yields.

Acknowledgements

First and foremost, I would like to thank God (Allah) for his never-ending grace, mercy, and provision during what ended up being one of the toughest time of my life.

It is a pleasure to thank the many people who have made this thesis possible. I wish to express my deepest gratitude to my supervisor **Dr. Sinead Drea** for her support, and for giving me the opportunity to undertake this fascinating project in her laboratory.

Additionally, I would like to thank **Dr. Sofia Kourmpetli** for her explanations and for providing advice. Additionally, my particular thanks go to the members of Drea's lab group for creating a supportive and friendly atmosphere in the lab. My appreciation also goes to our lovely technician **Neelam Dave** for Spiritual and physical support.

I would like also to thank other members of the Department of Genetics who has offered me all help I needed especially Dave's members (**Dr.Diter** and **Dr.Lewis**). Also, I'd like to thank **Dr.Inna Guterman** and **Ramesh**. Also, special thanks to one of my panel reviewer **(Dr.Nicola)** for her notes and advice which encouraged me to improve my work.

I am also grateful to each member of the Genetics Department, Leicester University. In addition, I am grateful for everyone in in Taif University for supporting me during my PhD journey.

It is a pleasure to extend my deepest thanks to many people who have walked alongside me during my journey directly or indirectly.

Therefore, I am heartily thankful to (**Yosra Al-Hakeem**) but I can't find the word to express or appreciate all the support, advice, strength that she provided me. Also, I was lucky to find another kind person (**Aisha Amer**). She was extremely generous for anything I need from her. Really thanks for every single information, advice and for friendship before and during my writing up, we shared the food, talk, and help.

Also, I'd like to thank (**Jaybal**) for his support and advice, this person really stands with me.

Last but not least, I would like to dedicate the thesis to my family members as a token of thanks and love to **my parents**, whom I could never thank enough, who raised me with love for education and supported me, and to my husband (**Mohammed Saif**) who day and night supported me during my journey. Those efforts and advice always encouraged me. I would not be where I am today, without his support. Also, a big thank to my lovely sons (**Abdulaziz and Omar**) who supported me and afforded my busyness and of being me away the whole days during my journey in the UK. I will never forget my lovely sisters who gave me the happiness and the best memorize in my life. Also, many thanks to my brothers and for everyone which really without their support it was impossible to achieve this goal.

Dedication

‘To most respectable people in my life for all their love and support, firstly

My Parents, who do all the sacrifices for me. Hence, I should take a

moment in thanking for what they have done to make me successful.

Secondly, my husband (**Mohammed Saif**),

for his endless support, and his understanding of spirit during our journey in

the UK, and my lovely sons

(**Abdulaziz & Omar**).

Also, my sisters and my brothers who inspired me

throughout my study for their unconditional love, and encouragement.’

Table of Contents

List of Figures

List of Tables

List of Abbreviations

aa	amino acid
ANOVA	analysis of variance
AD	activating domain
AP2	APETALA2/ethylene-responsive element binding protein family
BCIP	5-bromo-4-chloro3-indolyl phosphate
BD	binding domain (also termed DBD)
BLAST	basic local alignment search tool BP
Bp	base pair
BR	basic region
BLZ2	Basic Leucine Zipper Factor
bZIP	Basic zipper protein
BZO2H	Basic leucine zipper O2 homolog
β-gal	β-galactosidase
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
CaMV 35 S	Cauliflower mosaic virus (CaMV) 35 S Promoter
cDNA	complementary DNA
cPCR	cloning PCR
CDS	coding sequence
DNA	deoxyribonucleic aci
Dof	DNA binding with one finger
DAA	Day after an-thesis
DAP	Day after pollination
DDO	Double dropout –Leu/–Trp.
DBD	DNA binding domain
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylene diaminetetraacetic acid
IM	Inflorescence meristem
ISH	In-situ Hybridization
(EST)	expressed sequence tag
FM	floral meristem identity
FLuc	firefly luciferase acitivty
FLuc/RLuc	normalised dual luciferase activity
Fmol	fentamole
G	gram
GAL4	galactose-gene activating transcription factor
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescent protein
His	Histidine
HMW	High molecular weight
Kb	kilobase pair
LB media	Luria Bertani media
LFY	LEAFY
LMW	Low molecular weight
Mg	milligram
miR159	microRNA159
ml	millilitre
mM	millimolar
mRNA	Messenger RNA
Mya	Million years ago
NBT	Nitroblue Tetrazoluim
nptII	neomycin phosphotransferase II
O2	OPAQUE-2
SBEI	starch branching enzyme
S. cerevisiae	Saccharomyces cerevisiae
DBE	Starch debranching enzyme
SSI	soluble starch synthase I
SSIIa	soluble starch synthase II
SSIII	starch synthase III
P2H	protoplast two-hybrid
PCD	Programmed Cell Death
PCR	polymerase chain reaction

PBF	prolamin-box binding factor
Pol II	Polymerase II
QTL	quantitative trait loci
qRT-PCR	quantitative RT-PCR
RenLUC	Renilla luciferase protein
Rluc	Renilla luciferase activity
RNA	Ribonucleic acid
(RNAi)	RNA interference
RISBZ	Rice seed b-Zipper
RT	room temperature
RT-PCR	Reverse transcription-polymerase chain reaction
RSR1	Rice Starch Regulator1
SDO	Single dropout (–Trp) or (–Leu) media
SDO/X Agar	Single dropout (–Trp) media containing 40 µg/ml X-a-Gal
SDO/X/A	Single dropout (–Trp) media containing 40 µg/ml X-a-Gal
SPA	Storage protein activator
SAM	shoot apical meristem
TAD	transactivation domain
TF	transcription factor
Tris	Tris (hydroxymethyl) amino methane
TFs	Transcription factors
TFL1	TERMINAL FLOWER1
µg	microgram
µl	microliter
UTR	untranslated region
WT	wild type
Xgal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YPD	Yeast Extract-Peptone-Dextrose
YPDA	Yeast Extract-Peptone-Dextrose Agar
Y2H	Yeast-2-hybrid
QDO	SD/–Ade/–His/–Leu/–Trp

Nucleobases:

A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil

Amino acids

A/Ala	alanine
R/Arg	arginine
N / Asn	asparagine
D / Asp	aspartic acid
C / Cys	cysteine
Q / Gln	glutamine
E / Glu	glutamic acid
G / Gly	glycine
H / His	histidine
I / Ile	isoleucine
L / Leu	leucine
K / Lys	lysine
M / Met	methionine
F / Phe	phenylalanine
P / Pro	proline
S / Ser	serine
T / Thr	threonine
W / Trp	tryptophan
Y / Tyr	tyrosine
V / Val	valine
%	percentage

Chapter 1

Literature review

Chapter 1 Literature review

1.1 The Poaceae Family

The grass family, Poaceae has been studied in detail because of its economic importance. It has approximately 11,000 species, some are herbs, while others are trees such as bamboos, and species occupy habitats from warm and cold deserts to rainforests (Gibson, 2009). The Pooideae subfamily one of the subfamilies of the Poaceae which exhibits a major lineage of the Poaceae family including almost all of the world's major temperate zone cereal, biofuel and forage crop species **(Fig.1.1)** (Bremer, 2002; Shewry, 2009) on which much works have been done on the evolution and development of such characters as inflorescence architecture. These analyses span the cereal phylogeny and extend into uncultivated and wild relatives (Doebley *et al.,* 1997; Vollbrecht *et al.,* 2005; Preston and Kellogg, 2008). Although the close relationship between species such as wheat, barley and rye and is expected to exhibit a strong physiological and genetic similarity between these species (Huo *et al.,* 2009), they show several variations of physiological, morphological, and genetic diversity such as a genome size which reflects the number of repetitive DNA in the grass genomes (Kellogg, 1998; Sorrells *et al.,* 2003; Bruggmann *et al.,* 2006). In addition, one of the differences between the grasses genomes is chromosome number. On the phenotypic traits, the variation within the species in the architecture of inflorescences leads to distinctive morphological diversity.

Furthermore, grasses are considered as a source of food for humans who depend on grain crops such as wheat (*Triticum aestivum*), rice (*Oryza sativa*) and maize (*Zea mays)* which together account for over 70% of the total production. Also, the grasses contain other cereals including barley, sorghum, millets, oats, and rye. Most of these cereals belong "Core pooids" lineage includes the Poeae, Aveneae, Triticeae, and Bromeae tribes (Catalan *et al.,* 1997; Doring *et al.,* 2007) **(Fig.1.1).** This sub-group was first identified and named in a study by Davis and Soreng (1993), who found evidence to support this grouping based upon chloroplast

restriction site variation. It has been investigated that this group has seven large chromosomes (Catalan *et al.,* 1997).

It has been confirmed that most species which are outside this group show significant differences to core pooids such as the Brachypoideae, which is located between Pooideae and Oryzoideae. *Brachypodium sp.* diverged from the Pooideae approximately 35–40 Mya (Catalan *et al.,* 1997). The genome sequence analysis could provide more knowledge about the evolution of *B. distachyon* (Opanowicz *et al.,* 2008). However, *B. distachyon* has been studied in detail because it has one of the smallest genomes to be found in the entire grass family, and it shares many physical and phylogenetic features, that make it suitable as a temperate grass model, also an interesting point of comparison in that it is a wild undomesticated grass (Hands and Drea, 2012).

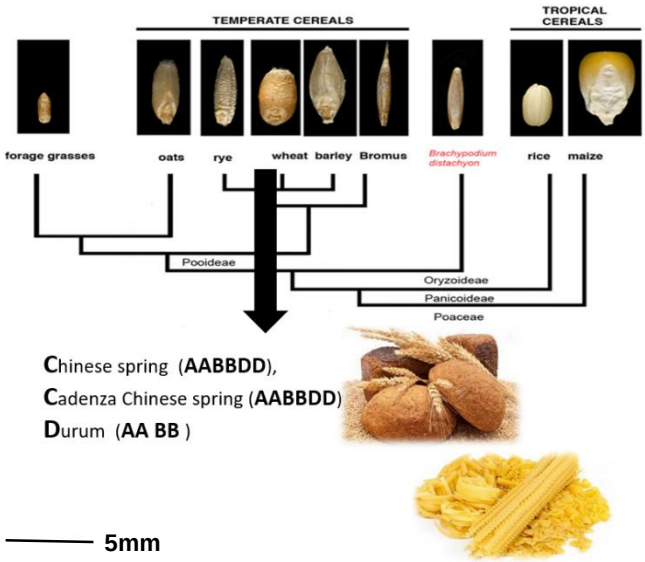


Figure 1.1 Phylogenetic relation between members of the Poaceae.

Brachypodium (red) which is located between the rice and temperate cereals. Bar 5mm; seed size. (Adapted from Opanowicz *et al.,* 2008).

1.1.1 **Wheat**

Bread wheat (*Triticum aestivum* L.) is one of the world's most important cereal grain crops and is the staple food source for 30% of the human population (International Wheat Genome Sequencing, 2014). Globally, wheat produces > 60% of the calories and protein for daily life (Gill *et al.*, 2004). It spreads worldwide as a global food because is its adaptability to a wide range of climatic conditions. There are several types of wheat such as diploid einkorn (14, AA), also 13 diploid species and 18 allopolyploid species (Morris and Sears1967; Van Slageren 1994). It was domesticated ~10,000 years ago, so that the modern wheat cultivars belong to two main species: hexaploid bread wheat, *Triticum aestivum* (42 chromosomes, AABBDD), and tetraploid, durum-type wheat, *T. turgidum* (28, AABB), that is known with the hard grain of all kinds of wheat it is due to the hard endosperm, so that traditionally, durum wheat has been thought to have the superior pasta making quality compared to that of bread wheat (Quaglia, 1988). Consequently, the texture of grain is one of the primary commercial classifications in world trade and can be

considered the single most important determinant of overall end-use quality of wheat.

The International Wheat Genome Sequencing Consortium (IWGSC), (2014) reported that around 17GB of hexaploid bread wheat (*Triticum aestivum*) genome has been produced by sequencing isolated chromosome arms, with 124,201 gene loci annotated that distributed nearly across the homologous chromosomes and sub-genomes. In addition, it was confirmed that the high sequence similarity and structural conservation are retained between diploid and tetraploid wheat relatives. The allohexaploid bread wheat genome contains three closely related sub-genomes are known as A genome, B genome, and D genome, but it is still unclear of their phylogenetic history (Marcussen *et al.*, 2014). The A and B genomes diverged from a common ancestor ~7 million years ago and that these genomes gave rise to the D genome through homoploid hybrid speciation 1 to 2 million years later. Hexaploid wheat genome arose as a result of two polyploidization events.

Hybridization and polyploidization have a significant impact on the evolution of species in both natures or under domestication. The wheat group (genera Aegilops and Triticum) are good examples of the effects of hybridization and polyploidization on the evolution of a species. The first diploidization event gave raise a tetraploid wheat when wild diploid wheat *T. urartu* (2n=2x= 14, AA genome) hybridized with a donor B, which is closely related to goat grass *Aegilops speltoidis* (2n=2x= 14, SS genome), 30,000-50,000 years ago. This was then domesticated around 10,000 years ago and became known as emmer wheat (*T.turgidum* L.2n=4x=28 AABB genome). A second diploidization event led to bread wheat took place between tetraploid emmer wheat and the wild diploid species *Aegilops tauschii* (2n= 2x= 14, DD genome). This gave rise to the fertile hexaploid wheat (AABBDD genome), is known as *T. aestivum* or bread wheat **(Fig.1.2) (**IWGSC, 2014). The huge diversification and adaptability in wheat are generally related to polyploidy (Dubcovsky & Dvorak, 2007). Polyploidisation participates significantly in overall plant biodiversity and evolutionary history. Genetically, hexaploid wheat behaves as a diploid because of the homologous pairing is prevented through the action of *Ph*

genes (pairing homoologous) located on the long arm of chromosome 5B (Martinez-Perez *et al.,* 2001), the major regulator of chromosome pairing and recombination in wheat. Ph1 plays an important role during meiosis recombination only occurs between pairs of homologous chromosomes. Each of the sub-genomes is large, about 5.5 GB in size and a high proportion (>80%) of highly repetitive transposable elements (TEs) (Eilam *et al.,* 2007; Wicker *et al.,* 2011).

Domestication is one of the factors that result in the increased adaptation of plants or animals to cultivation for human usage (Brown, 2010). In wheat, domestication plays a significant role in the genetics of transformation of several traits from the wild to domesticated types such as the brittle rachis (Br), tenacious glume and non-free thresh-ability in durum and common wheat (Dubcovsky and Dvorak, 2007). Also, it modified yield and yield components in the grain. As a result, when the domesticated plants moved into different regions their range of cultivation expanded, not only for adaptation to new conditions but increased desirable traits. The variation in the selection and the expansion of crops from their domestication origin gives rise to the many local landraces and varieties of crop plants (Diamond, 2002).

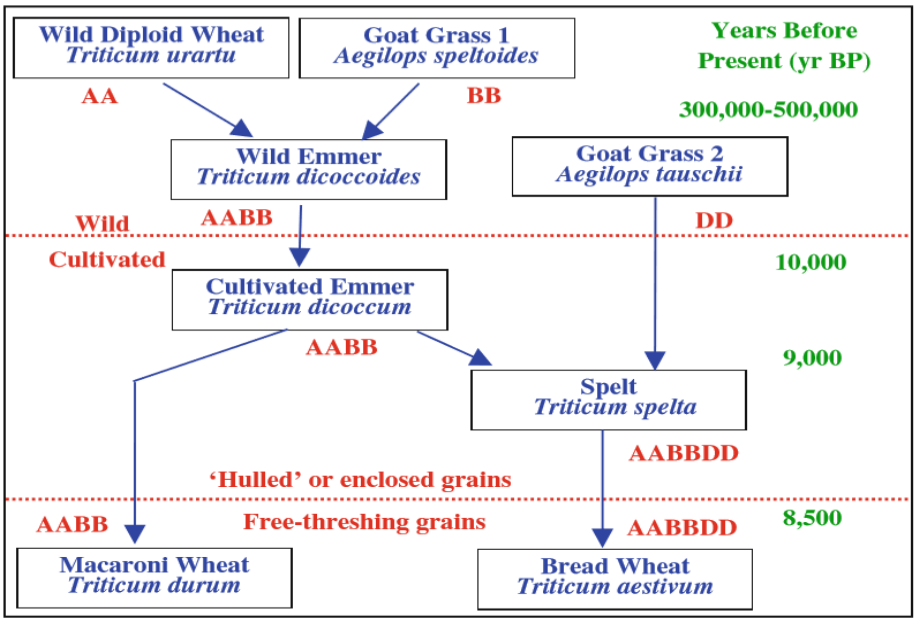


Figure 1.2 Aschematic diagram illustrates the evolution of wheat and allopolyploidization events. (Source from http: //www.newhallmill.org.uk/wht-evol.htm).

1.1.2 **Triticale**

Triticale was the first synthesized cereal allopolyploid generated by Wilson in 1876 (Wu *et al.*, 1978). It resulted from the inter- genera hybridization between wheat and rye (Lima-Brito *et al.*, 1996). The primary reasons to combine these two genomes were to obtain a plant with the grain quality of wheat and with rye's growth vigor and tolerance to grow in non-optimal environments (Coppock, 1975). However, the parents of wheat and rye influence the allopolyploid genome of triticale, where triticale was observed to have significantly more rye genomic changes than wheat. In addition, it has been reported that gene expression in triticale changed, compared with parents (Lacadena *et al.,* 1984; Somers and Gustafson, 1994; Gustafson and Flavell, 1996). As in many species, allopolyploidization has significant effects on gene expression (Chen, 2007).

The more common triticales are the hexaploid (AABBRR, six sets of chromosome) which was obtained from hybridization between diploid rye *Secale cereale* (RR) with tetraploid wheat (*T. Durum*); octaploid (AABBDDRR, eight set of chromosome), obtained by hybridization of hexaploid wheat, and rye (Ma and Gustafson, 2008). The hexaploid triticales are more successful than the octoploid since they are more fertile and higher levels of polyploidy are rare (Kuleung *et al.*, 2004; Hao *et al.*, 2013). The modern triticale has a soft grain with a high starch but with low in protein compared to wheat. However, the protein content is still higher than other cereals such as maize (Boros, 2002; Van Barneveld and Cooper, 2002).

1.1.3 **Rye**

Rye (*Secale cereale* L.) is also one of the temperate cereals, belonging to the tribe Triticeae, which is grown mainly in Europe and Northern America. It is known for being particularly tolerant to biotic and abiotic stress, a feature generally lacking in other temperate cereals. Thus, rye remains an important grain crop species for cool temperate zones and is closely related to wheat and barley. In terms of the yield, it is very low, around 2-5 tonnes per hectare with only 6,000 hectares being grown. However, its uses include grain, hay, pasture, cover crop, green fodder, and green manure. More than 50% of the annual rye harvest is used for bread making,

resulting in rich, dark bread that holds its freshness for about a week. In addition, rye is considered some important crops because it is one of the parents of Triticale, and the (short arm of rye chromosome 1) (1RS) has been introgressed into several hundred wheat cultivars (Schlegel and Korzun, 1997; Rabinovich, 1998).

Recently, reference sequences of grass genomes have become available for rice (Goff *et al*, 2002), sorghum (Paterson *et al*, 2009), Brachypodium (International Brachypodium Initiative, 2010), and maize (Schnable *et al.*, 2009). Although rye has economic importance, relatively is known about its genome sequence. As a first comprehensive sequence resource for rye was reported by Haseneyer *et al.*, (2011), more than 2.5 million reads were assembled into 115,400 contigs, representing an expressed sequence tag (EST) library which allowed the development of the Rye5k genotyping array. Recently, a whole-genome draft sequence of rye (*Secale cereale* L.) with a 7.9 GB genome was reported. The genome assembly was predicted 27 784 rye gene models based on homology to sequenced grass genomes (Bauer *et al.*, 2017).

1.1.4 **Barley**

Barley (*Hordeum vulgare.* L), belongs to the genus Hordeum, also in the tribe Triticeae (Briggs, 1978). Barley is a diploid inbreeding crop species and a model for temperate cereals especially for the species that have more complex genetics such as durum, bread wheat, and rye. It is considered the closest species to the wheat. It has been domesticated for over 10,000 years and is the fourth most grown cereal generally after wheat, rice and corn. Barley is considered a global production due to the ability to adapt to a wide range of diverse environments. Globally, 70% of barley production was used for livestock feeding, 16% changed into malt for beer-making, and 14% is consumed by a human (Zhou, 2010).

Mature barley grain has the oval shape, with tapering at each end. Most of the barley grains have a protective outer husk surrounding the grain. In terms of the structure of grain, it has five distinct tissues **(Fig.1.3).** The whole grain is covered with an outer layer with a long awn, this layer is not for protecting the grain, but also

it is semi-permeable membranes which allow for some substances, such as water into the grain, but restricting the flow of others, such as hormones. In the center, the endosperm which has a significantly higher starch content than the other products (Anderson *et al.,* 2003; Flores *et al.,* 2005). The endosperm contains around 10–20% proteins, is known as the hordeins (52%) which function as the main storage proteins. Glutelin is another major storage protein of the endosperm, which represents 23% of the total barley proteins (Sullivan *et al.,* 2013). These proteins are known as good candidates for application of food supplements owing to their functional properties, including elasticity, cohesiveness, and water holding capacity (Yalçın *et al.,* 2007; Wang *et al.,* 2010). Consequently, barley proteins have a significant impact on food development. In terms of the starch, barley grain has starch in the endosperm fraction of the barley kernel **(Fig.1.3)**, a little represents in the aleurone, subaleurone and germ tissues (Flores *et al.,* 2005). Indeed, the starch in barley grain has a high amylose-to amylopectin ratio, meaning that there is a significantly higher proportion of amylose than amylopectin. As a result, the starch in barley has lower gradation than wheat starch (Van Amelsvoort and Weststrate, 1992).

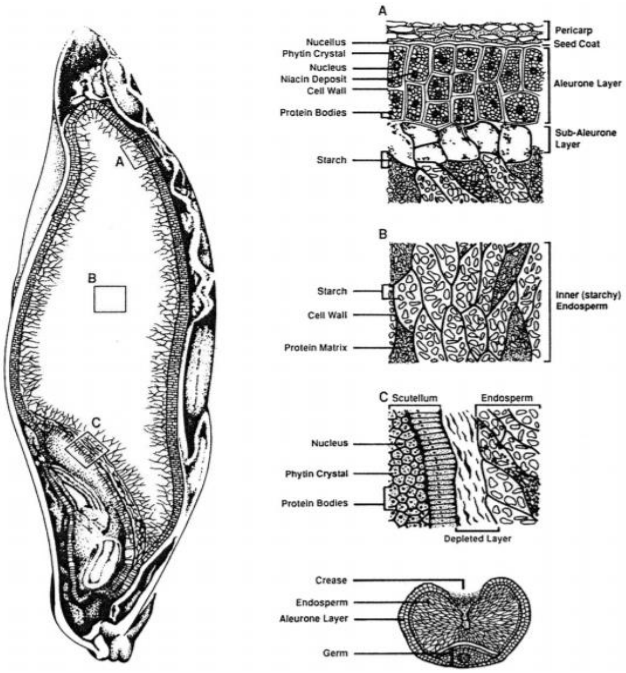


Figure 1.3 Structure of the barley grain. (From, Sullivan *et al.,* 2013).

1.2 Grain in the cereals

Globally, the grain of cereals (including wheat, barley, rice, and maize), is an important nutrient source for human because it contains starch and proteins providing the energy source for seed germination, carbon and nitrogen. Also, it can be used for feeding animals, constituting over 50% of worldwide crop production (http://www.fao.org/). The grain is a complex structure which has three genetically distinct compartments; the pericarp and seed coat (testa) are fused to form a caryopsis and associated maternal tissues, the embryo and the endosperm. During the development of grain and germination, the endosperm differentiates into many regions, starting with the outer layer of the endosperm, aleurone that has been referred as peripheral aleurone, whereas the cells on center endosperm will be as starch- and protein reserves cells.

The aleurone layer is the only living cells in the mature endosperm, unlike the cells of the starchy endosperm which undergo programmed cell death (Young and Gallie, 1999). Also, this layer shows a difference in the number of layers between the cereal grains, for example, in wheat there is one layer of cuboidal cells, distinct from the thin-walled, irregular and larger cells of central starchy endosperm, surrounding the endosperm cells, while barely has three layers.

Another significant feature in the grain is the crease region known to have a role during grain filling as part of the nutrient flow pathway (Olsen *et al.,* 1992; Wang *et al.,* 1994). Overlaying the crease in the center of the grain, is a distinct region of the endosperm named the modified aleurone (Drea *et al.,* 2005) **(Fig.1.4).** It involves developing transfer cells. According to Evers (1971), the modified aleurone can be distinguished from the endosperm cells at an early stage of 4DAA. The endosperm has no ability to connect directly to the parent tissues so that, the modified aleurone region acts in conjunction with the maternal vascular connection to facilitate efficient transfer into the endosperm (Bewley & Black, 1994; Wang *et al.,* 1994a, 1994b, 1994c). Solutes are transported from vasculature into the nucellar projection, then to the endosperm cavity where they spread across a mucopolysaccharide filling this space before uptake into the modified aleurone

(Cochrane & Duffus, 1980). The modified aleurone diffuses rapidly into the central endosperm (Bewley & Black, 1994). Consequently, this region is especially important during grain filling to improve the grain quality for commercially important grains. These distinct tissues differentiate functionally during grain development. For example, the grain has filial tissues which change rapidly during development within the endosperm (Percival, 1921). In addition, each distinct tissue has a unique temporal pattern of gene expression during grain filling because of the gene expression patterns are central to differentiation pathways. The development of grain can be classified into several stages: early development, accumulation of storage compounds, and maturation–desiccation. However, in other studies, they described the three stages as division and expansion, grain filling and maturation, and desiccation (Shewry *et al.*, 2012).

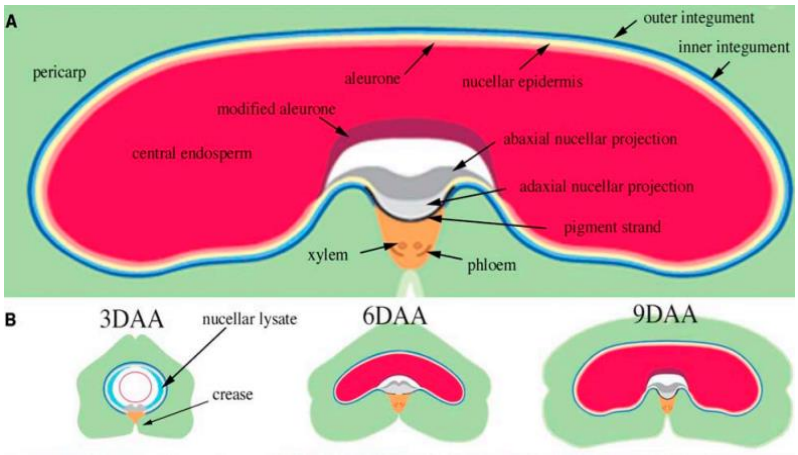


Figure 1.4 Scheme of the caryopsis in transverse section.

Showing the terminology of the cell types, based on the structure at 9 DAA. (Drea *et al.,* 2005).

1.2.1 Endosperm development and its role in seed development

During double fertilization in angiosperms, the polar nuclei (2n) fuses with the sperm nucleus (n) to produce a (3n) endosperm cell, which is triploid. This cell will develop to be endosperm tissue that shows variation between different types of cereal (Olsen, 2004). The development of endosperm in wheat and barley has been described in six stages: syncytial, cellularization, endosperm cell division, differentiation (which includes the formation of the main cell types: transfer cells,

aleurone, starchy endosperm, and embryo surrounding cells), the accumulation of storage components; and maturation (which includes programmed cell death (PCD), dormancy, and desiccation) **(Table.1.1)** (Sabelli and Larkins, 2009).The first stage is the formation of nuclear endosperm. In this stage, the nuclear endosperm cells divide repetitively to form a mass of nuclei without separation by cell walls, which happens around three to four days after fertilization. The second stage is cellularization, during which stage the preclinical cell wall forms between the sister's nuclei. There are four main types of cells in the endosperm, which are the starchy endosperm, the aleurone layer, transfer cells and the region surrounding the embryo. In this period of cell expansion, the water content increases, and starch and protein accumulate. Kowles and Phillips (1988) reported that the endosperm development in maize requires three different cell cycles: the first is cytokinetic mitosis which results in a syncytium; the second is mitosis which occurs after cellularization; the last one is endoreduplication, which is required for reiterated rounds of DNA replication without chromatin condensation.

Generally, the cell division patterns are conserved in the cereal endosperm. Several studies reported that the cell division is important, in particular the timing initiation, duration and the rate of development of endosperm with regard to the grain yield (Reddy and Daynard, 1983; Ober et al., 1991; Commuri and Jones, 1999) because the number of starch granules is dependent on the number of endosperm cells (Brocklehurst, 1977; Chojecki et al., 1986a, 1986b; Jones *et al.,* 1996). Furthermore, there are many factors have a big impact on the accumulation of starch and protein in the grain such as the final size of the cells, cell-wall extensibility, rate and duration of grain fill (Egli, 1998). It has been reported that the cell wall can determine the extent of cell enlargement (Chanda and Singh, 1998). Both the cell expansion and water accumulation stop to be replaced with starch and protein synthesis and accumulation; thereafter, the grain starts to desiccate (Rogers and Quatrano, 1983; Lopes and Larkins, 1993; Berger, 1999). This stage is called the dry stage, giving dry weight or physiological maturity **(Figure 1.5).**

Table 1.1 Main phases of endosperm development and the major change / events in wheat. (Stone and Morell, 2009).

Stages	Timing (DAF)	Major Cellular Events	Major Events in Starch Granule Development
Syncytial	0-5	Fertilization starts. Vacuoles form in cytoplasm. Proplastids are found.	Spherical A granules start initiation
Cellularization	6-8	Cell walls occur. Amyloplasts differentiate and divide rapidly.	Equatorial plates develop on A granules.
Cell division	9-14	Cells divide rapidly	Initiation of B granules occur, and equatorial plate A granules is completed.
Differentiation	15-21	Aleurone layer forms. Endosperm cells divide rapidly	Deposition of starch on A granule continues, also initiation of B granules continues.
Maturation	22-35	Both endosperm and aleurone cell divisions have stopped. Grain storage components are deposited.	The growth of A granules completes through lateral deposition of starch. Equatorial groove is less prominent. Also, the initiation and expand radially of B granules continue, C granules develop radially.
Desiccation	35 to maturity	Desiccation occurs.	Starch granules are compressed into the protein matrix. Amyloplast membrane integrity is lost.

The final cycle is endoreduplication, the reduplication rounds of DNA replication within the nuclei that do not include chromatin condensation, sister chromatid segregation, or cytokinesis, and which results in endopolyploid cells. This phase has been described as having a significant impact on the final grain size in maize (Engelen-Eigles *et al.,* 2000). The endoreduplication phase and part of the cell division phase correspond to dramatically increased growth of the endosperm and the accumulation of storage contents.

In the final stage of development, the grain undergoes Programmed Cell Death (PCD), where it has been reported that this phenomenon is similar to that observed in other plant and animal tissues (Beltrano *et al*., 1994; Young and Gallie, 1999). Generally, PCD is thought to play an important role in cereal endosperm development because it is believed to facilitate the hydrolysis of nutrients absorbed by the embryo during germination (Nguyen *et al.,* 2007). The first PCD has happened when hydrolysis starts during the seed germination in the aleurone layer

to allow the mobilization of stored nutrients which are essential for the growth of the plant (Young and Gallie, 1999). In addition, PCD has a specific pattern in the endosperm of maize, which are the central starchy endosperm cells and apical cells near the silk scar where, PCD starts at around 16 DAP. Then PCD continues from these two regions to merge at the ends. So that, by 28 DAP nearly the top half of the endosperm is dead (Young and Gallie, 1999). However, PCD has a similar process in wheat but with a more random spatial pattern. By roughly 30 DAP all endosperm cells and the aleurone have undergone PCD (Young and Gallie, 1999). Finally, grains start to dry out rapidly by losing around 10–15% of their water content, at which point they are ready for harvest.

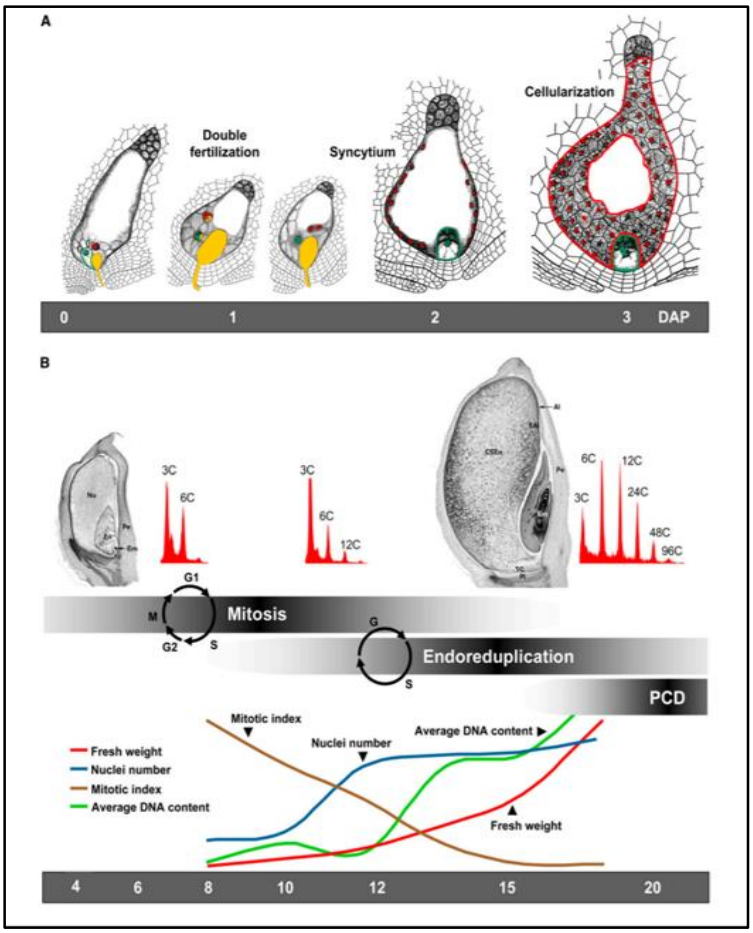


Figure 1.5 Endosperm development stages in maize.

A, the development stage of endosperm. The pollen tube and sperm nuclei are introduced in yellow, in red are polar nuclei in the central cell of the female gametophyte and endosperm nuclei, and the egg cell nucleus and embryo nuclei are mentioned in green. **B,** The endosperm undergoes a phase of mitotic cell proliferation (from 4 to 20 DAP), followed by endoreduplication (from around 8–10 DAP). The key parameters during mid-endosperm development, such as fresh weight (red line), nuclei number (blue line), mitotic index (brown line), and average DNA content. Modified from Kiesselbach (1999) with permission. **AI,** Aleurone; **CSEn,** central starchy endosperm; **Em**, embryo; **En**, endosperm; **Nu,** nucellus; **Pe,** pericarp; **PI,** placentochalaza; **SAI,** subaleurone layer; **TC,** transfer cells. (Sabelli and Larkins, 2009).

1.2.2 Structural diversity of the starchy endosperm in mature grain

The starchy endosperm is the tissue reserved for the accumulation of starch and protein, which is subsequently a nutrient source for monocots during seed germination and seedling establishment, while in dicots it supplies the embryo with nutrients during development. Additionally, it is considered an essential source of nutrients in the human diet (Lopes and Larkins 1993).

The structure of the endosperm in different species shows a considerable degree of variation. For example, the endosperm in the tomato (*Solanum lycopersicum*) has a thick, hard cell layer in the mature seed (Nonogaki *et al.,* 2000), while in Arabidopsis, the endosperm is restricted to a peripheral aleurone layer in the mature seed (Lee *et al.,* 2012). In contrast, the endosperm in a mature seed such as soybean (*Glycine max*) or pea (*Pisum sativum*) is very small (Lackey, 2010). However, cereal seeds have more complex structures with a large starchy endosperm and living aleurone layers.

Furthermore, the starchy endosperm shows some variation in the morphological in different species such as the shape and size of starch granules (Becraft and Asuncion-Crabb, 2000). For example, the endosperm cells in rice are radially symmetrical and so have the appearance of being tube-like (Srinivas, 1975); by contrast, the endosperm cells in sorghum are hard or translucent tissue, surrounding a softer, opaque core (Waniska, 2000).

1.2.3 **Starch bio-synthesis and accumulation**

The starch is a polymeric carbohydrate, it is up to 70% in the grain. It is formed and collected into semi-crystalline granules which are starch granules in the starchy endosperm cells. These granules contain two water-insoluble homoglucans, amylose and amylopectin. In the grain of the cereals, the starch is around 25–28% amylose and 72–75% amylopectin (Colonna and Buléon, 1992). Amylose is formed from polysaccharides, which is known to be linear (Hizukuri *et al.,* 1981; Shibanuma *et al.,* 1994) alpha-(1–4) linked D-glucose units. However, amylopectin is also polysaccharide, but it has a highly branched structure of short alpha-(1–4) chains linked by alpha-(1–6) bonds (Zobel, 1988).

The starchy granules can be characterized depending on their size, shape and its hierarchical structure, which are important to starch functionality (Lindeboom *et al.,* 2004; Park *et al.,* 2004). The small granule is known as (B), the average size is 5 µm, while the big granule is named (A), is around 20 µm (Karlsson *et al.,* 1983). Additionally, both have different chemical components and functions (Bechtel and Wilson, 2003). Approximately 98.5% of the chemical components of A- and B-type

starch granules are amylose and amylopectin, while amylopectin constitutes more than 50% of their chemical composition. Furthermore, various studies confirmed that both A-type and B-type starch granules have different establishment times, A-type starts at around 4-15 days after anthesis (DAA) when the endosperm cells are still actively dividing (Peng *et al.,* 2000), while B-type starch granules are established during the endosperm cell enlargement stage, which is at 15 DAA (Darlington *et al.,* 2000).

However, several enzymes are required for starch biosynthesis in the cereal endosperm including adenosine 50 diphosphate-glucose (ADP-Glc), pyrophosphorylase (AGPase), granule-bound starch synthase (GBSS), starch branching enzyme (SBE), starch debranching enzyme (DBE), and soluble starch synthase (SS) (Ball and Morell, 2003; Hannah, 2005; Keeling and Myers, 2010). The starch synthesis genes have many different isoforms, with 29 and 32 cDNA sequences encoding for starch biosynthesis proteins in rice and maize, respectively (Ohdan *et al.,* 2005; Yan *et al.,* 2009), while 26 genes in wheat (Kang *et al.,* 2012).

The enzyme involved in the synthesis of amylose is Waxy (GBSSI), while SS, is required for the synthesis of amylopectin. Also, SSI, SSIIa, and SSIII likely play important roles in the synthesis of amylopectin. There are many studies investigated the function of these enzymes such as in wheat, the result of RNAi suppression targeting by SBE IIb has been shown to increase the amylose content from 25% to 35% (Yamamori *et al.*, 2000). Similarly, in durum wheat, the RNAi construct of SBE IIa led to a 30% to 75% increase in amylose content (Sestili *et al.*, 2010). Consequently, the temporal and spatial expression of these starch synthesis enzymes can be related to photosynthetic products and starch in grain storage organs. These genes are differentially expressed in different grains of plants because of there are some differences in the functional features of starch (Kang *et al.*, 2012). Numerous studies have attempted to analyze the transcription profiles of starch synthesis enzyme genes in rice, barley, maize and wheat were determined and identified key genes that play significant roles during the development of the endosperm (Ohdan *et al.*, 2005; Radchuk *et al.*, 2009; Kang *et al.*, 2012).

The starch synthesis genes are regulated by transcription factors binding their promoters, however, the regulatory network involved in starch synthesis is not well-characterized. Several reports have shown that several TFs regulate starch synthesis genes in cereals; for example, in rice, OsbZIP58, belong to the bZIP (basic leucine zipper transcription factor) (Wang *et al.,* 2013), which is expressed in the filling stage of the grain in the endosperm (Hannah, 2005; Keeling and Myers, 2010). Another example, Rice Starch Regulator1 (RSR1), a member of the AP2/EREBP family, negatively regulates starch synthesis genes in rice and wheat (Fu, and Xue, 2010; Kang *et al.,* 2012). Knocking out *rsr1* in rice showed an increase in amylose content and the change in the structure of amylopectin, and also caused the starch granules to be loose and take on a rounded shape. In addition, *rsr1* affects the size of the seed, which became larger and showed an increased seed mass and yield. However, the metabolic pathway for starch and protein biosynthesis are coordinated in the grain development which is related in their physical association forming complexes in the wheat and maize endosperm amyloplasts (Tetlow *et al.,* 2004; Hennen-Bierwagen *et al.,* 2009).

1.2.4 Seed Storage Proteins in the grain

Cereals are considered one of the principal sources of dietary protein (Shewry and Halford, 2002). In addition, the protein in the endosperm controls the cohesive and viscoelastic properties of the dough, which are essential to the quality of such products as bread and pasta (wheat) and have the functional features of other baked goods (Sabelli and Larkins, 2009). The cereal grains contain around 8–11% protein, however, there is a significant variation between cereals in terms of protein content due to the fact that a large number of factors can have an effect including genotype, cereal type, species, and various environmental conditions such as soil, climate, fertilization and the time at which nitrogen fertilization (Belitz *et al.,* 2009).

The distribution of the proteins within the whole grain can affect protein compositional profiles. For example, the content of protein is more than 30% in the germ and aleurone layer of wheat grains while the starchy endosperm contains ~13%, and the bran ~7%. However, the starchy endosperm contains the most

proteins of the grain which, in fact, are the sources of flour. Nevertheless, proteins have a significant impact on end-use properties.

Seed storage proteins have been reported in the grain specific to the endosperm of the grain (Miflin *et al.*, 1981, 1983; Shewry *et al.,* 1983). Also, they have been given many names in different species of cereals such as gliadin in wheat, secalin in rye, hordein in barley, avenin in oats, zein in corn, and kafirin in millet, sorghum, and oryzin in rice. The prolamins in cereals have been identified as the main protein except in oats and rice, where they are considered to be a minor protein component. In rice and oats, the main protein is gluten and 12S globulin (Zhao *et al.,* 1983). Additionally, it was confirmed that prolamins of species such as wheat, barley and rye are related to each other in terms of their amino acid sequences (Kreis *et al.,* 1985), as compared to other species such as maize, sorghum, millets to Panicoideae which are much more variable in structure, possibly due to separate evolutionary origins (Shewry and Halford, 2002).

The storage proteins in the cereals can be broadly divided into four categories (albumins, globulins, prolamins and glutelins) depending on their solubility and other features (Osborne, 1907). Consequently, the various classifications of storage proteins are dependent on their different features such as the amino acid sequences (Miflin *et al.,* 1983). Another classification of proteins which are the high molecular weight (HMW), sulphur poor (S-poor) and sulphur rich (S-rich), prolamins based on differences in MW and sulphur (cysteine, methionine) content (Shewry and Tatham, 1990). Consequently, they can be categorized into three groups which are a high molecular weight (HMW) group, a medium molecular weight (MMW) group and low molecular weight (LMW) group according to related amino acid sequences and molecular masses (Wieser and Koehler., 2008).

SSPs (seed storage proteins) genes are highly expressed in the grain particularly during the filling stage. The spatial and temporal expression patterns for those genes are expected to be regulated by TFs by binding specific motifs in their promoters. The accumulation of these proteins is controlled primarily at the transcriptional level, during the middle and late stages of endosperm development according to the

specific spatial and temporal patterns (Woo *et al.,* 2001; Shewry *et al.,* 2003; Halford and Shewry, 2007; Xu and Messing, 2008). Several SSPs genes have been characterized such as their promoters and their expression patterns (Hood *et al.,* 2003; Qu and Takaiwa, 2004; Furtado *et al.*, 2008). For example, many O2 proteins can bind the target elements in the promoters of SSPs such as O2 can bind the target element (Ciceri *et al.,* 1997) such as the endosperm-box which is very highly conserved, located 300 bp upstream of the transcription initiation site which is around 30 bp long (Forde *et al.,* 1985). Generally, the endosperm box has two motifs separated by a few nucleotides (Forde *et al.,* 1985). These motifs are (5*-TGTAAAG-3*) and G (A/G) TGAGTCAT, with shifting regions in between. The first consensus sequence is named the prolamin box (P box), or E motif (Hammond-Kosack, *et al.,* 1993; Müller and Knudsen, 1993). Prolamin box binding site for TF (PBF) encodes a member of Dof class of plant TFs in Arabidopsis, pumpkin, and tobacco (Paolis *et al.*, 1996; Yanagisawa, 1996) **(Fig.1.6)**. The second motif in the endosperm box is GCN4-like (GCN) motif or N motif (Diaz and Carbonero, 1998). The GCN4-like sequence is bound by basic Leu zipper transcription factors, such as OPAQUE2 (O2) in maize (Kawakatsu and Takaiwa, 2010) **(Fig.1.6)**. In addition, it has been reported that other bZIPs can heterodimers with (OHP1) and OHP2 (Pysh *et al.,* 1993), barley BLZ1 can transactivate the B-hordein promoter by GCN4, (Vicente-Carbaj-osa et al., 1998) and rice RISBZ2/REB trans-acti-vates from thea-globulin promoter via an ACGT motif (Nakase *et al.,* 1997). All these TFs can form complexes with O2 orthologs in vitro or in vivo, but these complexes is still unknown.

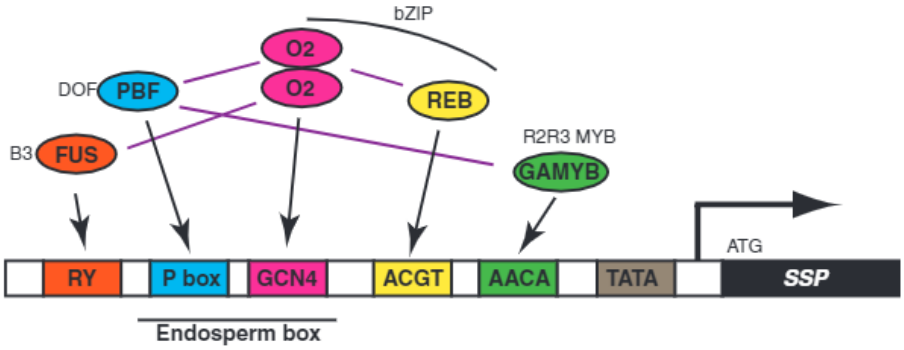


Figure1.6 Schematic diagram of core promote of storage protein genes.

Trans- and cis-elements of promoters of (SSP) gene regulation. (From Kawakatsu and Takaiwa, 2010). Showing the TFs and binding sites on the promoter. For example, bZIP such as O2 binds to O2-binding sites (GCN4 motif); DOF TFs such as PBF binds to P box; REB from bZIP binds to (ACGT); MYB TF: barley GAMYB binds to the AACA motif; B3 TF (FUS3) binds to RY.

1.3 Grain Quality and Yield

The term “quality” has a complex definition and covers a wide spectrum of variable concepts and characteristics. “Yield grain” can refer to the mass of product at final harvest, or it refers to multiple phenotypic traits of the grain, flour, dough, and final products which can determine an overall quality and end-use quality of the product (Guzman *et al.,* 2016). The quality evaluation involves several features of a product that are related to various quality characteristics such as milling efficiency, end-use, grain shape and appearance, nutrition, etc. However, there is no specific assessment of the quality of grain; each country has its own protocol depending on their needs (Bao *et al.,* 2014). Wheat grain quality can be assessed by identifying either physical or chemical characteristics. For example, physical characteristics such as grain vitreousness, color, weight, shape, and hardness (Gaines *et al.,* 1997), while chemical characteristics include protein content, SDS-sedimentation value, and gluten strength.

On the other hand, the yield refers to the amount of agricultural production harvested per unit area of land. Furthermore, the yield can be divided into its two major features: the number of grains set per ground area, and the average grain

weight (Reynolds *et al.,* 2009; Pedro *et al.,* 2012). The number of grains depends on a complex process that begins pre-anthesis. Before anthesis, each spikelet produces florets, which later become fertile and capable of bearing grains, so that the number of grains is related to the number of fertile florets (Miralles *et al.,* 2000). It seems possible that the grain numbers can be determined by the number of fertile of florets rather than the actual number of florets produced (Ferrante *et al.,* 2012). The grain yield is associated with enhanced starch biosynthesis of grain which is known as the main storage reserve in the endosperm of grain. The properties of starch granules in cereals have an effect on the end-use such as the eating and cooking qualities, even on industrial processes such as paints, varnishes, paper coverings, superabsorbers and adhesives (Nakamura, 2002; Morell and Myers, 2005). Thus, it is essential to understand the processes of starch biosynthesis to improve grain yield and quality in cereals. Both carbohydrates and storage proteins strongly impact on the processing quality of flour and the hardness of the endosperm and improve the nutritional quality and yield of grain. While the coordination of starch and protein metabolism in the grain was noted earlier in terms of their physical association in the endosperm, recently, it has been confirmed that both are linked by regulation via transcription factors such as O2, PBF (Zhang *et al.,* 2016).

1.4 General features of TFs in plants

In cells, some genes are expressed constitutively (Albert *et al.,* 1992), whilst others respond to specific signals (Adam *et al.,* 1994; Tonoike *et al.,* 1994). In both situations, the transcription factors act in regulating the expression of genes by interaction with cis-acting elements and with each other in regulatory complexes. Consequently, the variation in the expression of genes could be related to the mechanism of regulating these genes by TFs, or the changes which happen during the development (Chen *et al.,* 1997; Kater *et al.,* 1998). Also, a rapid change in the structure of these genes during evolution (Doebley and Lukens, 1998).

The classification of TFs depends on many features such as the conserved domains, and they can be divided further depending on the conserved residues

within, adjacent and between these conserved domains (Riechmann and Meyerowitz, 1998). In addition, TFs can be divided into many functional groups. For example, TFs can be classified as activators or repressors. These proteins regulate their targets by binding specific motifs in the promoter of their targets. A further classification is either TFs are co-activators and/or co-repressors, these do not bind the promoter of their targets directly, but they interact with other proteins (activators and repressors) to bind the promoter (Singh, 1998).

1.4.1 **bZIP family**

Transcriptional regulatory factors of the basic leucine zipper (bZIP) have been characterized in all eukaryotes (Wingender *et al.,* 2000). The name of this family comes from the bZIP domain which is present in all members **(Fig.1.7).** Several studies of genetics and molecular show that some bZIP factors in *Arabidopsis thaliana* (AtbZIP) have diverse biological processes such as pathogen defense, responses to environmental or pathogen challenge, seed maturation and flower development (Jakoby *et al.,* 2002). Some members of the bZIP transcription factors can be expressed constitutively while others are expressed in specific organs (Schindler *et al.,* 1992; Rodriguez-Uribe and O'Connell, 2006) in a stimulus-responsive (Vetten and Ferl, 1995), development-dependent (Chern *et al.,* 1996), and cell cycle-specific manner (Minami *et al.,* 1993).

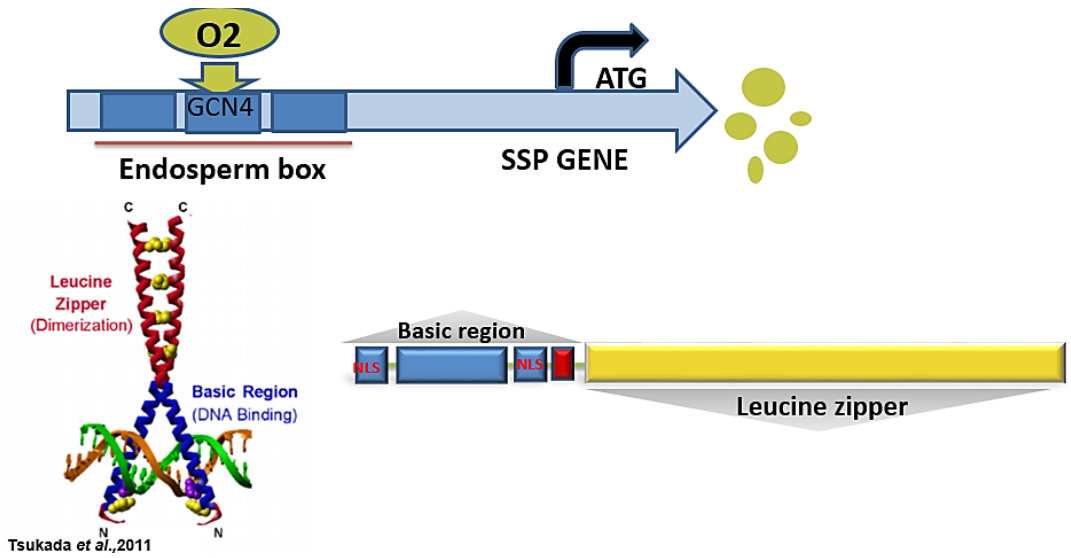


Figure1.7 The structure of bZIP domain which present in all bZIP members.

However, it was confirmed that there are bZIP TFs can be grain-specific, especially those expressed in the endosperm of the seed (Neto *et al*, 1995; Vicente-Carbajosa *et al.,* 1997). One of the endosperm-specific TFs is OPAQUE2 (O2) from maize which plays a role in regulating the α- and β-zein genes in maize grain (Schmidt *et al.,* 1987; Schmidt *et al.,* 1992). O2 was studied in detail because it has an important role in improving the protein composition of the maize endosperm with mutants containing higher lysine and tryptophan in addition to the decreased zein (Gibbons and Larkin, 2005). In wheat, SPA was identified as a key regulator of storage protein genes (Albani *et al.,* 1997). In 2009, Ravels confirmed the SPA has a big impact on he composition of protein, dough viscoelasticity and grain hardness.

In rice, it has been identified several bZIPs members which play a role during the seed development such as RITA/OsbZIP20 (Izawa *et al.,*1994), RISBZ4/ OsbZIP15, RISBZ5/OsbZIP52 and RISBZ1/OsbZIP58 (Zhang *et al.,* 2014). OsbZIP58 rice was found to regulate a broad range of roles in starch synthesis (Wang *et al.*, 2013). However, in dicots such as Arabidopsis were identified two Opaque2-like (O2-like) factors which are AtbZIP10 and AtbZIP25 (Lara *et al.,* 2003), are close to the O2 monocot members. Furthermore, bZIP53 has been reported to play a crucial role in controlling seed maturation gene transcription in cooperation with several TFs (Alonso *et al.,* 2009).

In Barley, HvBLZ1 was identified with similarity to various bZIP members of the O2 subfamily of bZIPs, especially OsREB in rice and ZmOHP1 maize with approximately 70% amino acid identity compared to 35% with ZmO2 and TaSPA (Diaz and Carbonero, 1998). Additionally, Onate *et al.,* (1999) characterized and analyzed HvBLZ2 as paralogous to HvBLZ1. Both have similarities in the protein sequences, although HvBLZ2 is an endosperm-specific TF, while HvBLZ1 is expressed in other tissues such as in leaves and roots (Diaz and Carbonero, 1998). Moreover, it has been confirmed that both HvBLZ1 and HvBLZ2 undergo hetero-dimerization *in vivo* (Onate *et al.,* 1999). It is known HvBLZ2 is very close to TaSPA wheat (Albani *et al.,* 1997) and ZmO2 maize (Schmidt *et al.,* 1992).

1.4.2 **AP2 transcription factors**

One of the largest families of transcription factors in plants is AP2/ERF (APETELA2/ Ethylene-Responsive Element Binding Factor) (Song *et al.,* 2005; Golldack *et al.,* 2011). It was thought even quite recently that the AP2 family was restricted to plants; but it has been reported that the AP2/ERF family members can be found in various protists and ciliates (Dietz *et al.,* 2010; Licausi *et al.,* 2013). The name of this family was derived from the AP2 domain which occurs in all of its members. The AP2 superfamily can be divided into three separate families (ERF, AP2, and RAV) **(Fig.1.8)** (Okamuro *et al.*, 1997; Licausi *et al.,* 2013; Lata *et al.*, 2014). There are variations between the families such as the number of AP2 domains, so that the genes with a single AP2 domain and the fewer introns belong to the ERF family (Nakano *et al.,* 2006). Genes with an AP2 domain and an associated B3 DNA binding domain belong to family RAV. The AP2 family has two AP2 domains. Further analysis of the members of the AP2 family allowed their division into the euAP2 and ANT lineages depending on the amino acid sequence of the double AP2 domain (Wolfe *et al.,* 2000). However, both groups have AP2 domains that show higher similarities to the one contained in double-AP2 proteins than to the AP2 domain of the ERF proteins. In terms of the double AP2-domain in members of the AP2-like genes it was reported that it is unclear if the functional and structural significance of having two AP2 domains can act as a single domain in the AP2-like proteins, however, Jofuku *et al.,* (1994) confirmed that mutant alleles of AP2 show that both domains (R1 and R2) are required for AP2 function.

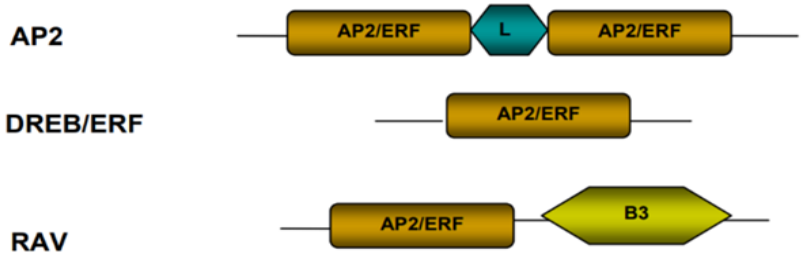


Figure 1.8 Schematic diagram of three families of AP2 superfamily domains (AP2, DREB and RAV) (Source, Saleh, 2003**).**

The evolution of gene structure in the AP2-like gene family is shown in **(Fig.1.9).** The members of the euAP2 lineage have miR172-binding sites in their post-domain region, the ANT lineage has a 10-aa insertion in R1, while R2 has a 1-aa insertion (Kim *et al.,* 2005).

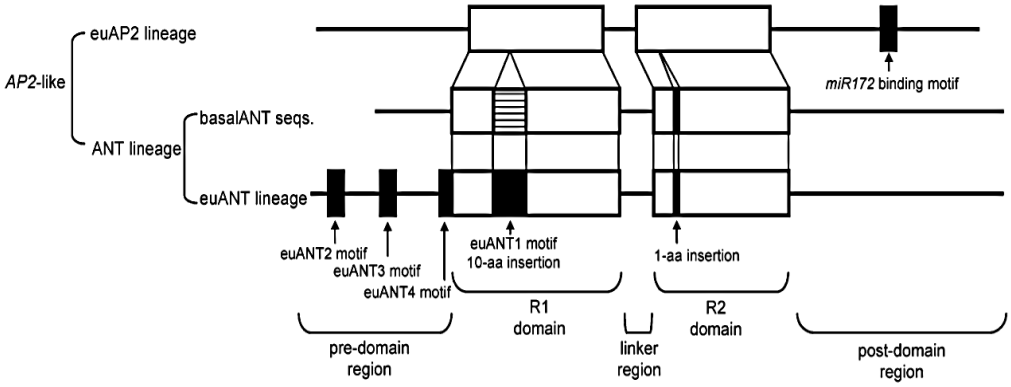


Figure 1.9 Schematic illustrates the structure of AP2-like genes. (From, Kim *et al.,* 2005).

However, Kim *et al.,* (2005) reported that phylogenetic analysis of the AP2-like family showed a duplication event resulting in the euAPETALA2 (euAP2) and AINTEGUMENTA (ANT) gene lineages, which can be recognized between these lineages by the miRNA172 (miR172) binding site, which is present in the euAP2 lineage.

Despite AP2 members share a similar structure, they show a wide diversity of functions. Various studies suggested that the members of the AP2 family play significant roles in several processes such as biotic and abiotic stress responses, primary and secondary metabolism, and developmental processes by regulating gene expression (De Crozé et al., 2011; Aoyama *et al.,* 2012; Licausi *et al.,* 2013). In addition, AP2 genes are involved in the early stage of floral development by enhancing the formation of the floral meristem. This function is controlled by cooperation with a group of floral meristem genes, *APETALA2 (AP2), APETALAl (AP7), LEAFY (LFY), and CAULIFLOWER (CAL)* (Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993). In addition, AP2 genes involved in the specification of sepal and petal are considered as a part of the combinatorial

genetic mechanism of floral organ identity determination (Komaki et al., 1988; Kunst et al., 1989; Bowman et al., 1989; 1993). Also, they are required for normal ovule and seed development (Jofuku et al., 1994; Leon-Kloosterziel et al., 1994; Modrusan et al., 1994), and they have a function in embryo development by controlling the size and cell number (Jofuku et al., 1994; Ohto et al., 2005).

Generally, most AP2 functions are related to flower development but AP2 genes are expressed in both floral and vegetative tissues such as inflorescence meristem and throughout the floral primordia from the early stages to the late stages of flower development. Jofuku et al., (1994) provided an overview of AP2 being expressed in all organs but more highly in petals and specific tissues of stamens and carpels, including ovules. In addition, AP2 genes are expressed in the vegetative leaves and in the stem.

Fu and Xue (2010) reported that one of the euAP2 genes which are called Rice Starch Regulator1 (RSR1). It regulates starch synthesis genes, and it has a significant effect on in the starch content and structure, starch granule morphology, and gelatinization properties. Kang et al., (2012) characterized transcription levels of Rice Starch Regulator1 (RSR1) to show it was significantly and negatively associated with the transcription levels of eleven starch synthesis enzyme genes during the wheat grain-filling stage.

One of the most important euAP2 genes in grasses is the major domestication gene is Q which presents in varieties of cultivated wheat, while the wild-types of wheat have q allele, which is related to the speltoid spike (a spear-shaped spike with an elongated rachis) and non-free-threshing grains (Fans et al., 2005). It is known that the Q gene is located on the long arm of chromosome 5A (5AL) (Simons et al., 2006), it was reported that there is polymorphism on chromosome 5A (AP2-5A) for both Q/q alleles, the 5B homologous (AP2-5B) is a pseudogene, while the 5D homologous (AP2-5D) plays a role in encoding a functional protein that participates in the suppression of the speltoid phenotype (Zhang et al., 2011).The *TaQ* gene in wheat (*Triticum aestivum*) is related to *ZmIDS1/OsIDS1* genes (Zhang *et al.,* 2011). Q has highly pleiotropic functions that influence a wide range of domesticated

characters such as inflorescence compactness, fragility, and free-threshing (Simons *et al.,* 2006), the number of grains on the inflorescence (Komatsuda et al., 2007) and floret fertility (Ramsay *et al.,* 2011). Furthermore, TaQ also has an impact on wheat height (Zhang *et al.,* 2011). It has been reported that there was a significant increase in height that was also observed in Q-5A mutants by Greenwood *et al.,* (2017). Recently, Liu *et al.,* (2018) confirmed the role of Q in the controlling spike architecture by interacting with co-repressor (TOPLESS) through EAR motif (LDLNVE). Furthermore, Xu *et al.,* (2018) confirmed the effects of Q on two significant features, spike density and grain protein content (GPC) by detecting a new allele, Qc1. Also, it suggests that the overexpression of Qc1 lead to the increase the expression of storage protein activator gene (SPA) which increase the synthesis of storage protein (Ravel *et al.,* 2009). These results provide new insight into the role of the TaQ gene, which is considered one of the most important domestication genes for wheat and breeding. TaQ gene is the target of miR172 as euAP2 transcripts, providing another mechanism in the regulation of spike and floret development (Zhu and Helliwell, 2010; Wang *et al.,* 2015).

1.5 miRNA in the Poacea

miRNAs are classes of small, non-coding regulatory RNA of around 20-22 nt (nucleotides). They could regulate the expression of their targets. Briefly, miRNAs are encoded by endogenous miRNA genes and transcribed by RNA polymerase II into primary miRNAs (pri-miRNA) having partially double-stranded stem-loop structures (Jones-Rhoades *et al.,* 2006). In a two-step process, pri-miRNA is cleaved by a DCL1 enzyme resulting in the production of a pre-miRNA and a *mature* miRNA duplex. The mature miRNA is joined with RISC and mediates the degradation of the mRNA target. The mechanism for down-regulation of the level of protein of their target genes via miRNA can be either translational repression (Chen, 2004; Brodersen *et al.,* 2008), through cleavage of the transcript (Llave *et al.,* 2002) or transcriptional inhibition (Bao *et al.,* 2004; Khraiwesh *et al.,* 2010) **(Fig.1.10).**

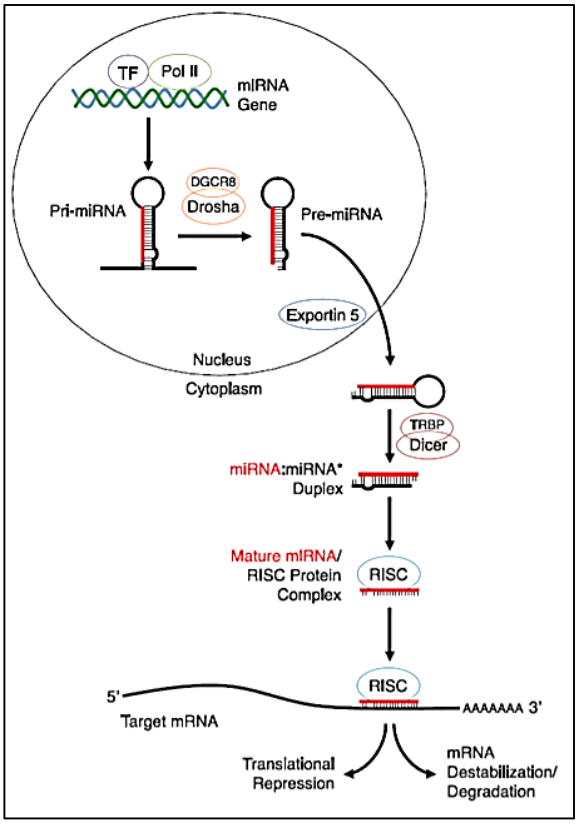


Figure 1.10 A schematic model of miRNA biogenesis and its role in plants. (From, Zhang *et al.,* 2006).

miRNAs are evolutionarily conserved across species. The first miRNAs were identified in *Caenorhabdities elegans* (Lee *et al.,* 1993; Wightman *et al.,* 1993). However, several miRNAs in plants have been identified and characterized by experimental and computational approaches were used successfully to identify various conserved miRNAs (Wang *et al.,* 2004; Dezulian *et al.,* 2005). In the plant microRNA database (PMRD), there are approximately 8500 mature miRNAs from 121 different plant species which had been detected (Zhang *et al.*, 2010). The majority were identified in species including rice (*Oryza sativa*) at 242, and 131 in *Arabidopsis* (Sun, 2012). 96 were identified in maize, 72 in Sorghum, 22 in soybean, and 16 in sugarcane. Recently, Meng *et al.,* (2013) identified 540 conserved miRNA families and 182 novel miRNA families during grain-filling in wheat grain. However, the most conserved miRNAs in grain are miR156 and miR172 whose functions have been studied in detail. For example, miR172 regulates the *AP2-like* genes that play a role in controlling floral organ identities in rice, maize, and barley (Lauter *et al.,* 2005; Zhu *et al.,* 2011).

miRNAs regulate their target genes by binding to complementary sequences located in the transcripts produced by these genes. Their binding sites are in the 3' untranslated regions of the target mRNAs (Slack *et al.,* 2000; Reinhart *et al.,* 2002). Comparing the plant miRNAs sequences and their target binding sites, miRNA sequences have a high similarity or nearly completely complementarity with their targets (Rhoades *et al.,* 2002). Most of the putative TFs targets of miRNAs have been implicated in various developmental processes such as the control the root and shoot architecture, transitions from the vegetative to reproductive phase (Chen *et al.,* 2009; Curaba *et al.,* 2012) and leaf and flower morphogenesis and cell differentiation (Rhoades *et al.,* 2002). Consequently, the miRNAs are likely to control the timing of flowering and development of plants by down-regulating their targets of genes.

1.6 Aims and justification

Grain quality in wheat is determined by the combination of multiple factors at several levels. From macroscopic features including grain position, size, and shape, through to more detailed features within the grain including internal tissue organization (encompassing the maternal fruit tissues and filial embryo/endosperm) and the starch/protein profiles within these tissues, particularly the endosperm. Protein and starch accumulation in the cereal grain endosperm are intertwined processes, temporally and physically, and in the maturing grain, the storage protein bodies are packed between the starch granules. Within the endosperm, virtually all the accumulation of reserves is in the central starchy endosperm excluding the aleurone layers (peripheral and modified).

The qualitative manipulation of nature's central food source depends on the detailed understanding of the regulatory mechanisms controlling metabolism. Identification and characterization of the maize OPAQUE2 (O2) bZIP transcription factor regulating zein storage protein genes has meant that the gene is the center of grain quality programs in maize. But despite extensive knowledge about cereal grain protein and starch metabolism and biochemistry we know relatively little about how protein and starch accumulation is regulated genetically, particularly in wheat. In the

rice and maize (RISBZ1 and O2, respectively) (Kawakatsu *et al.,* 2009; Zhang *et al.,* 2016) have recently been shown to regulate starch biosynthesis genes in addition to its role in storage protein gene regulation. Therefore, we proposed to study regulators of starch and protein metabolism in parallel. If we can establish detailed accounts of the transcriptional regulation mechanisms of genes in storage protein and starch metabolism, we will be able to examine variations in regulatory processes between cultivars and species and explain variation in quality profiles with the ultimate possibility of being able to manipulate them. Consequently, the aim of this thesis was to understand the conservation/ diversification of candidate TFs regulating protein and starch metabolism in the endosperm of the grain.

Our central hypothesis is that the variation in the TFs and the mechanism of regulation storage protein and starch genes in different cultivars and species could explain the variation in quality profiles. For that, in order to produce a complete picture of the two groups of candidate TFs which are O2 genes (TaSPA/TaBLZ1), (HvBLZ2/HvBLZ1), and euAP2 genes (RSR1, and Q). Therefore, the general objective is that comparing the orthologous and paralogous to identify the similarities and differences between them in several aspects in different varities of wheat (two hexaploid wheats such as (*Chinese spring* and cadenza) and tetraploid wheat such as Durum, barley, rosner and rye. Therefore, the aim was pursued through the following objectives:

1. Clarify and identify the TFs either the O2 or euAP2 members by constructing phylogenetic trees, and analyzing the protein structure to detect the similarities and differences between them which can reflect the relation between them and their common function.
2. Comparison analysis of O2 family (TaSPA/TaBLZ1) and their orthologous (HvBLZ2/HvBLZ1) in terms their spatial and temporal expression patterns during the grain development and other tissues in different varieties of wheat, barley, Rosner, and rye via various techniques such as RT-PCR,qRT-PCR, and mRNA In Situ Hybridization.
3. Investigate the function of TFs which regulate SSPs genes and starch synthesis genes in the endosperm by identifying cis-elements in their

promoters, the common function between the members of O2 such as (TaSPA/TaBLZ1) and their orthologous (HvBLZ2/HvBLZ1), and TaRSR1/HvRSR1.

4. Examine the interaction between the TFs in wheat via yeast two hybrid system, and their role in the activation of the promoter by transient assay.
5. Assess the general structure of grain by using SEM. Also, the total starch and in the endosperm was measured by using a commercial kit which is the Megazyme total starch assay procedure (Megazyme Total Starch Assay Kit), and for amylose content, (Amylose/ amylopectin assay procedure kit) was used.

Chapter 2

Phylogenetic & Protein sequence analyses

Abstract

Background: Transcription factors are the DNA-binding proteins that regulate genes expression that involve in a variety of developmental processes. To understand the function of these TFs, and possible variation between species, protein sequences analysis and phylogenetic tree could be informative focusing on of the putative functional domains which are DNA binding region/domain (DBD), a transcription regulation domain (RD) and a nuclear localization signal (NLS).

Result: Identify the members of the TFs which belong to O2 family (bZIP), and euAP2 (AP2 superfamily). The phylogenetic and protein analyses were used to confirm the similarities between them, which show that both are specific monocots lineages. Although they share conserved structures, there are variations between them in terms of amino acids substitution and conserved motifs. Furthermore, the microRNA172-binding sequence was identified as a conserved motif which restricted to the euAP2 lineage. However, structural changes to these TFs may represent a significant evolutionary drive (Doebley and Lukens, 1998).

Conclusion: the main features of TFs were highly conserved within the members of each family, suggestive of common functionality. However, the variation in the protein sequences, particularly the conserved domains is identified and discussed.

Chapter 2 Phylogenetic & Protein sequence analyses

2.1 Introduction

Transcription factors in the endosperm have been studied in detail in different species of plant and some cases have been functionally characterized. TFs can be divided into different families depending on the DNA binding domain and other conserved regions as follows:

2.1.1 DNA-binding domains (DBD)

The DNA-binding domain is one of the most important regions in TFs and they are characterized by the basic amino acids that contact DNA bases at cis-acting elements. This region plays a significant role to determine the specificity of binding DNA (Hung *et al.,* 1996). In addition, some residues in the domain are highly conserved between the members of the same family, for example, a single arginine residue in bZIP domains (Aukerman *et al.,* 1991), which it likely plays an important role in the function of these TFs. The spatial arrangement of these residues of amino acids in the domain plays a significant role in terms of the binding DNA and the specificity for TFs (Liu *et al.,* 1999).

2.1.2 Oligomerization domains

Homo-dimerization and hetero-dimerization influence significantly on the DNA binding specificity, the affinity of transcription factors for a promoter and nuclear localization (Katagiri *et al.,* 1992; Sainz *et al.,* 1997). The oligomerization residues generally are highly conserved between the members of the same family of TFs however, a variation between the members in terms of the length of the oligomerization domain can be observed in some cases. Consequently, this variation could play a significant effect on the mechanism of transcription and the ability of these members to modulate the gene expression in the plant.

2.1.3 Transcriptional regulation domains (RD)

TFs can be classified into several functional types such as repressors or activators (Singh, 1998). The mechanism of the repressor of transcription factors in the plant is unclear but there are many possibilities including the repression of gene expression by preventing the activators of transcription factors to bind their targets of promoter via competitive binding. Another possibility is the dimerization

between the regulation domains, as well as, the interaction between the regulators and repression domains (Chen *et al.,* 1996). For example, two bZIP of rice (OsZIP-2a and 2b) have the ability to dimer with wheat EmBP1 in vitro which prevent them to bind their target of promoters (Nantel *et al.,* 1996).

2.1.4 **Nuclear localization signals**

Most TFs have conserved regions of amino acids which are rich in arginine (R) and lysine (K) which are known NLSs (nuclear localization signals) (Klinge *et al.,*1996; Lyck *et al.,* 1997). Others lack in NLSs, and it is likely those TFs are imported into the nucleus by dimerizing with other proteins that have this signal (Goldfarb and Lewandowska, 1994). In plants, the first description for NLSs was in the tobacco TGA1 TGA1a and TGA1b bZIP proteins (van der Krol, and Chua, 1991).

Up to now, far too little attention has been paid to investigate in the O2 family and euAP2 in monocots in terms of the phylogenetic analysis and the protein structure. Therefore, in this chapter, I focused initially on O2 members to identify the orthologous and paralogous which are (TaSPA/TaBLZ1), (HvBLZ2/HvBLZ1) in monocots such as *Triticum aestivum, Hordeum vulgare*, *Oryza sativa, Zea mays* and *Brachypodium distachyon,* in dicots such as *Arabidopsis thaliana, Brassica sp.* to establish the background on their protein sequences in terms of (the structure of bZIP domain and conserved motifs) which could reflect the common features and functions. The second part is related to euAP2 proteins (RSR1 and Q) confirm the relation between them as paralogous and the relation with euAP2/TOE3 lineage by constructing a phylogenetic tree, and analyzing the protein sequences to address whether the sequences play role in determining the regulation specificity of these proteins. Additionally, the euAP2 genes possess microRNA172 (miRNA172) binding site which has a crucial role to regulate their target genes. So that, I identified the type of the miRNA in most the members of euAP2 in monocots and dicots to investigate the role of miR172 in regulating their targets.

2.2 Materials and Methods

2.2.1 Phylogenetic analysis

The protein sequences of bZIP and AP2 orthologous of wheat and other species of monocots and dicots were obtained by (BLAST X) searches using known rice, maize, and *Arabidopsis* as queries, by using available data in TAIR (https://www.arabidopsis.org/);NCBI(https://www.ncbi.nlm.nih.gov);and Ensemble Plants (https://www.plants.ensembl.org/);Phytozome v12.1(https://phytozome.jgi.doe.gov/) The orthologous of the genes are listed **(Appendix 7.Tables 7.2.3;7.2.4).**

The bZIP and AP2 full length of amino acid were aligned using Clustal X (Thompson *et al.,* 1997), the unrooted phylogenetic tree was reconstructed using Neighbour-Joining (NJ) (Dayhoff, 1972). Another method was used, Maximum Likelihood (ML) analysis, and was carried out using RAxML (Stamatakis, 2006) with the distances computed using the JTT+T (Jones *et al.,* 1992). Bootstrap values from 1000 replicates were indicated at each node. To confirm this result, another method was used to construct a tree, taking a Bayesian approach. The MAFFT alignment tool was used for protein sequences (Nuin *et al.,* 2006) while WAG+G was used as the protein substitution model. Multiple aligned protein sequences were exported to MEGA 6 to choose various attributes of the substitution models for the protein sequences, and which was used to ensure that a reliable tree was computed, then the final output of the tree was produced using the Geneious software (V.10.0.2) (https://www.geneious.com) (Kearse *et al.,* 2012).

2.2.2 Structure of bZIP and AP2 genes

By using the gene structure display server v2.0 (http://gsds.cbi.pku.edu.cn/), the full length of CDS sequence and genomic DNA sequence were loaded to obtain the distribution of exons and introns in all our targets (Guo *et al.,* 2007). In addition, one of the features of these TFs is the distribution of introns and exons, so that we analyzed the intron distribution patterns and intron splicing phase within the genes by using GeneWise website (https://www.ebi.ac.uk/Tools/psa/genewise/) (Birney *et al.,* 2004).

2.2.3 **The Leucine zipper analysis**

The leucine zipper was identified in the C-terminus of the domain, and it is composed of several repeats of Leucine or other bulky hydrophobic amino acids (Ile, Val, Phe, or Met) (Nijhawan *et al.,* 2008). The dimerization specificity properties of the O2-like family was analysed considering three structural properties of the dimeric α-helical leucine zipper coil structure: (i) the length of the leucine zipper; (ii) the replacement of asparagine or a charged amino acid in the hydrophobic interface; and (iii) the presence of interhelical electrostatic interactions. To identify the C-terminus boundary, three criteria were used: (i) the existence of a proline or a pair of glycines in the α-helical structure of a leucine zipper in the C-terminal; (ii) the existence of a leucine in the d position; and (iii) the presence of charged amino acids in the g and e positions. All these features reflect the significance of the leucine zipper, which plays an important role in homo- or heterodimerization. The same criteria were used for Arabidopsis (AtbZIP) and *Oryza* (OsbZIP) to define the N-terminal and C-terminal boundaries because there are no studies in this family. The leucine zipper was divided into many small fragments called heptads whose length is around seven amino acids which were given the letters g, a, b, c, d, e and f in order to analyze the dimerization. Also, the detailed analysis of the type of amino acids present at the a, d, e and g positions allowed for a comparison with Arabidopsis and *Oryza sp*.

2.2.4 **Protein properties and detection of additional conserved motifs**

Searching for conserved domains within proteins or coding regions was achieved using online tools such as searching for conserved domains within proteins or coding regions, NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer *et al.,* 2014), SMART (Simple Modular Architecture Research Tool, (http://smart.embl-heidelberg.de/), and Pfam (http://pfam.xfam.org/). The sequences were further analyzed to confirm the presence and integrity of the conserved domain through the ExPASy Proteomics Server (http://prosite.expasy.org/) and Interpro (http://www.ebi.ac.uk/). From the aligned proteins sequences, the conserved regions were detected by using the BioEdit software. Additionally, to identify additional conserved motifs outside the bZIP and AP2 domains, the Motif Elicitation tool (MEME version 4.9.1, http://meme.nbcr.net/meme/) (Bailey *et*

al., 2009) was used. The limits for maximum width, minimum width and the maximum number of motifs were specified as 50, 10 and 25, respectively.

2.2.5 Identified miRNA in members of euAP2 (RSR1 clade, Q clade)

To confirm that euAP2 genes have miRNA, the mRNA sequences of wheat, barley and other species were extracted from different sources. Then, these genes sequences were aligned to detect a conserved sequence in the 3' UTR region. miRBase website (http://www.mirbase.org/) was used for identify the type of miRNA.

2.2.6 Subcellular Localization Experiments

All details of this experiment have been mentioned in chapter four.

2.3 Results

2.3.1 (O2 members); bZIP family

2.3.1.1 *Identification and phylogenetic analysis*

Amino acid sequences of the orthologous bZIPs present in the sequenced genomes of the following species were selected as being representative of major monocots and dicots in plants. For example, the species have been selected in the monocots especially in the Poaceae are *Triticum aestivum*, *Hordeum vulgare*, *Zea mays*, *Sorghum bicolor*, *Oryza sativa ssp. japonica*, *Brachypodium distachyon* and *Musa acuminate* as an example of non-Poaceae. Among the dicots Arabidopsis thaliana, *Brassica rapa, Carica papaya*, C*ucumis melo, Prunus persica, Vitis vinifera, Solanum lycopersicum.* O2 maize was used as a model for identifying bZIP members because of it was studied in detail and it was considered is the central grain quality programmes in maize, so that this family will be referred as the O2 family, which is composed of 60 proteins, 44 from monocots such as two in wheat (TaSPA/TaBLZ1), in barley(HvBLZ2/HvBLZ1, HvBLZ1-like and HvRF2a-like),in rice(OsOs1, OsbZIP, OsREB, OsCRF2, OsO2 like, OsRISBZ1, OsbZIP9 and OsRIZBZ4/5), in Brachypodium (BdbLZ1, BdO2 and BdRF2a. In dicots, the closet to the monocots is AtBZO2H3, AtbZ22, AtbZIP25, AtbZIP10, AtbZIP10, AtbZIP9, and AtbZIP1. These protein sequences were used to construct a phylogenetic tree **(Fig.2.1).** Three methods were used, Neighbour-Joining (NJ), Maximum likelihood (ML) and MrBays methods, which produced nearly identical phylogenetic trees; therefore, only the MrBays tree was used for further analysis. bZIP 60 group entries inputted as out groups. Many criteria are considered, such as a posterior probability was approximately 1, which was represented on the branch.

The results showed that the O2 family was divided into two main monophyletic clusters (A and B), most monocots species represented in A group which diverged approximately 60 million years ago (Kellogg, 2001). The A group was further divided into three subgroups, A1, A2, and A3. The A1 group has the Poaceae species, which can be divided further into two subgroups; A1 contains two branches, first is (HvBLZ1, ZmO2) and their orthologous from the Panicoideae, while the second subgroup has TaSPA, HvBLZ2 and their orthologous which includes some species of Temperate cereals exclusively. HvBLZ1 was more closely to members of O2 protein, as found by Varagona *et al.* (1992), also it was found HvBLZ1 was close to

those of REB from rice (Nakase *et al.,* 1997) and OHP1 from maize (Pysh *et al.,* 1993), than to those of ZmO2 or TaSPA (Hartings *et al.,* 1989; Schmidt *et al.,* 1992; Albani *et al.,* 1997). The second group (A2) has AtBZO2H3, which is very close to the previous group, Parsley CRF2, Solanaceae CRF2 and others of the dicot species The last group is A3, which includes all the dicots of Arabidopsis such as AtBZIP10, AtbZIP25 and AtBZO2H1 diverging about 112-156 MYA (Yang *et al.,* 1999). The second cluster is **B**, which contains monocots such as ZmBZO2H2, ZmBZO2H1, OsRISBZ45, OsRISBZ5 and ZmCRF2, and dicots such as AtbZIP9. This means group A1 is specific to Poaceae. However, Arabidopsis and other dicots represent as an individual distinct clade; supporting the earlier analyses by Vincentz *et al.,* (2003). More recent duplication within the Poaceae specifically, gave rise to the TaBLZ1 and TaSPA lineages. Also, these gene duplication events lead to the diversification in the function of the bZIP factors (Walsh 1995; Wendel *et al.,* 2000).

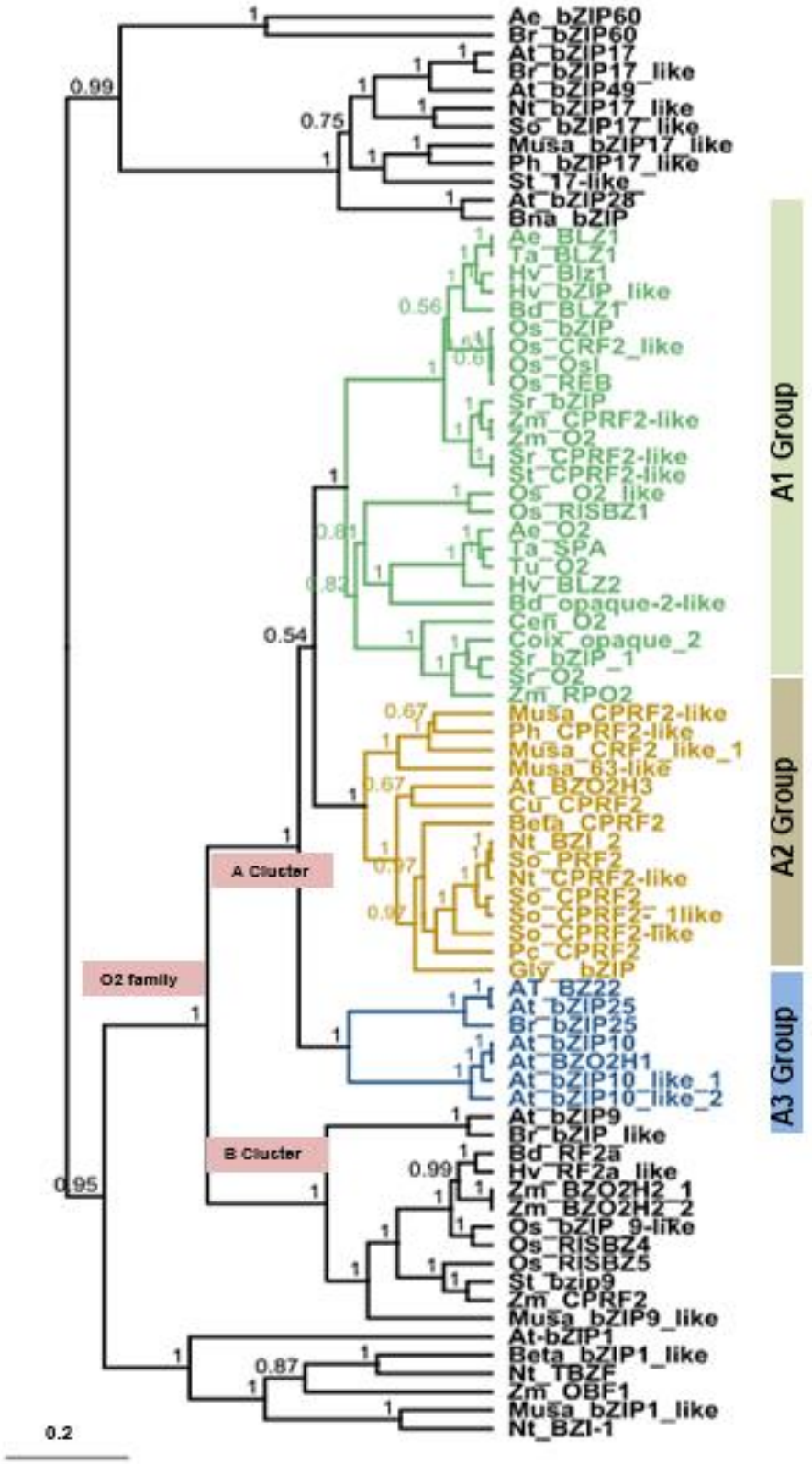


Figure 2.1 Phylogenetic relationship among the O2-like proteins based on the whole sequence of proteins. The unrooted tree was generated using MAFTT alignment tool, MarBays method.

2.3.1.2 *Characterization of bZIP genes*

The structure of bZIP genes has been studied in different species (Nijhawan *et al.,* 2008). In this study, I examined the numbers and the organization of exons and introns of O2 members (TaSPA/HvBLZ2) (TaBLZ1/HvBLZ1) in wheat, barley, maize and Brachypodium. The result showed the *TaSPA* gene has six exons and five introns and encodes a protein of 400 amino acid residues. In addition, *HvBLZ2* barley has a similar number of exons and distribution; also, it encoded 400 amino acids. *ZmO2* maize has five exons and encoded 441amino acids. In addition, the orthologous of *TaSPA* in Brachypodium has a similar structure to *TaSPA* wheat. Additionally, the *TaBLZ1* gene showed similarity to its paralogous, *TaSPA* or *HvBLZ2*, in terms of the number of exons, but it appeared a difference in their distribution especially *HvBLZ1*. However, *BLZ1* in Brachypodium has only five exons and four introns. Generally, there are differences in the size of introns, but they showed equivalent positions with regards to those in the *TaSPA* and *HvBLZ2* genes. **(Fig.2.2).**

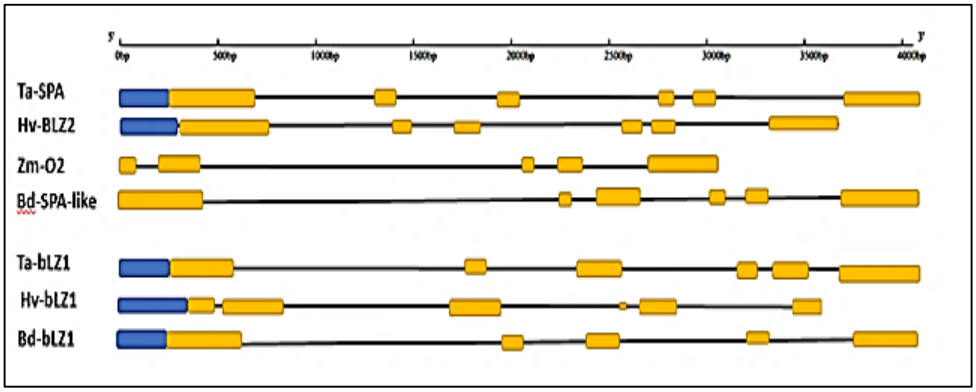


Figure 2.2 The schematic illustrates the numbers and the distributions of exons and introns for TaSPA/HvBLZ2 and TaBLZ1 and their orthologous in different species.

The yellow colour indicates exon; the black line indicates the introns, while the blue colour indicates the 5' UTRs. The scale mention to the length of the genes.

Based on the intron presence, position, and splicing phase, all O2 members have only one position of an intron within the basic region of the bZIP domain. This pattern has one intron in phase (1) except ZmO2 and MusaCPRF2 that have phase (2) **(Fig.2.3).** -However, the introns presented in the basic region had a splicing phase 1 (P1) with the exception of ZmO2 and MusaCPRF2. Overall, the positions of the introns reflect the phylogenetic topology of these bZIP family

genes. The splicing phase has been well conserved during the course of the evolution of the bZIP genes.

It has been reported that the importance of the intron patterns within the basic and the hinge region are very important for their functional evolution due to different exon splicing patterns in these regions. In plant species such as rice, maize and Arabidopsis, the patterns of those motifs exhibited regular conservation and diversity (Nijhawan *et al.*, 2008; Wei *et al.*, 2012). Furthermore, it was suggested that the sites of insertion introns sites especially in the conserved domain could be important for the evolution and divergence of function in the gene family (Liu *et al.,* 2014). In *B. distachyon,* the splicing phase remained conserved during the evolution of BdbZIP family genes, which were very consistent with those in rice (Nijhawan *et al.,* 2008).

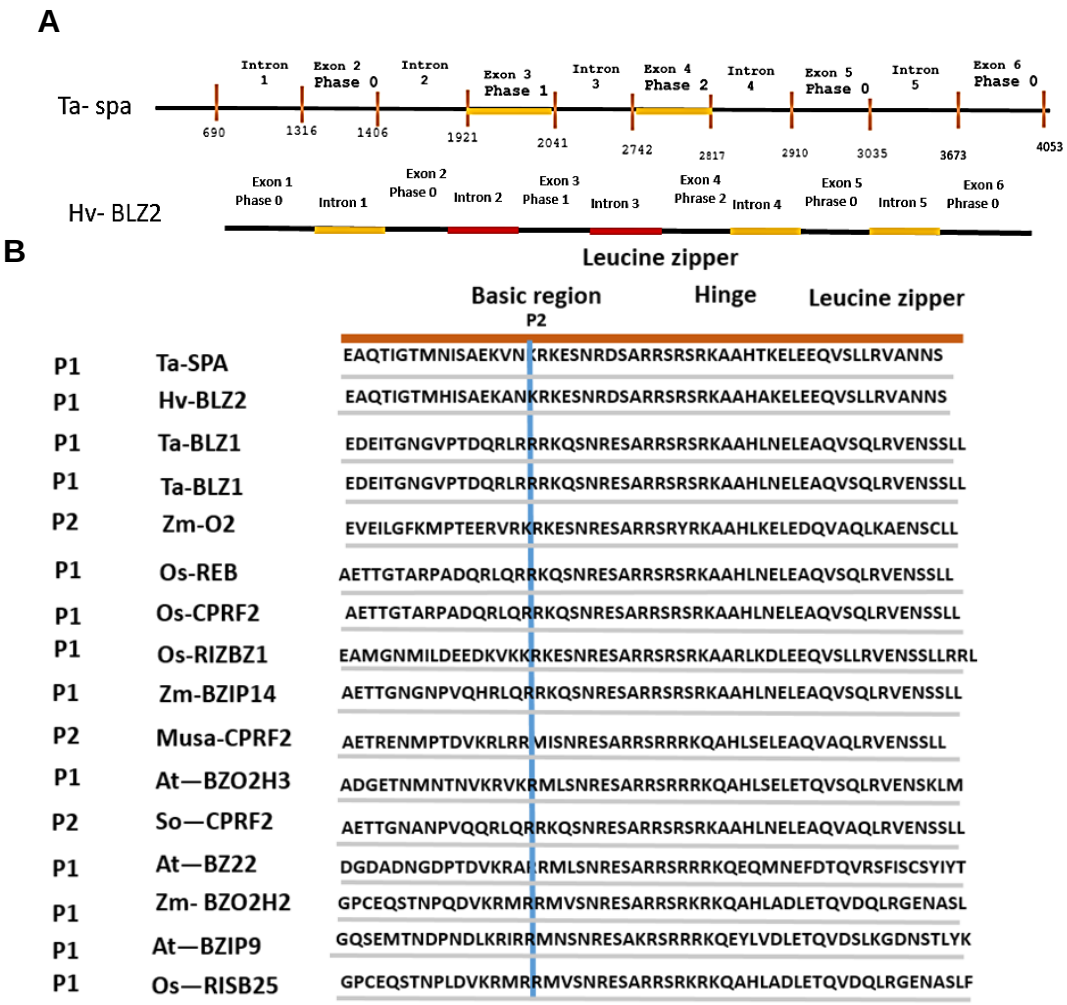


Figure 2.3 A. The schematic illustrates the number of exons (black line) and intron positions (yellow line) of both orthologous genes (TaSPA, HvBLZ2) and the position of introns within the exons.

B. Intron distribution within the basic region of the bZIP domains of O2 proteins. The structure of bZIP domain that contains (the basic, hinge regions and leucine zipper) was shown at the top with vertical blue line indicating the position of the intron P1 and P2 indicate the splicing phases of the basic region of the bZIP domains.

2.3.1.3 *Analysis of the sequences of bZIP domain of O2 family*

2.3.1.3.1 *DNA-binding domain features*

To analyze the O2 proteins sequences, each group of O2 was aligned separately to identify the bZIP domain and conserved motifs **(Fig. 2.4); (Fig. 2.5).** There are a number of residues within each group which are highly conserved across the species, which indicates their importance in the function of these proteins and the ability to bind the DNA. For example, the arginine residue in the basic region is very highly conserved in all bZIP proteins (Liu *et al.,* 1999). In addition, the number of Arginine (R) and Leucine (L) replications in the domain may be involved in the dimerization. In addition, the spatial arrangement of these amino acids in the domains is very important for specificity; this family has one bZIP binding domain, which has the specificity to interact with the DNA of the targeted genes (Liu *et al.,* 1999).

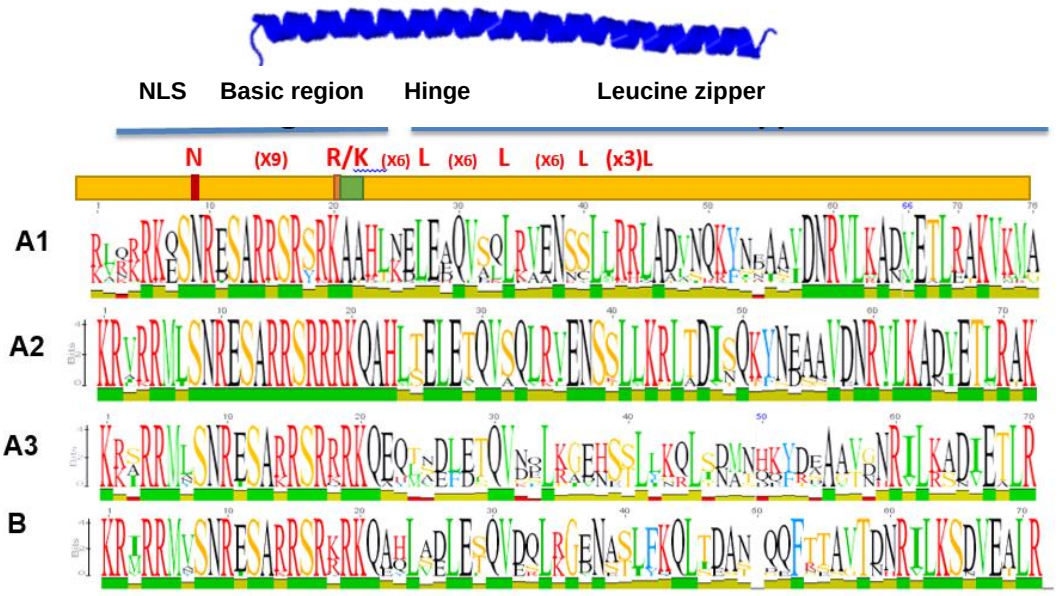
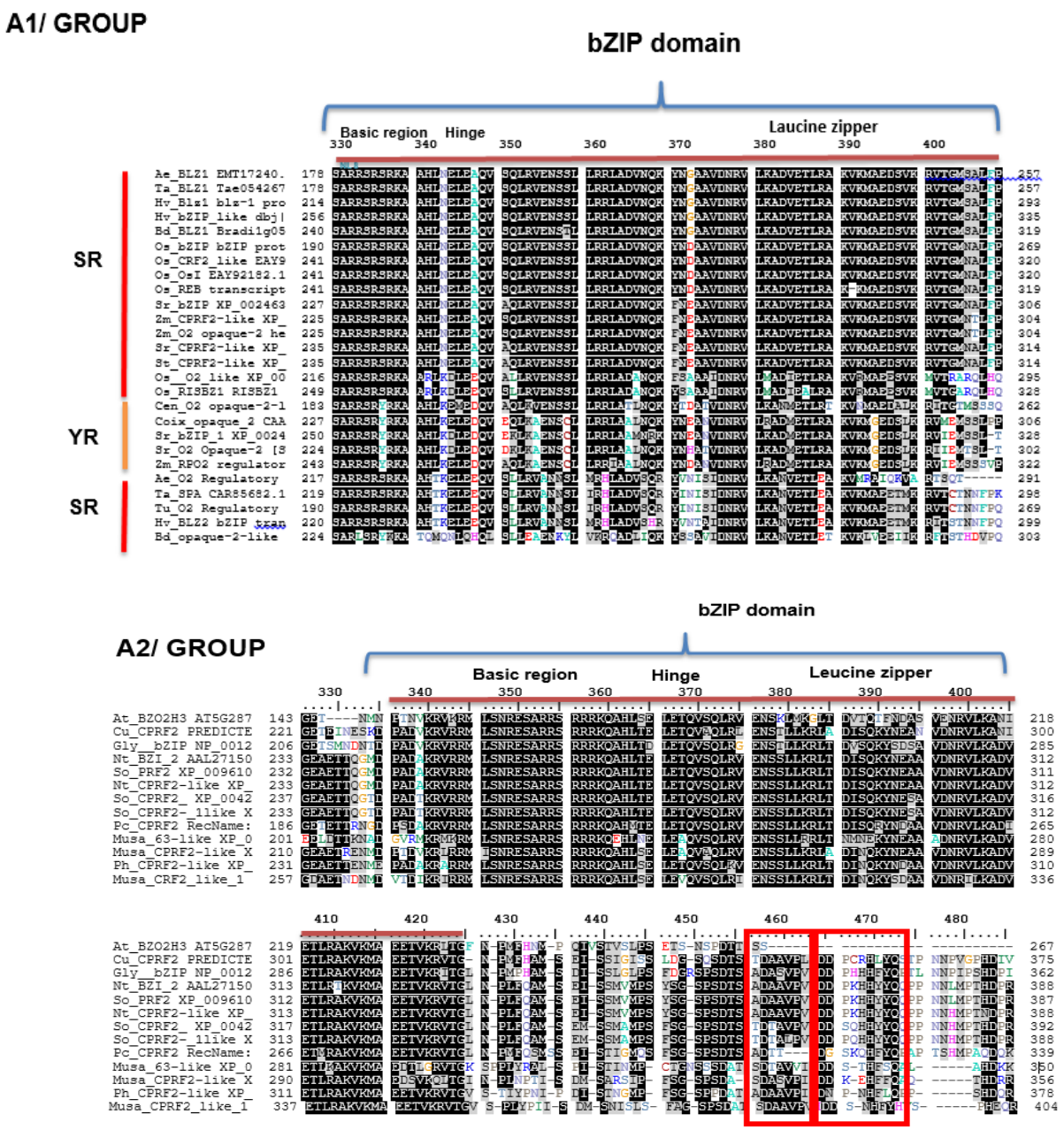


Figure 2.4 The 3D structure of bZIP domain by using Phyre 2, four sequence logo of bZIP domains for for groups of O2 family was determined by using MEME comparing. The dark green line represents the conserved amino acids.


A1/ GROUP
bZIP domain
Basic region
Hinge
Laucine zipper
SR
YR
SR
A2/ GROUP
bZIP domain
Basic region
Hinge
Leucine zipper

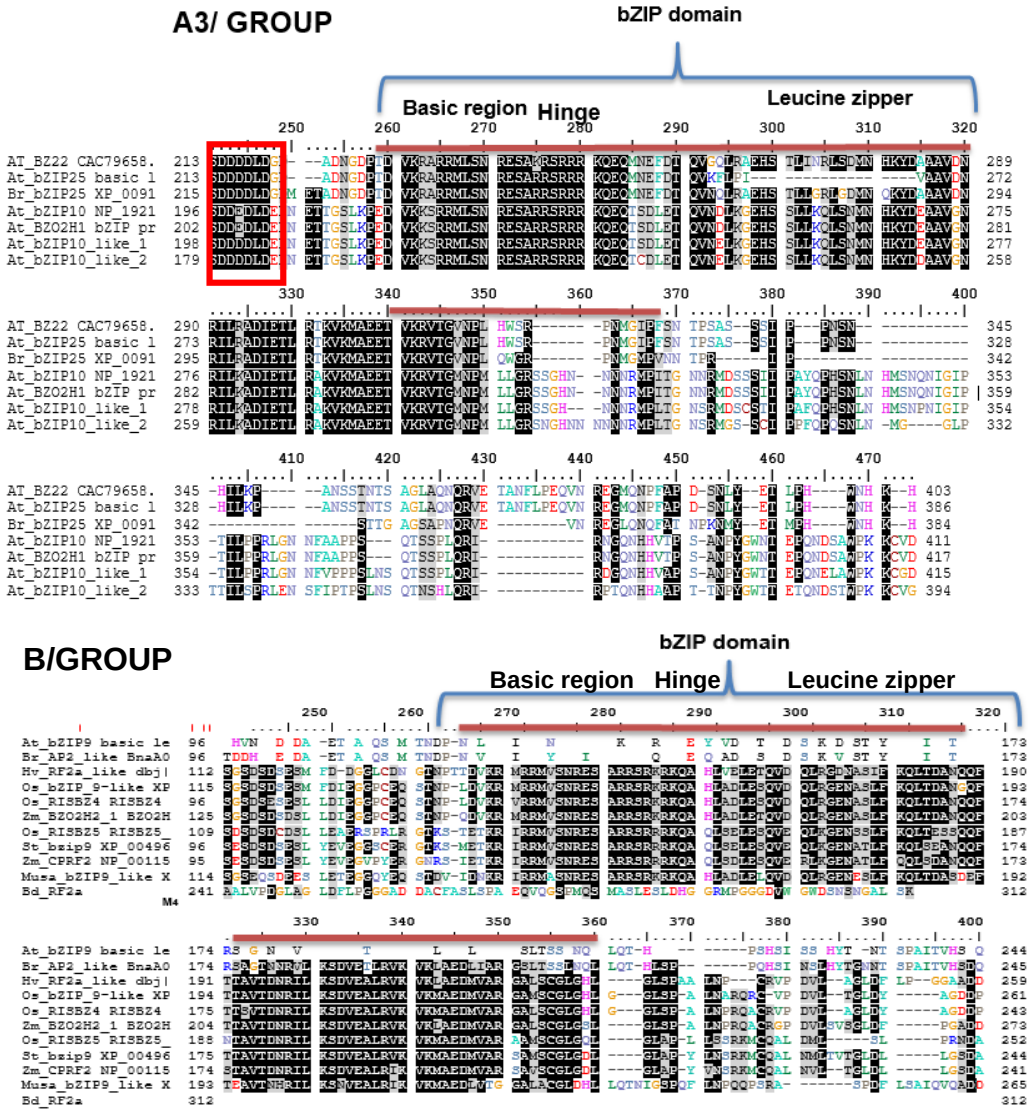


Figure 2.5 Alignment the full length sequences of proteins of 4 groups of O2 family, conserved motifs were boxed in red line, the bZIP domain upper red line.

To characterize the structure of the bZIP domain in detail, it was divided into several sub-regions, which are:

A) Basic region (BR)

The basic region consists of about 16 amino acid residues which are characterized by the presence of an invariant N-x7-R/K motif in the N-terminal boundary of the leucine zipper. It contains three main regions:

1.Basic region. The N-terminal boundary of the leucine zipper is approximately 22 amino acids in length. It is very highly conserved between species and is considered one of the two most highly conserved sites in the basic region of the bZIP domain, the conventional pattern of invariant residues was N-×…-R/K (termed the "N-R/K" bZIP class in this study) **(Fig. 2.4).**

2.NLS. is a short sequence of the protein which is characterized by arginine (R) and Lysine (K); all members of the O2-like family have two groups of NLS in the N-terminus of the basic region separated by non-conserved residues**.(Fig.2.6).**

3.Hinge*.* The second region of around nine amino acids is located between the basic region and leucine zipper (Williams *et al.,* 1992). Although it is very highly conserved between all Poaceae members, it shows differences outside the Poaceae. From the alignment the significant differences between the groups were:

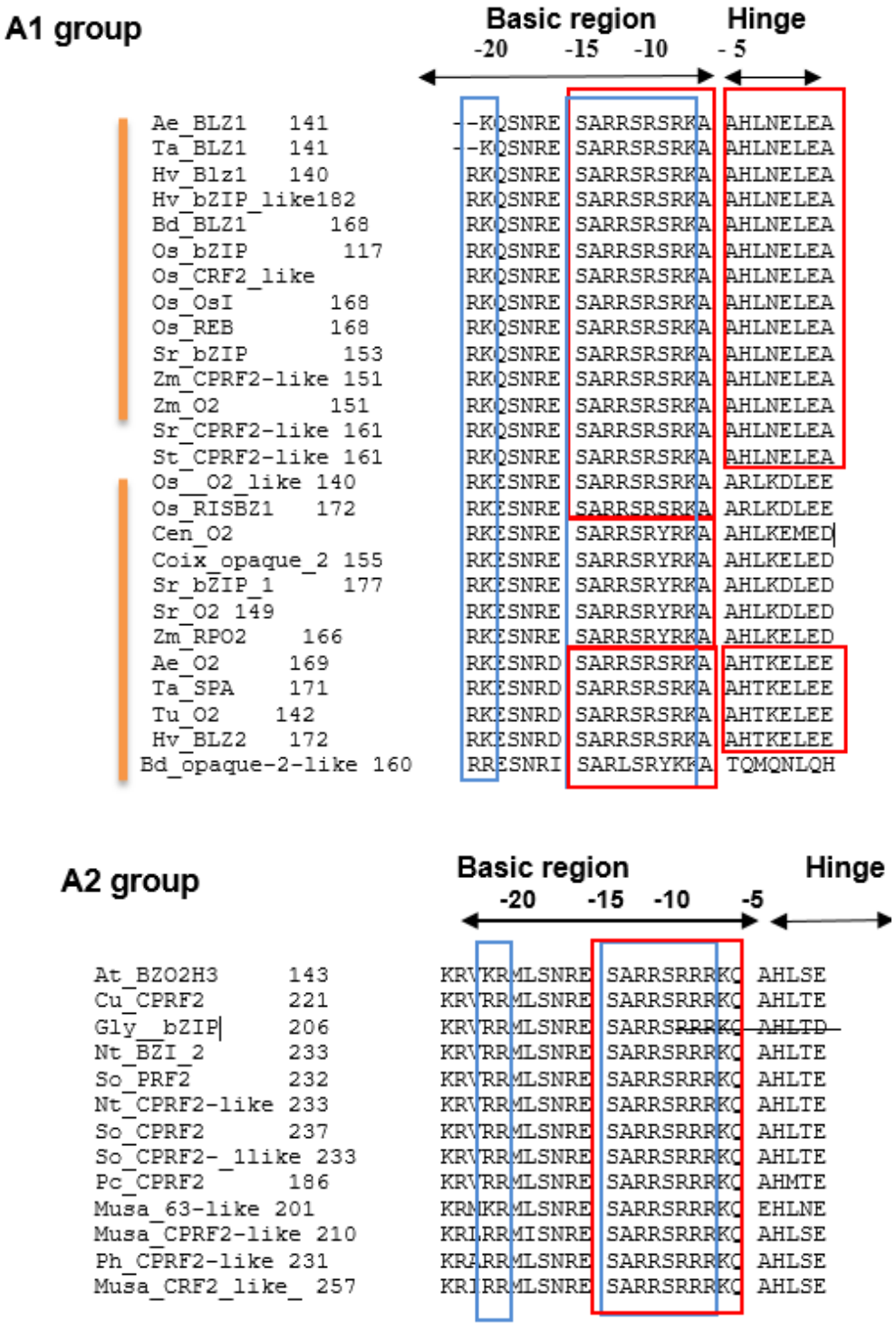
***A1 Group*:** there were replacements in the amino acid in the basic region. As a result, this group can be divided into three small subgroups: the first includes HvBlZ1, BdBLZ1, OsBzip, OsOSI, OsCRF2-like, OsREB, SrbZIP, ZmCPRF2-like, ZmO2, SrCPRF2-like, St CPRF2-like, except TaBLZ1which is different only in one residue which was represented by (RKQSNRE) The second subgroup contains OsO2, OsRISBZ1, CenO2, CoixOPEQUE 2, Sr bZIP/1, SrO2, ZmPro2, was represented by (RKESNRE), whilst the last subgroup has TaSPA, HvBLZ2 and AeO2, and was represented by (RKESNRI). In the 337 position, most of them have SR, while a few have YR **(Fig.2.6).**

A2 Group: **KRVKRM** in AtBZO2H3 and Musa 63-like, **KRVRRM** in others, while MusaCPRF2 has **KRLKRM** or **KRARRM, KRIRRM** in PhCPRF2, MusaCPRF2-like.

A3 Group: the basic region is identical in this group, with **KRARRM** in AtBZ22, AtbZIP25, and Br-bZIP25 while others have **VKKSRRM,** but both these amino acids are non-charged.

Cluster B, was divided into two groups, the first includes At-bZIP9-like, Br-bZIP-like, OsRISBZ5, ZmCORF2, StbZIP9 and Musa-Bzip9 represented by **KIRRMVSNRESAR AQLSELE**, while the second group has ZmBZO2H2/1, OsRISBZ4, OsBzip9-like and HvRF2a, represented by **KMRRMVSNRESAR AHLVELE (Fig.2.6).**

Furthermore, the significant difference between the members of the O2-like family is the sequence in the hinge region, which is more specific in the A1 group than other groups. Vettore, (1998) that the bZIP domain of O2-like proteins in maize and sorghum are highly conserved, whereas considerable divergence can be observed at the C-terminus, supports these results.


A1 group
Basic region
Hinge
-20
-15
-10
- 5
Ae_BLZ1 141 --KQSNRE SARRSRSRKA AHLNELEA
Ta_BLZ1 141 --KQSNRE SARRSRSRKA AHLNELEA
Hv_Blz1 140 RKQSNRE SARRSRSRKA AHLNELEA
Hv_bZIP_like182 RKQSNRE SARRSRSRKA AHLNELEA
Bd_BLZ1 168 RKQSNRE SARRSRSRKA AHLNELEA
Os_bZIP 117 RKQSNRE SARRSRSRKA AHLNELEA
Os_CRF2_like RKQSNRE SARRSRSRKA AHLNELEA
Os_OsI 168 RKQSNRE SARRSRSRKA AHLNELEA
Os_REB 168 RKQSNRE SARRSRSRKA AHLNELEA
Sr_bZIP 153 RKQSNRE SARRSRSRKA AHLNELEA
Zm_CPRF2-like 151 RKQSNRE SARRSRSRKA AHLNELEA
Zm_O2 151 RKQSNRE SARRSRSRKA AHLNELEA
Sr_CPRF2-like 161 RKQSNRE SARRSRSRKA AHLNELEA
St_CPRF2-like 161 RKQSNRE SARRSRSRKA AHLNELEA
Os__O2_like 140 RKESNRE SARRSRSRKA ARLKDLEE
Os_RISBZ1 172 RKESNRE SARRSRSRKA ARLKDLEE
Cen_O2 RKESNRE SARRSRYRKA AHLKEMED
Coix_opaque_2 155 RKESNRE SARRSRYRKA AHLKELED
Sr_bZIP_1 177 RKESNRE SARRSRYRKA AHLKDLED
Sr_O2 149 RKESNRE SARRSRYRKA AHLKDLED
Zm_RPO2 166 RKESNRE SARRSRYRKA AHLKELED
Ae_O2 169 RKESNRD SARRSRSRKA AHTKELEE
Ta_SPA 171 RKESNRD SARRSRSRKA AHTKELEE
Tu_O2 142 RKESNRD SARRSRSRKA AHTKELEE
Hv_BLZ2 172 RKESNRD SARRSRSRKA AHTKELEE
Bd_opaque-2-like 160 RRESNRI SARLSRYKKA TQMQNLQH
A2 group
Basic region
Hinge
-20
-15
-10
-5
At_BZO2H3 143 KRVKRMLSNRE SARRSRRRKQ AHLSE
Cu_CPRF2 221 KRVRRMLSNRE SARRSRRRKQ AHLTE
Gly__bZIP 206 KRVRRMLSNRE SARRSRRRKQ AHLTD
Nt_BZI_2 233 KRVRRMLSNRE SARRSRRRKQ AHLTE
So_PRF2 232 KRVRRMLSNRE SARRSRRRKQ AHLTE
Nt_CPRF2-like 233 KRVRRMLSNRE SARRSRRRKQ AHLTE
So_CPRF2 237 KRVRRMLSNRE SARRSRRRKQ AHLTE
So_CPRF2-_llike 233 KRVRRMLSNRE SARRSRRRKQ AHLTE
Pc_CPRF2 186 KRVRRMLSNRE SARRSRRRKQ AHMTE
Musa_63-like 201 KRMKRMLSNRE SARRSRRRKQ EHLNE
Musa_CPRF2-like 210 KRLRRMISNRE SARRSRRRKQ AHLSE
Ph_CPRF2-like 231 KRARRMLSNRE SARRSRRRKQ AHLSE
Musa_CRF2_like_ 257 KRIRRMLSNRE SARRSRRRKQ AHLSE

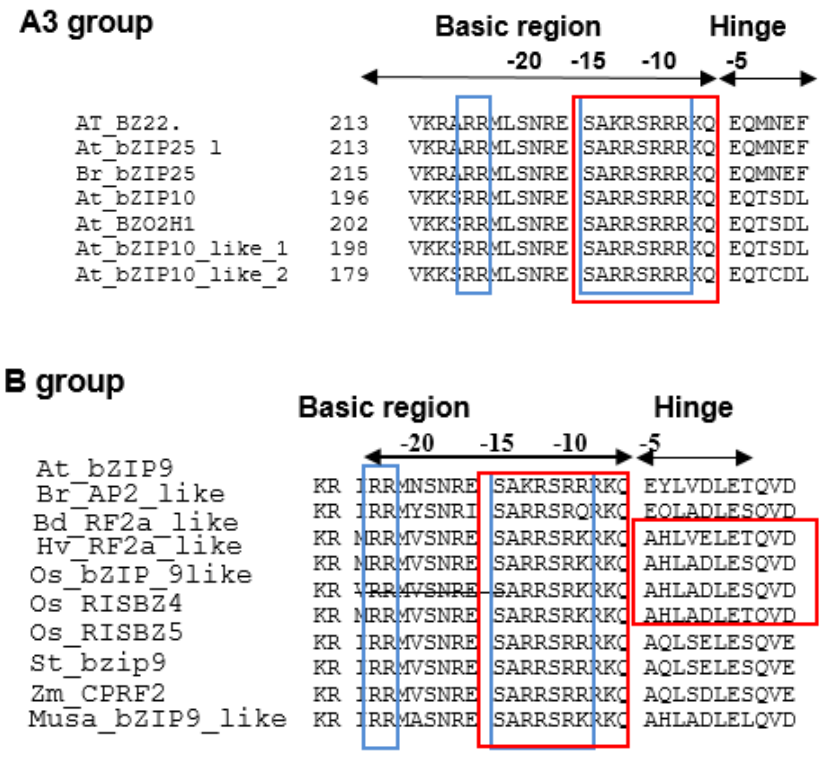


Figure 2.6 Alignment of basic and hinge regions of bZIP domain of O2 members.

The bZIP proteins are classified into 4 groups according to their amino acid sequences in basic and hinge regions. The conserved amino acids between the whole groups are boxed in blue line, while the red boxes indicate the conserved a.a. within the members of each group. Amino acid numbering is according to Suckow *et al.,* (1993).

2.3.1.3.2 *Nuclear localization signals (NLS).*

Most TFs in plants have an NLS is characterized by the presence of arginine (R) and Lysine (K). Nuclear localization signals (NLSs) are stretches of residues within a protein that has importance for regulating the nuclear import of the protein (Ba *et al.,* 2009). In ZmO2 has two NLSs, which have function independently. By comparing between the members of the O2 family, there are two NLS motifs in the N-terminus of the basic region, as separated by non-conserved residues (**Fig.2.7).**

From the alignment of the bZIP proteins, the NLS sequence is present in two sites as a short basic region of amino acids. The first NLS in the N-terminus of the basic region varies between the groups, while the second NLS is located within the basic DNA binding domain, which is highly conserved. For example, there is a difference in the NLS sequences between members of the **A group**. The first group, TaSPA, HvBLZ2 bZIP, and OsRISBZ1, have **NKRK**, but others, such as BdBLZ1, HvBLZ1, have **RRRK**, whilst the last group, such as ZmO2, have **QRRK**. **A2 group** is substituted with V **(KRVKR),** whilst **A3 and B groups** have **(KRARR).**

Figure 2.7 Nuclear localization signals (NLS) in the basic region, comparing the sequence of NLS between members of O2 family of different groups. (DB: double partiate).

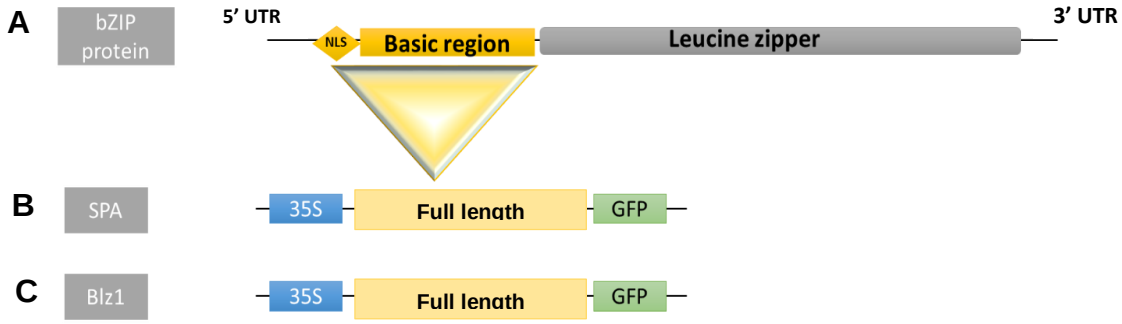


Figure 2.8 Structure of bZIP protein, the NLS which were amplified of Ta SPA and Ta BLZ1.

A**.** Diagram of the bZIP domain; B.Full-length of TaSPA was constructed with GFP fusion under control of promoter 35S; C. Full-length of TaBLZ1 was constructed with GFP fusion under control of promoter 35S, gateway-cloning technique (Invitrogen) used to create constructs.

To confirm the function of NLS in TaSPA/ HvBLZ2 and TaBLZ1 wheat, and barley were cloned this region, and then each construct of TaSPA/HvBLZ2 in wheat and barley and BLZ1 in wheat and barley were shown in **(Fig. 2. 8),** were infiltrated into *Nicotiana benthamiana* leaves were observed under the epifluorescence microscope after 2 days to determine whether the GFP marker was expressed in the nucleus and the cytoplasm. **(Fig.2.9)** shows a GFP-GFP was used as a negative control, which shows a signal in the cytoplasm rather than the nucleus. A striking difference can be found among two paralogous either TaSPA/ TaBLZ1 wheat or HvBLZ2/ HvBLZ1 barley. In this experiment, TaBLZ1 shows signal in cytoplasm and nucleus, while TaSPA/HvBLZ2 do not have any signal in the nucleus but there is a faint signal in the cytoplasm **(Fig.2.10).**

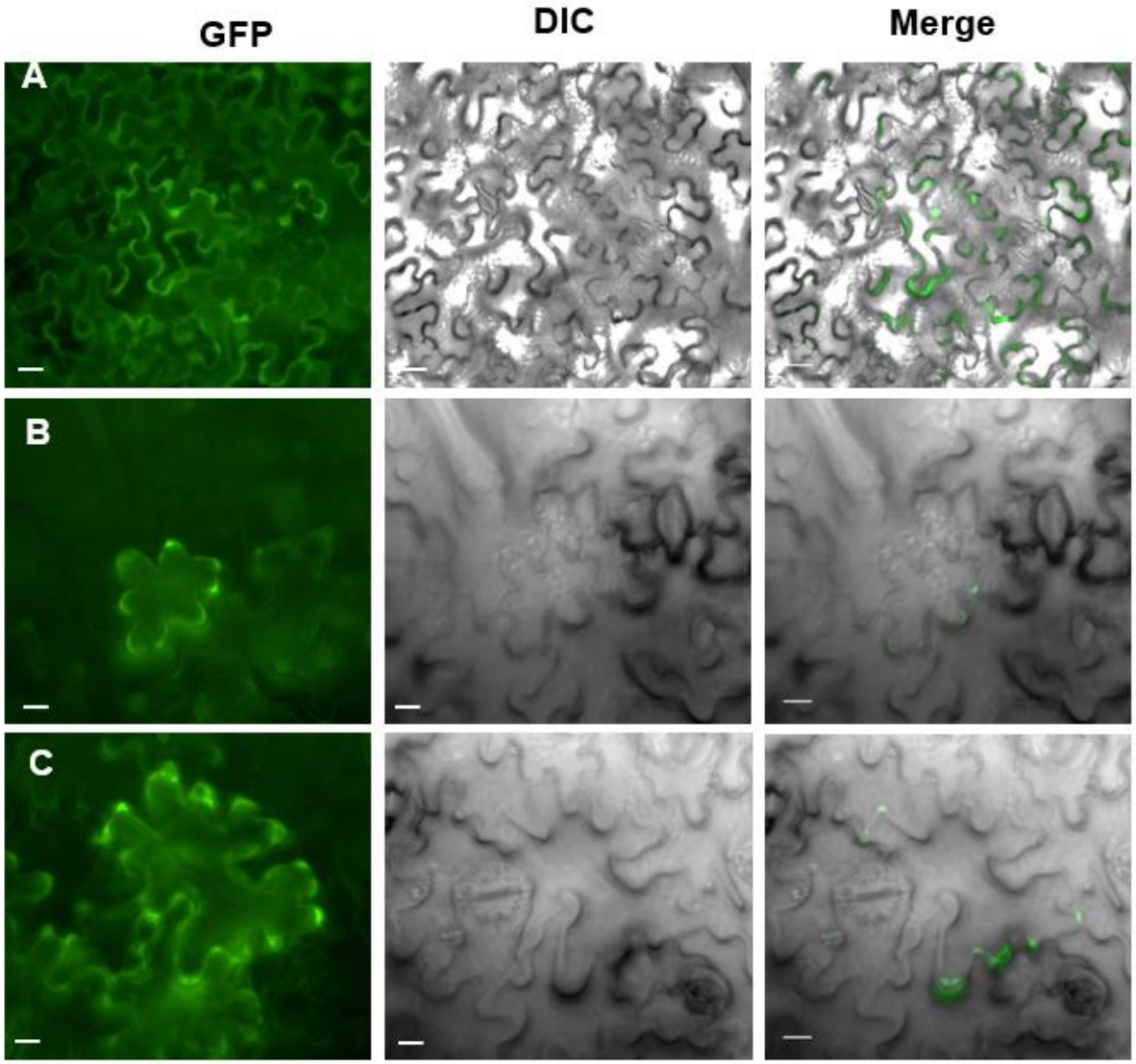


Figure 2.9 Subcellular localization of TaSPA/HvBLZ2 proteins variants.

Fluorescence microscopy analysis shows nuclear protein localization *N. benthamiana* leaves. A: promCaMV 35S:2XGFP (control), B: promCaMV 35S: TaSPA-GFP, C: promCaMV 35S:HvBLZ2-GFP.

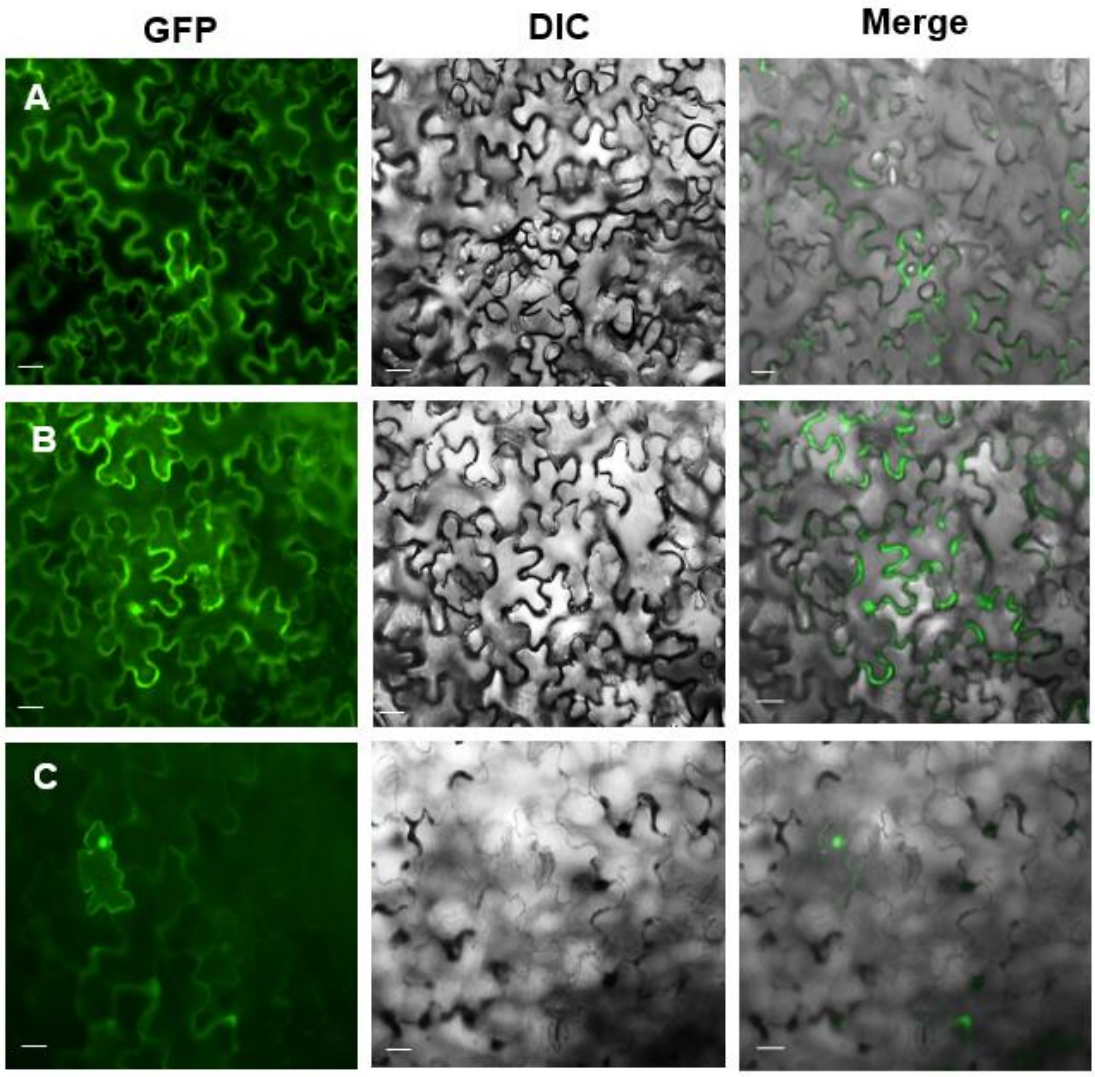


Figure 2.10 Nuclear localisation of Ta/HvBLZ1 proteins variants.

Fluorescence microscopy analysis shows nuclear protein localization *N. benthamiana* leaves. A: promCaMV 35S:2XGFP (control), B: promCaMV 35S:TaBLZ1-GFP, C: promCaMV 35S:HvBLZ1-GFP.

2.3.1.3.3 *The leucine zipper*

From the number of heptad repeats, the length of the leucine zipper can be determined and identified the charged amino acids are present in these positions which could play a significant role in determining the stability and specificity of the dimerization between the bZIP proteins **(Fig. 2.11)**.

For example, amino acids at **(a and d)** positions are on the same surface of the α-helices and are hydrophobic, which is essential to the formation of a hydrophobic core. They interact with the complementary positions **(a' and d')** in the opposite monomer and create a hydrophobicity, which is important to the overall stability of the dimer **(Fig.2.12).** Positions **(e and g)** usually have charged amino acids, which mediate the specificity of the dimerization as well as govern overall stability (Nijhawan *et al.,* 2008).

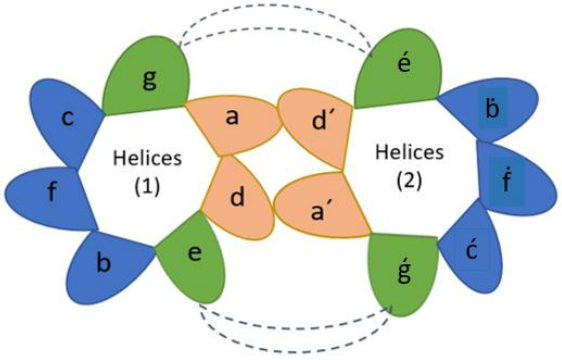


Figure 2.11 Outline of a section for interacting of two helices via two leucine zippers.

The diagram shows the position of the amino acids in (*a* and *d*) interface of the helices, which are the hydrophobic core, which are indicated with the pink color. Charged residues in positions *e* and *g* generate electrostatic forces, represented by green color. Others which are hydrophilic surface is formed by the amino acids in positions *b*, *c*, and *f*. (Adopted from, adaLlorca *et al.,* 2014).

2.3.1.3.4 *Leucine zipper length.*

The length of the zipper is usually quite variable between species **(Fig.2.12)**, but in this group the length of leucine zipper was conserved between species. Glycine presents in the C-terminus, which is considered a terminator of the α-helix. From these points, all these proteins have around eight to nine heptads, which was the longest leucine zipper in the bZIP superfamily (Jakoby *et al.,* 2002). The difference between the shorter and longer leucine zippers is the longer is better in the stability of dimerization because of the dimerization with short leucine zipper needs optimization for stability (Vinson *et al.,* 2002).

2.3.1.3.5 *Predicted dimerization in bZIP TF*

As described above, the leucine zipper contains heptad repeats, which is responsible for controlling the dimerization. To make this analysis easy and clear, the heptads were given different colors depending on the type of amino acids involved in the specificity of the dimerization.

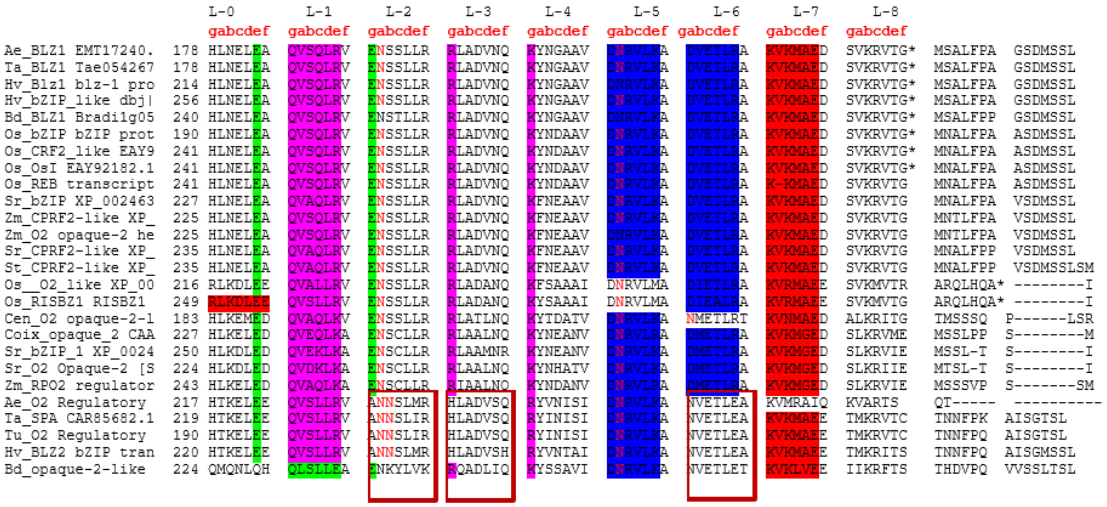
In **(Figure 2.12),** all the colors shown are classified into four groups according to electrostatic charges in the g and e positions. A blue box means that the **g** is acidic, and the **e** is the basic amino acid; for example, (E-xxxx-R) (E-xxxx-K) (D-xxxx-R) and (D-xxxx-K). Therefore, the interaction between two helices with two blue boxes could be the formation either the homodimer or heterodimer, depending on if the helices are identical or otherwise. An orange heptad indicates the **g** and **e** positions are basic and acidic, respectively, as (R-xxxx-E) (K-xxxx-E). In the same way, two orange boxes in two helices will interact. However, current knowledge regarding the dimerization between identical boxes is extremely limited. Those boxes, with their similar charges, would disfavor dimerization due to the consequent electrostatic repulsion. In **e** and **g** positions, which have the same charge, for instance (acidic-xxxx-acidic) (E-xxxx-E, E-xxxx-D, E-xxxx-Q, Q-xxxx-E), are assigned a green colour, and are known repulsive acidic pairs. Also, basic-xxxx-basic (K-xxxx-K) (R-xxxx-K) (Q-xxxx-K) (R-xxxx-Q) (K-xxxx-Q) repulsive basic pairs are coloured pink. If the two helices have two acidic-xxxx-acidic or basic-xxxx-basic boxes, this will prevent dimerization. In the opposite situation, when one of the

helices has one repulsive acidic heptad and another has a repulsive basic heptad, then they can form a heterodimer, and are referred to as being mutually complementary.

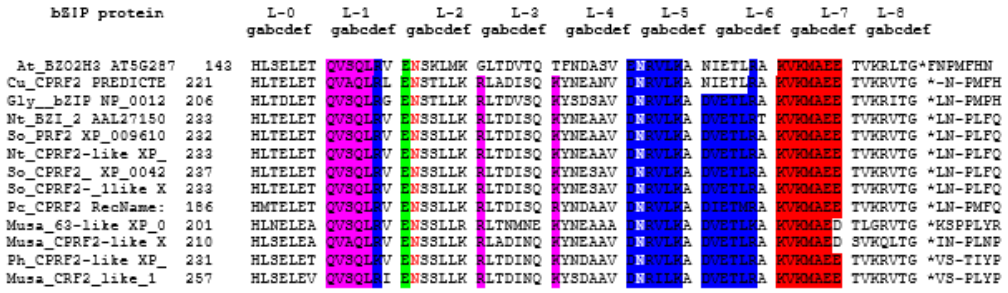
This combination of either acidic or basic repulsive heptad is often abundant in proteins such as AtbZIP10, AtBZO2H1, AtbZIP10-like, and AtbZIP10-like 1, which have (acidic–xxxx-acidic) at the 0 heptad, is a repulsive acidic heptad, while the 1st heptad, (basic-xxxx-basic), is a repulsive basic heptad. However, these proteins share other groups with the same number of the attractive acidic heptads in the 5th and 6th heptads, and a basic heptad at the 7th heptad. This means that the O2-like family members share the same number and distribution of attractive and repulsive heptads, as might be noted from (**Fig. 2.12).** However, there is a slight difference in some species of Arabidopsis and some OsbZIP in terms of having different heptads in the (acidic–xxxx-acidic), which indicates the possibility of hetero-dimerization that might otherwise be absent in these proteins.

Leucine zipper

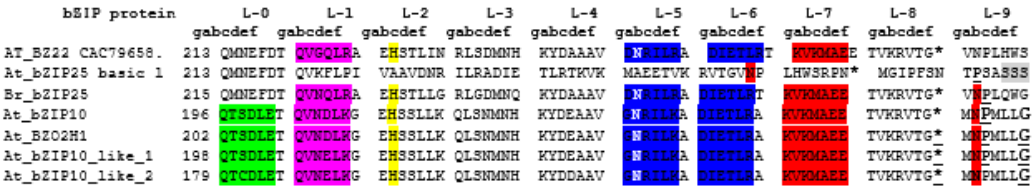
A1 GROUP



A2 GRPOUP



A3 GROUP



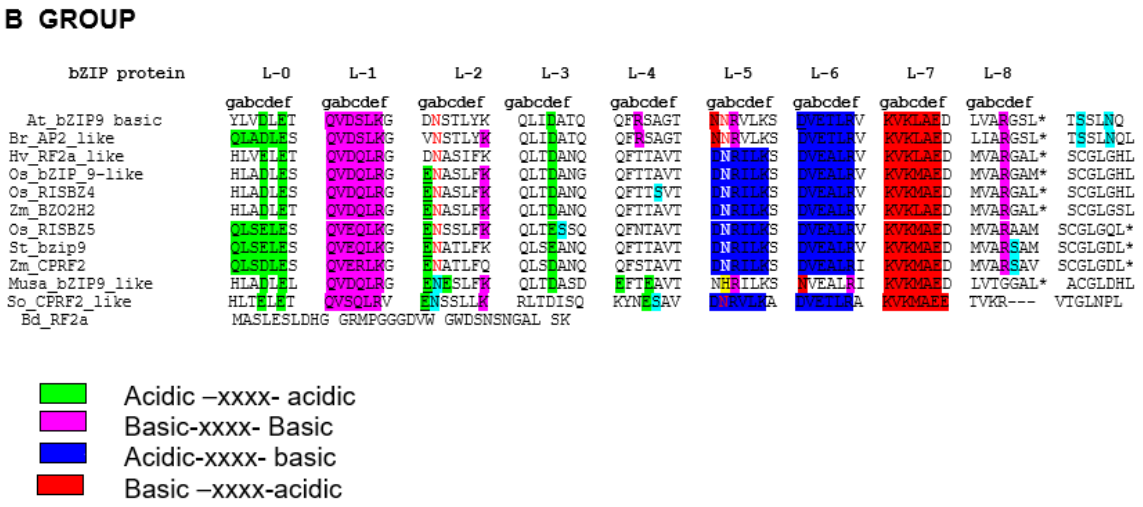


Figure 2.12 Amino acid sequence alignment of the leucine zipper region of 26 of bZIP proteins.

The bZIP proteins are classified into groups with similar predicted dimerization properties. The leucine zipper region is divided into heptads (gabcdef) to visualize the potential g↔e′ pairs. A colour code is used to highlight the amino acids predicted to regulate both the dimerization specificity and stability. Four colours are used to differentiate between different g↔e′ pairs. Attractive basic-acidic pairs (R↔E and K↔E) are coloured orange, attractive acidic-basic pairs (E↔R, E↔K, D↔R, and D↔K) are blue, repulsive basic pairs (K↔K, R↔K, R↔Q, Q↔K, and K↔Q) are pink and repulsive acidic pairs (E↔E, E↔D, E↔Q, and Q↔E) are green. If only one of the two amino acids in the g↔e′ pair is charged, the residue is coloured pink for basic(R/K) and green for acidic.. Asparagines at a position are coloured red. The predicted C-terminal boundary is denoted by asterisk.

Table 2.1 Characteristic features of the O2-like family classified into four groups according to with similar predicted dimerization specificity.

No. of group	Group	No. of member in the group	proteins	Heptad with N/K at a position	Length in heptads	Comments
A1	Subgroup –A1-a	14 in two groups; one of Triticeae, another one of Panicoideae.	HvBLZ1, OsCRF2_LIKE OsREB ZmO2	2-5	8	Ns in 2nd and 5th heptad (**a**) position supporting homo-dimerization, and presence of 2 attractive acidic in *g*↔*e′* pairs in 5th and 6th heptads. Also, attractive basic in 7th heptad. Many residues or incomplete pairs which could support the interaction with other proteins. One basic repulsive basic in 1st heptad.
	Subgroup- A1-B	12 in two groups: - Triticeae - Panicoideae	TaSPA HvBLZ2 OsRISBZ1	2-5	8	As described above, but the attractive *g*↔*e′* pairs in 6th heptad was absent, also, there are some of these residuces are absent in TaSPA, HvBLZ2 in the 2nd, 3rd. heptads. Both (RISBZ4) and (RISBZ5) have been shown to hetero-dimerize with (RISBZ1) (Onodera *et al.,* 2001).
A2	A2	15 contains dicots+monocots non-Poaceae	AtBZO2H3 NtCPRF2_like PcCPRF2_LIKE	2-5	8	As described in the first subgroup (A1-a)
A3	A3	7(Arabidopsis)	AtBzip25 AtBzip10 AtBZO2H1	5	8	Ns in 5th heptad *a* position, H replace in 2nd heptad. Presence attractive acidic in the 5th and 6th heptads, attractive basic in the 7th heptad only. There is repulsive basic in the 1st heptad, also, some replusive acidic in 0 heptad, which contribute to hetero-dimerization.
B	B	12 from dicots+ monocots non-Poaceae	ZmCPRF2, OsRISBZ5, OsRISBZ4, AtbZIP9	2-5	8	As described in above in A3 group

To analyze the contribution of change residues in governing dimerization properties of O2 members, the frequency of either attractive or repulsive amino acids were calculated corresponding histogram is represented in **(Fig. 2.13).** In addition, the occurrence of the type of amino acids has been calculated in four positions (a, g, e and d) in each group **(Fig. 2.14).**

In the **(a)** position, asn. (N) presents in heptads 2,5 in most members of O2 except A3, which has asn (N) in the 2nd heptad only. It is likely that asn. (N) effects the formation of an optical interhelical bridge between the amino acids (Vinson *et al.,* 2006). The occurrence of asn (N) has been suggested to limit the dimerization of the leucine zipper (Ellenberg *et al.,* 1992; König *et al.,* 1993). In addition, asn (N) prefers an interhelical interaction with another asn (N) in the **(a)** position of a partner protein, and will not interact efficiently with an aliphatic amino acid (Merier and Gruissem, 1994; O'Shea *et al.,* 1991). Furthermore, leu (L) presents in this position, which is at its highest in the A1 subgroup at around 50%, whilst other subgroups have a 25% occurrence of leu (L), which plays play role to form heterodimerization with a wide range of amino acids (Acharya *et al.,* 2002) (**Table 2.1).**

In addition, one of the most important positions, which has a pivotal role in dimerization, is the **(d)** position of each heptad. The frequency of leu (L) presenting in the **(d)** position, which is responsible for dimer stability in the bZIP, was found to be around 100% in A1, which is significantly higher than other groups. A2 has the lowest percentage of leu at around 5% in the **(d)** position (Nijhawan *et al.,* 2008). The frequency of leu in several heptads is 100%, which could be responsible for the stability of the longer zippers. However, leu was replaced with hydrophobic amino acids in the **(d)** positions for val (V) and ala (A), and other acids such as glu (E), arg (R), and ser (S). **(Fig. 2.14)** shows that there are other aliphatic amino acids in the **(d)** position that are more abundant in AtbZIP other groups. As a result, these replacements could lead to new binding specifications (Suckow *et al.,* 1994; Wei *et al.,* 2012). Furthermore, the **(g and e)** positions play a considerable role in the dimerization by forming attractive

electrostatic interhelical interactions between two oppositely charged amino acids (Cohen and Curran, 1990; Alber, 1992).

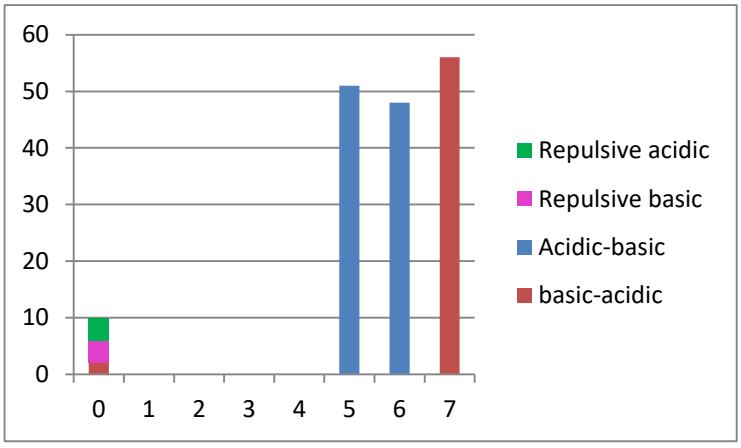


Figure 2.13 Histogram of the frequency of the attractive or repulsive g↔e pairs per heptad for all O2 proteins.

Figure 2.14 Charts illustrate the percentage and the type of amino acids which present in g, a, e, and d positions, respectively comparing between 4 groups.

2.3.1.4 *Conserved motifs outside bZIP domain*

These conserved motifs play an important role in terms of protein-protein interaction, transcriptional activity, and nuclear localization (Liu *et al.,* 1999). The additional conserved motifs may contribute to the functional diversity within the *bZIP* proteins. As the functional domain is conserved within the family or group of TFs, the conserved motifs indicate conserved functions within the same group or family.

To identify the conserved motifs outside the bZIP domain, two different methods were used:

A) Through the alignment process, these aligned proteins were used to construct the phylogenetic tree, also the results showed many conserved motifs between the members of bZIP (**Appendix 7. Fig. 7.2.1).**

There are differences between the A and B groups in terms of conserved motifs which can be used as characters to distinguish between them; furthermore, this result corresponds to the topology of proteins in the phylogenetic tree. They can be considered as informative shared derived characters that were used to classify members of the O2 family. However, most motifs are conserved among the members of the A cluster in the O2 family, but the biological functions of these motifs are mostly unknown, but most of these proteins share similar functions (Rashid *et al*., 2012). For example, the number of conserved motifs in A1 is more than other subgroups; although there is replacement within these motifs in the first group **(Table.2.2).**These substitutions divide A1 into two subgroups. The first subgroup includes TaBLZ1, which has very highly conserved motifs between its members. Furthermore, some motifs are rich in acids such as Ser (S), Asp(D), Glu(E) which thought these play a role in the transcription activation of *bZIP* genes (Vincentz *et al.,* 2003). Additionally, a set of conserved motifs which are rich with Ser/Thr (TSS) may play an important role in the transcriptional activation (Onodera *et al.,* 2001). Also, the phosphorylation sites, which are required for post-transcriptional modification to regulate nuclear translocation and DNA-binding activity (Ciceri *et al.,* 1997; Liu *et al.,* 2014). By comparing between the subgroups of O2, the members of A1 have motifs which are M1, M3 and M5 that can

participate in the transcriptional activation (Vincente-Carbajosa *et al.,* 1998; Onodera *et al.,* 2001), while A2 subgroup shares A1 subgroup in one motif which is (M1). However, A3 subgroup has different motifs, while it shares A1 in M5 is slightly different in two residues. By contrast, B has only one conserved motif and another small motif is SSI-TSA**.**

Table 2.2 The main conserved motifs between the members of O2-like family and the substitution amino acids within these motifs.

GROUP	MOTIF 1	MOTIF 2	MOTIF 3	MOTIF 4	MOTIF 5	MOTIF 6	MOTIF 7	MOTIF 8	MOTIF 9
A1									
TASPA	MEPVF	SLEE	TEWC	SVDE	EVDPVAY	HLAAVAMLRN	SSSSLE	GDMEGE	CLDHHPQ
HVBLZ2		LLEE	TELLF	SMEE	VLDPVGY	HLAAVAMWRT	SSSSWE	GDMEGE	CLDHHPR
HVBLZ1	MERVF	SVEE	CPSEWY	FLEE	VVDPVEY	DLAAVAMWARA	SSSSRE	DDMEGE	SLEHLQ
ZMO2 AND OTHERS	MERVF	SVEE	CPSEWY	FLEE	VVDPVEY	DLAAVALWRA	SSSSRE	DDMEGD	SLEHLQ
A2 GROUP									
	M[E/D]RVFS	NRS[A/S]SE WAFRF[I/L]Q E	E[Q/S/A]AFLK [S/R/Q]	LACAAVA	TSGSSR	DDD[A/L]	SPSDTS	VASLEHLQKR	
A3 GROUP									
	FSVDD	SQ[S/P]EW	Q[S/R]L	LK[S/N]KLE	RVG[T/S]VK PEDS[S/T]	SDDDDLD			
B GROUP									
	MKK..SEL	SSI..	GLSP						

One of common online tool is 9aaTADs "Nine Amino Acids Transactivation Domain 9aaTAD Prediction Tool" (http://www.med.muni.cz/9aaTAD/index.php) for predicting the presence of motifs that could be involved in the transcriptional activation. The main features of 9aaTADs are their acidity (Hope and Struhl, 1986, Gill and Ptashne, 1987), hydrophobicity (Drysdale *et al*., 1995), or both (Regier *et al*., 1993, Sainz *et al*., 1997, Sullivan *et al*., 1998). A relationship between transcriptional activation and hydrophobic amino acids (AlmlÖf *et al.,* 1997).

I analyzed three homologous of TaSPA comparing to other bZIP. So that, the result showed that TaSPA-B and TaSPA-D are very close to HvBLZ2, while TaSPA-A is completely different from others. TaSPA-B/D and HvBLZ2 had got a score around 7 out of 9 identical residues to known 9aaTAD members (Piskacek *et al.,* 2015) **(Fig. 2.15).** Furthermore, all these substitutions of amino acids in 9aaTAD can reduce transcriptional activation (Sainz *et al*., 1997). Although there are still some conserved amino acids, it is difficult to predict which residues can have a significant role in the motif. **(Fig.2.16).**

SPA-A	LMLQLVARV
SPA-B	EQVSLLRVA
SPA-D	EQVSLLRVA
Hv-BLZ2	EQVSLLRVA
Zm-O2	DQVAQLKAE

Figure 2.15 Sequences of motifs which present in some O2 members by using 9aaTAD.

Additionally, all these substitutions between the motifs could play a role in controlling the level of transcriptional activation. Although there are no studies to confirm the function of these motifs on the transcriptional activation. So that, identifying these key residues was crucial to confirm their function and confirm potential TAD property.

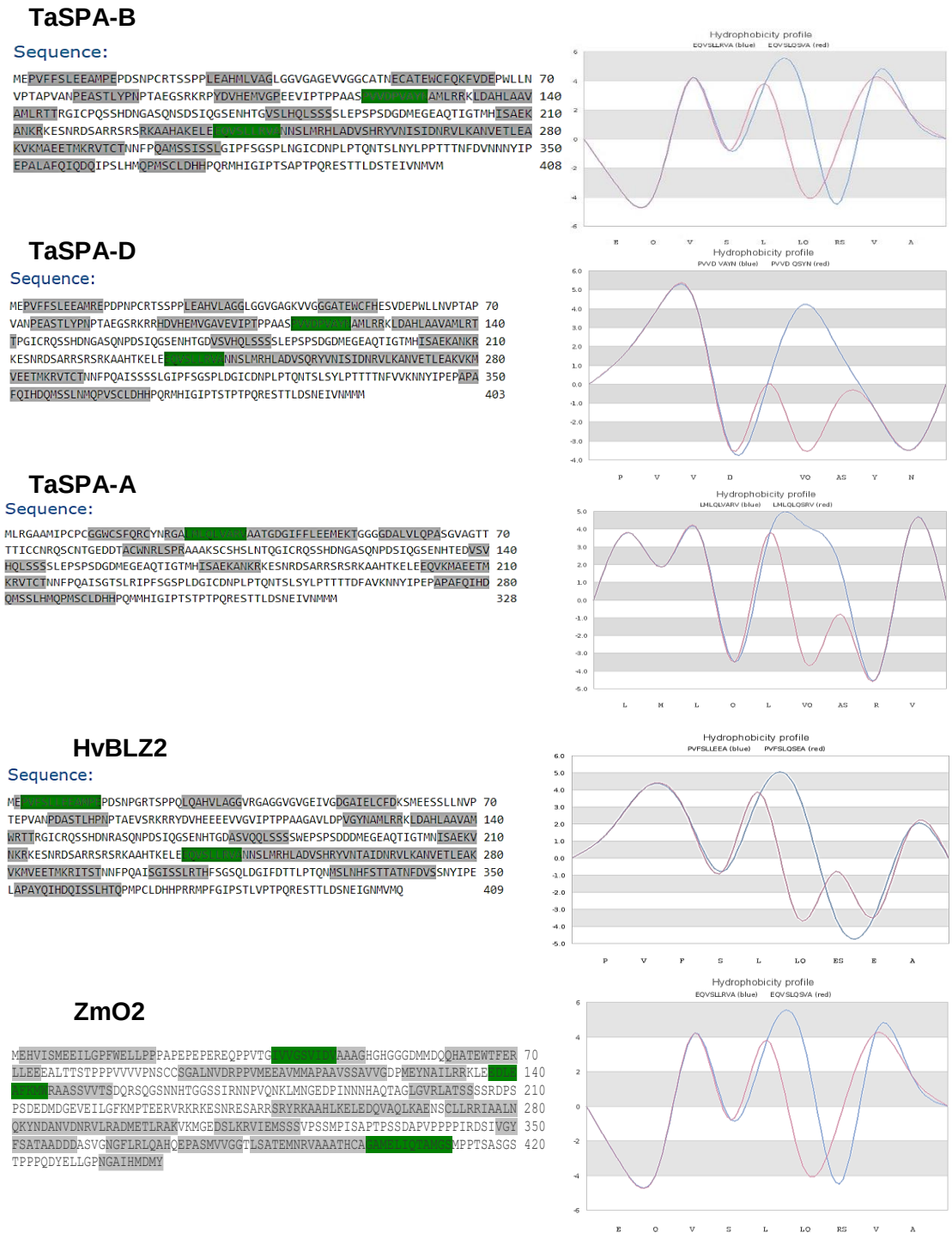


Figure 2.16 Consensus of three homologous of TaSPA comparing with TaBLZ2 and ZmO2.

The motifs were boxed in green which can be play role in the transcriptional activation. Their hydrophobicity profiles (blue curves) fits well with some of the known 9-amino-acid transactivation domain.

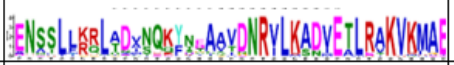
B) By using MEME tool (Bailey and Elkan, 1994). (http://meme.sdsc.edu/meme/website/), the total number of motifs were detected were 25 motifs **(Appendix 7.Table 7.2.6).** By comparing between A and B groups, they share the conserved bZIP domain **(Table.2.3). M3** was found in 37 members of A except for At-bZIP10. However, **M5** was found in a small subgroup of TaBLZ1 of A1 and members of A2. **M6** is specific to the A cluster except TaSPA and its orthologous. **M7** was found in 34 members of A cluster except BdO2, CenO2, Coix-O2, SrbZIP-1, SrO2 and ZmPRO2. **M8** was found in all members of the **A** cluster except the TaSPA subgroups (TaSPA, AeO2, TuO2 and HvBLZ2). All members of the A cluster have **M3**, **M6, M7** and **M8** except the TaSPA, which displays certain difference to the other members. **M9** is restricted to the **A** cluster, except for TaBLZ1, AeBLZ1 and OsRISBZ1 in A1, Musa63, AtBZO2H3 in A2, and AtbZIP25 and AtbZIP10 in A3. **M10** is restricted to A1 and A2 except for the subgroups of TaBLZ1, and TaSPA and its orthologous. **M11** was found in subgroup BLZ1, OsO2 and OsRISBZ4. **M12** was found in most O2 members except AeO2, OsbZIP, CenO2, SrbZIP-1, SrO2 in group A1. **M13, M15 and M20** were restricted to certain members of A2 which are NtBZ1-2, SoPRF2, NtCPRF2-like, SoCPRF2, PcCPRF2 and GlybZIP.

All members of A1 except BdO2, CenO2, SrbZIP, and SrO2 have **M14**, and all of A2 have this motif except Musa-63, AtBZO2H3, CuCPRF2 and GlybZIP. **M16** is restricted in cluster **B**. **M17and M25** were found in seven members of B cluster (BdRF2, HvRF2, ZmBZO2H1, MusabZIP9, OsRISBZ4 and Os-bZIP9). **M18** was found in a subgroup of TaBLZ1 of A1. **M19** is in the B cluster except AtbZIP9, HvRF2-like and OsRISBZ4. **M21** is restricted in 12 members of A1 except AebLZ1, TabLZ1 and BdbLZ1. **M23** is exclusively found in three members of A1, which are AeO2, TaSPA TuO2 and HvBLZ2. **M24** is specific to AtbZIP10 and AtBZO2H1 of A3.

Consequently, from the distribution of the conserved motifs, some motifs are specific in some groups such as M10 and M14 are conserved in A1 and A2. TaSPA and HvBLZ2 have a specific motif (M23). In addition, HvBLZ2 and its

orthologous have many specific motifs such as M11, M18 and M21. In addition, several motifs were found in B cluster such as M25, M17 and M19. Interestingly, members of A1 share various conserved motifs which divide this group into two subgroups, the first being TaBLZ1 and ZmO2 and the second TaSPA and HvBLZ2. In addition, A2 proteins can be divided into two groups, one for monocots especially for Panicoideae and another for the dicot species.

Table 2.3 Some conserved motifs in O2 family by using MEME which have functional associations.

No. of motifs	Sequence logo	Value	size	width	function
M1		4.6e-1274	59	31	bZIP domain (basic region)
M2		4.3e-1589	54	41	bZIP domain (leucine zipper)
M3		1.6e-234	37	15	Phosphorlation sites
M4		5.1-067	5	39	Phosphorlation sites

2.3.2 Analysis the Phosphorylation sites in bZIP proteins

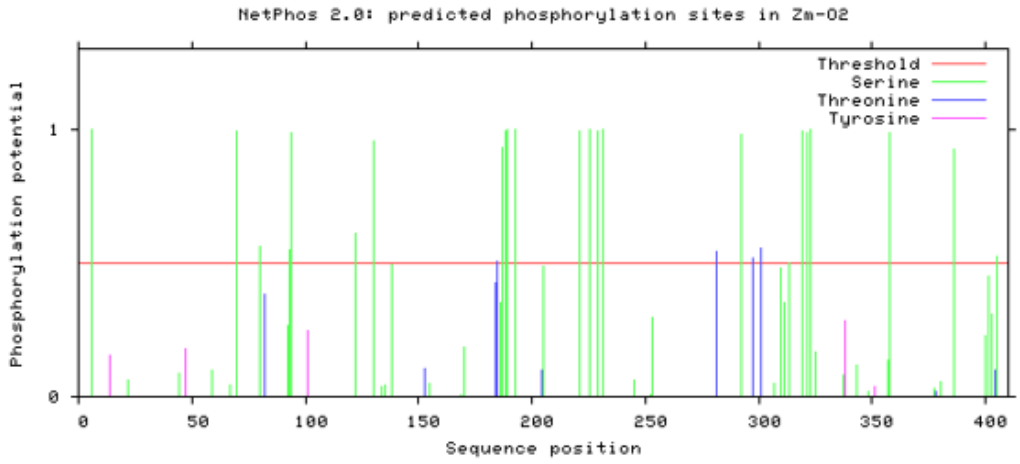
It is known that the activity of bZIP members can be regulated by phosphorylation. This kind of mechanism can control the function of the TFs at the post-translational level. In addition, the phosphorylation has the ability either to modify the dimerization specificity (Lee *et al.*, 2010). Also, it could control DNA binding by adding a phosphate group in the binding site of the bZIP domain (BD), this phosphate group will prevent the protein binding with its target because of the negative charge on the phosphate which results in a repulsive force between it and the DNA molecule (Deppmann *et al.,* 2004).

It has been confirmed that the efficiency of the protein phosphorylation was controlled by particular residues of the protein that play a role in regulation such as serine presents in DNA-binding site. There are many serine residues are conserved across all bZIP members. For example, the DNA binding region has a conserved serine amino acid in the basic region (BR), which is likely it is phosphorylation site Furthermore, phosphorylation has the role of controlling the bZIP subcellular localization, targeting bZIP either for nuclear import (Djamei *et al.,* 2007) or for cytoplasmic retention (Ishida *et al.,* 2008).

In the current study, a number of conserved motifs have been identified which are involved in phosphorylation, such as **M1** and **M2** represent RKX SNR, SARRxSR, K/RxxxD, KxxxED, respectively, and **M 6** presents as T/SxxE/D DDDD. Also, **M7** (TSx SSR [E/D] XSDDD/ E) which appeared in most of the bZIP genes as HvBLZ2, HvBLZ1 and ZmO2, and **M8** contains one of these sites, presented as SxxD/E, However, TaSPA and its orthologous such as HvBLZ2 have **M19** (SRSRxxDxxE) and **M23** (RxSS), while TaSPA has **M14** exclusively. The dicots such as Solaneceae and Nicotine have **M22**, represents as SSSxxxxxxK.

However, the analysis of the residues for phosphorylation in whole protein sequences of these transcription factors detected all serines are the presence in the proteins. The prediction of phosphorylation sites by NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos-2.0/output.php) (Blom, Gammeltoft and Brunak, 1999) which suggested numerous other potential phospho-sites in all bZIPs such as threonine (T) and tyrosine (Y). There are two examples of the O2 family (TaSPA and ZmO2) (**Fig. 2.17)** which shows differences in their phosphorylation sites in terms of the types of amino acids present, the numbers of these residues, and their position. However, both have a large number of serines.

Ta-SPA



Zm-O2

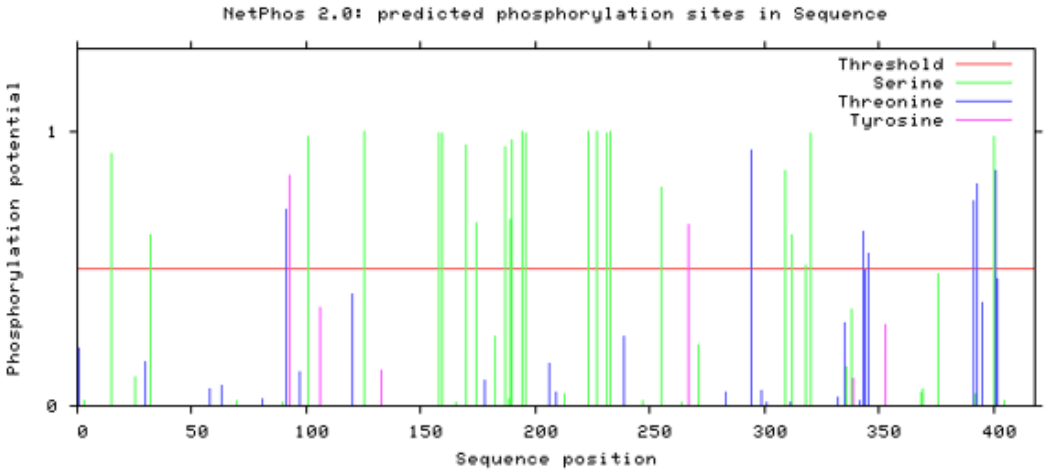


Figure 2.17 Diagrams show predicted phosphorylation sites in the whole sequences of TaSPA and ZmO2.(http://www.cbs.dtu.dk/services/NetPhos-2.0/output.php).

2.3.3 **(RSR1, Q) euAP2 clade**

2.3.3.1 ***Identifying the orthologous of euAP2 proteins and Phylogenetic analysis***

All members euAP2 in different species such as *Triticum aestivum*, *H. vulgare chilense, Oryza sativa, Brachypodium distachyon* and other species of monocots and dicots such as Arabidopsis were identified by conducting a BLAST X search against published full-length cDNA datasets. Around 49 proteins of dicot such as Arabidopsis and Brassica were identified that belong to euAP2 family. The results showed TaQ has high homology with many floral homeotic proteins such as *H. vulgare chilense*, *AP2*-like mRNA, *Zea mays* indeterminate spikelet 1 (*ids1*) mRNA, and *Oryza sativa AP2D2* mRNA. In addition, TaRSR1 and its orthologous were identified in *Oryza sativa*, *H. vulgare chilense*, *Sorghum bicolor* and *Brachypodium distachyon* (BdRAPD2-7). All these proteins belong to the euAP2 family of putative transcription factors were characterized by the presence of two DNA binding motifs, referred to as AP2 domains.

Moreover, *euAP2* genes have a miRNA binding site at the C-terminus is highly conserved across within *euAP2* members in dicots and monocots. The multiple alignments of proteins were used to construct a phylogenetic tree using MrBayes tool, the model of substitution protein Wag was used to which represent the relative rates of amino acid replacement at homologous sites in a protein, to accurately estimate the true evolutionary distances and relationships among species. (Le *et al.,* 2008). From the construction of the phylogenetic tree that infers the relation between the orthologous and paralogous among the monocots and dicots. The euAP2 proteins appear to be monophyletic, with a high posterior probability approximately (1). Bayesian phylogenetic tree of euAP2 proteins of Poaceae and dicots showed that they share a common ancestor **(Fig. 2.18).**

Furthermore, two major clades can be recognized, one of which included all the members of the Poaceae and was named (clade A). Another clade has the dicot species only (clade B). In addition, from the topology of these proteins, it may be noted that two groups present in the first clade (A), represented by RSR1 and Q, which means both are paralogous genes with a specific lineage in the monocots.

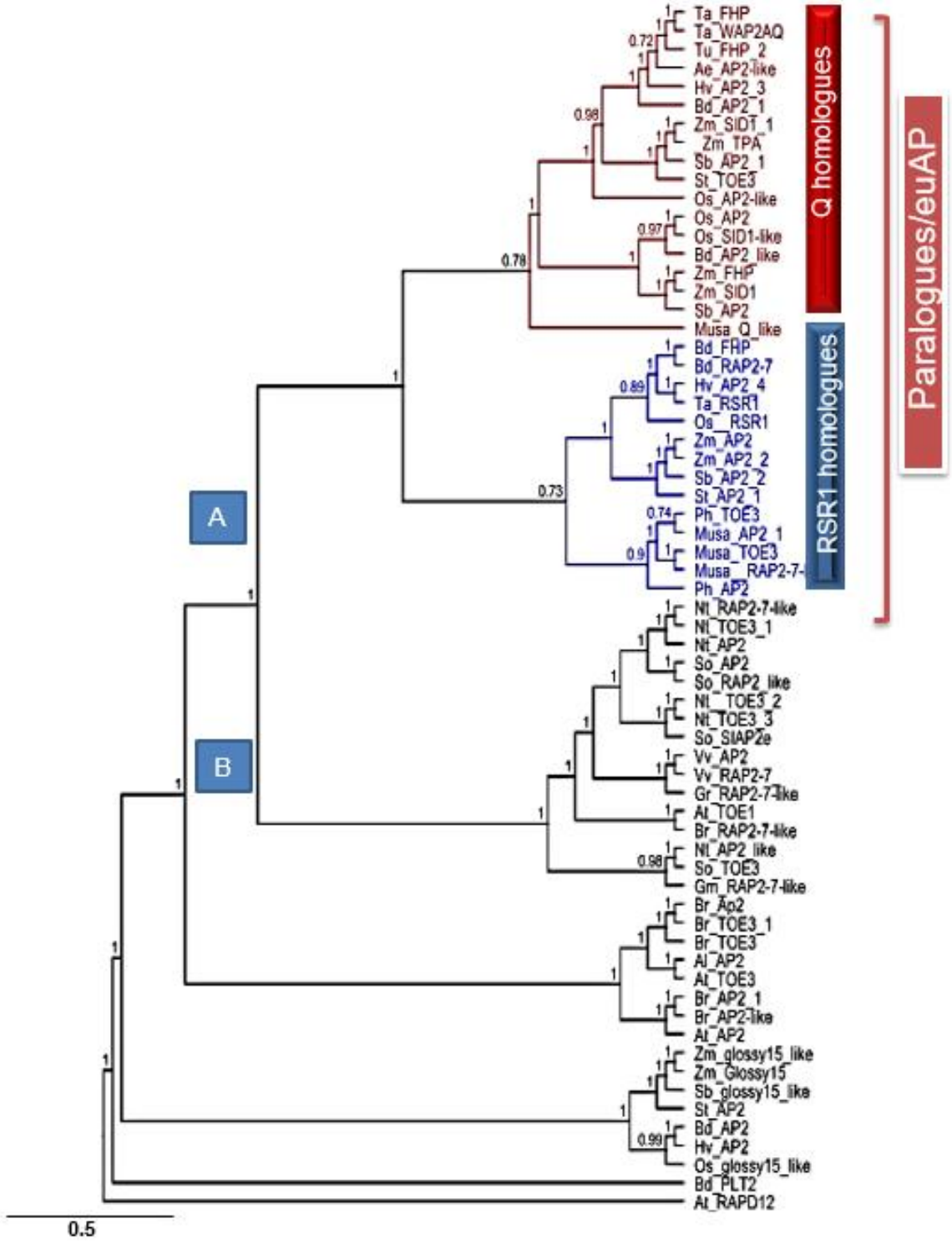


Figure 2.18 Phylogenetic relationship among the euAP2 proteins based on the whole sequence of proteins. The unrooted tree was generated using MAFTT alignment tool, MarBays method. Two groups of euAP2 family are colored red and blue respectively.

2.3.3.2 *Characterization of (RSR1, Q); euAP2 genes*

The distribution and the numbers of exons and introns of *Ta-RSR1* and its orthologous in different species such as Barley, Rice and Brachypodium showed some considerable variation. For example, the *TaRSR1* has nine exons and eight introns and encodes a protein of 498 a.a., whereas *Hv \RSR1* has seven exons and six introns, and Os*RSR1* has eight exons and seven introns. In Brachypodium, *BdRSR1* has nine exons and eight introns. Generally, the orthologous of *RSR1* in different species show variations in the size and the number of exons **(Fig.2.19).**

1. Furthermore, the *TaQ* gene has six exons and five introns and encodes for 447 a.a. The orthologous of *TaQ* was identified in barley, which shows differences in terms of the number of exons and position, and it has five exons and four introns. In (*Zea mays*), floral homeotic gene *APETALA 2* has nine exons and eight introns and encodes for 465 a.a. Also, floral homeotic gene *APETALA 2 (Brachypodium distachyon*) has nine exons and eight

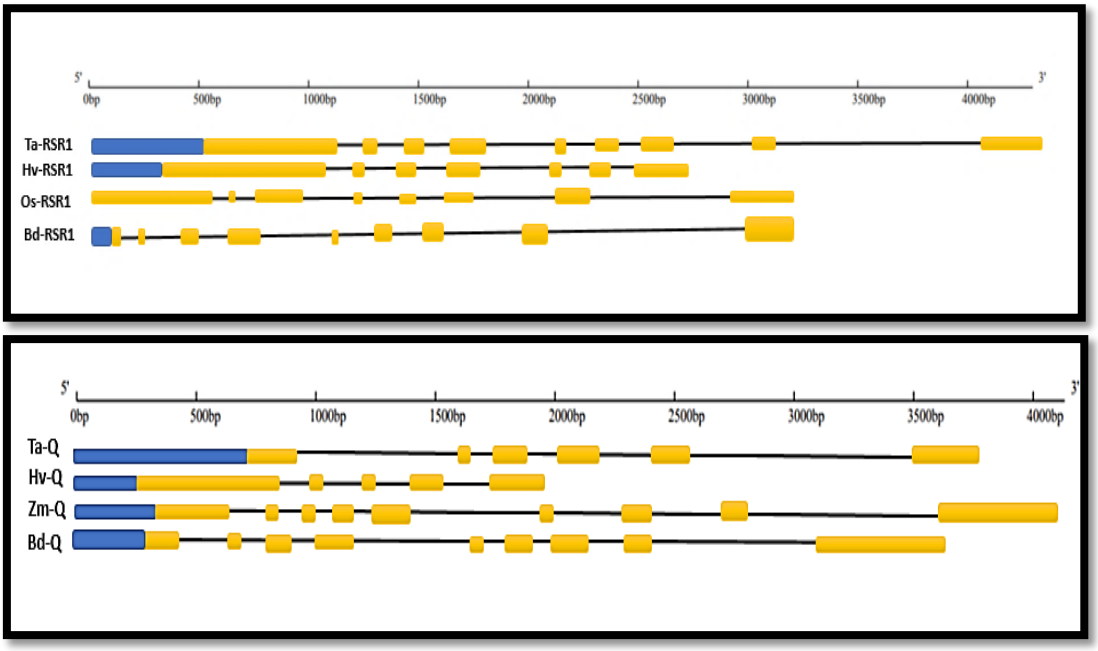


Figure 2.19 The number of exons and introns that comprise TaRSR1 and TaQ genes in wheat and its homologous in different species.

The yellow colour indicates exons, and the black line indicates introns, while blue colour indicates the 5'UTR region.

As it is known the position of exons and introns is considered as a key to evolutionary relationships of genomes (Nakano *et al.*, 2006). In euAP2 members have introns especially in the conserved AP2 domains. Comparing between two groups of paralogous (RSR1 and Q), the aligned proteins showed conserved positions of introns within the AP2 domain. Two intronic markers were found conserved in both the AP2-R1 and AP2-R2 of all euAP2proteins **(Fig.2.20).** Consequently, the finding of conserved intronic sites suggests the evolution of introns before gene duplication.

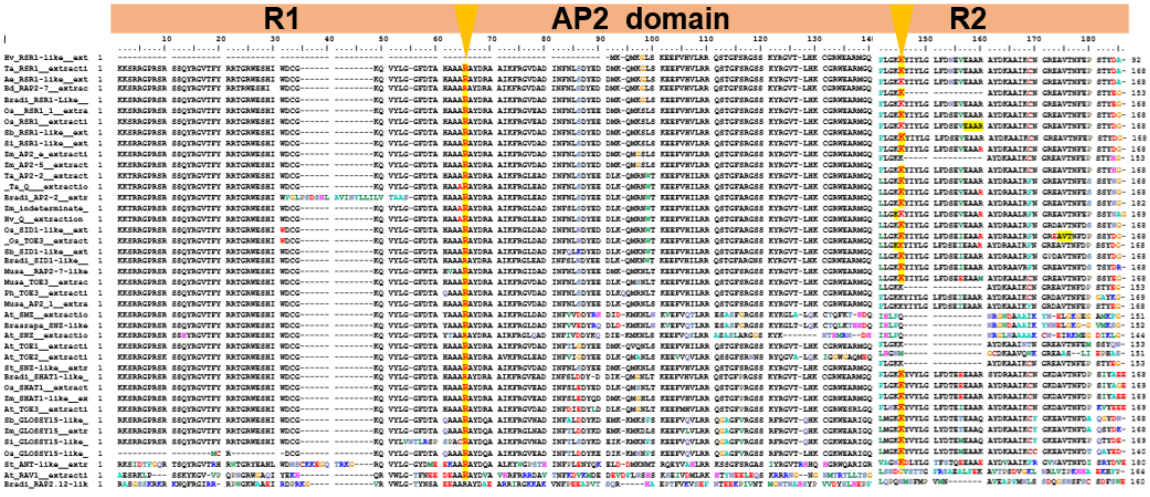


Figure 2.20 Analysis of the introns in the euAP2 domain sequence of proteins.

The euAP2 domains were aligned and detected the intron position by using GeneWise. The position of intronic sites is highlighted in yellow in the protein sequence.

2.3.3.3 *AP2 domain analysis*

The aligned full length of euAP2 proteins of monocots and dicots showed highly conserved domains and other conserved motifs within different species. **(Appendix 7. Fig.7.2.2)** shows aligned sequences of all the euAP2, highlighting the conserved motifs and the AP2 domain. In comparing between the AP2 proteins, they have a similar structural organization such as the motifs in the N-terminal (M1, M2, M3), and the nuclear localizing signal, the first AP2 domain (AP2-R1), the second AP2-domain (AP2-R2), M4 and the miRNA172-binding site **(Fig.2.21).**

Furthermore, AP2 domains (AP2-R1 and AP2-R2) are strongly conserved among species in the monocots and dicots. Consequently, the AP2 domain in plants is considered a specific domain (Magnani *et al*., 2004). The euAP2 in the monocots can be seen to have many conserved motifs after the AP2 domain in the C-terminus comparing to dicots.

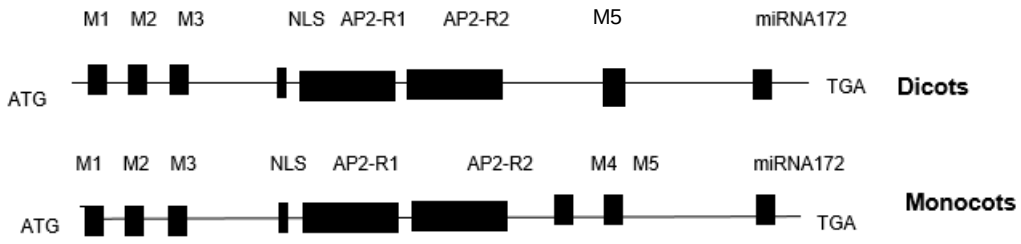


Figure 2.21 The general structure of euAP2 proteins in monocots and dicots.

The black boxes are indicated distribution and position of the conserved motifs.

The euAP2 proteins contain two AP2 domains, which are known as AP2-R1 and AP2-R2 (around 60, up 70, amino acids, respectively). Each one contains three β-sheets and one α-helix. The DNA binding region is highly positively charged, and it binds to the major groove of DNA via its three-stranded *β*-sheets. By using Phyre2 (www.sbg.bio.ic.ac.uk/~phyre/) to predict the 3D structure of the AP2 domain comparing between the paralogous TaRSR1 and TaQ may be noted that both have the same structure as the AP2 domain but with replacements in three positions which might play a role in the functions of those proteins **(Fig.2.22).** However, the 3D structure of AP2 domain shows variation between the TaRSR1 and the model of AP2 in Arabidopsis which has been obtained from the RCSB PDB (https://www.rcsb.org/) to predict 3D structure for AP2 family. This prediction structure of the AP2 domain by comparing between TaRSR1 and AP2 Arabidopsis shows the short part in the region between the *β*-sheets of AP2 domain in Arabidopsis.

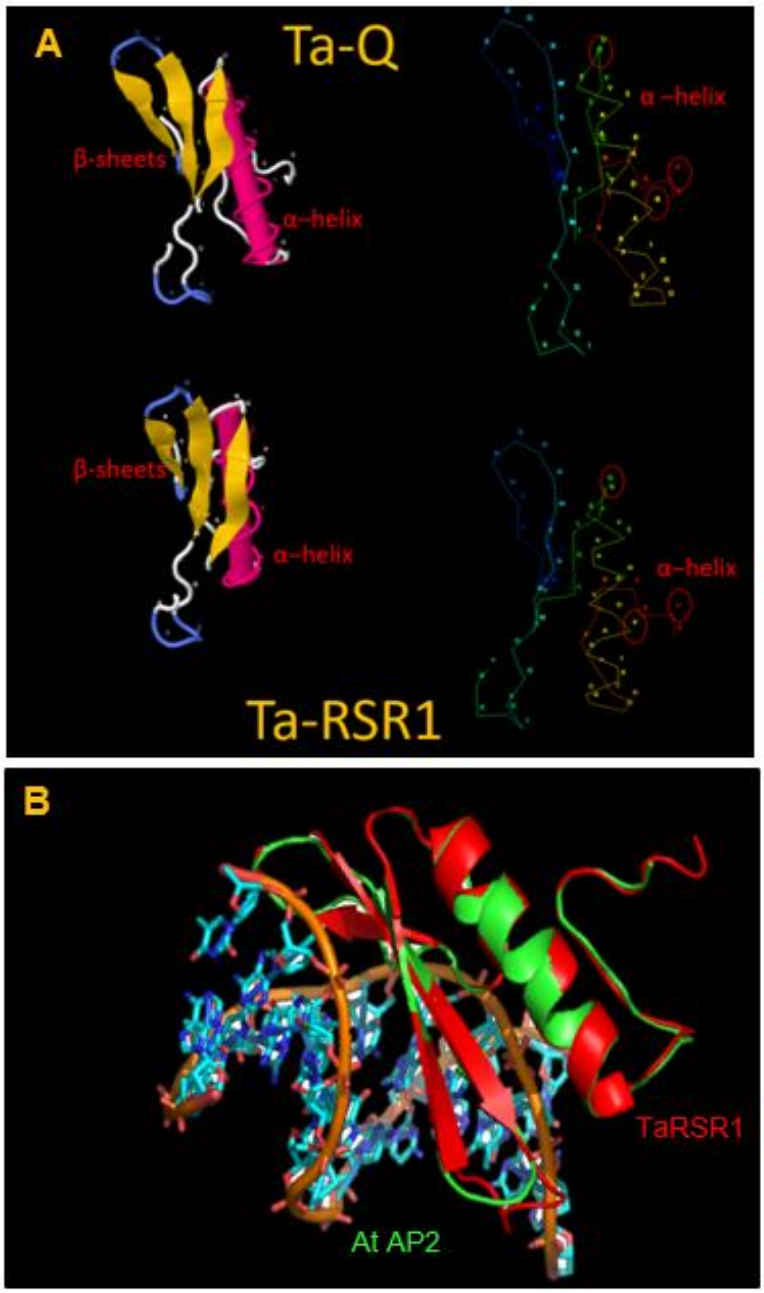


Figure 2.22 A-The 3D structure of AP2 by using Phyre 2, comparing between TaRSR1 and TaQ in wheat.

B- AP2 domain of TaRSR1 wheat in red color, comparing to AP2 in Arabidopsis in green color.

In addition, one of the conserved regions of the AP2 domain is a linker is around 25 amino acids, and it connects two AP2 domains (R1 and R2). It is could have a critical role in the function of AP2/ERB to contact DNA (Wolfe *et al.,* 2000). Thus, if a mutation occurs inside this region, it can affect the ability of the domain to bind DNA. (Saleh, 2003) **(Fig.2.23).**

Additionally, the aligned euAP2 proteins showed two conserved regions within the two AP2 domains. The first region is the YRG element which is approximately 20 amino acids and is rich in basic amino acids (Arg, R) and hydrophilic residues (Tyr,Y; Gly, G and The, T). YRG is highly conserved in both R1 and R2; these residues likely have the ability to bind protein with the DNA by contacting their sugar phosphate backbones because they contain basic amino acid. Also, the hydrophobic residues in this region are conserved, which play a role in stabilization (Magnani *et al.,* 2004).

The second region is RAYD, which has around 40 amino acids at the C-terminus of the AP2 domain. In this region, 18 amino acids of AP2-R1 are identical, which is likely has a functional role, while in AP2-R2 showed a slight variation especially between TaQ group and TaRSR1 group. Jofuku *et al.,* (1994) described this region in *Arabidopsis thaliana* as showing 72.2% identity between the two AP2 domains, R1 and R2; also, this region could participate in protein-protein interactions. Segrest *et al.,* (1990) confirmed that one of the functions of this region might be to mediate AP2 dimer formation through the interaction of these amphipathic helices (Flemington and Speck, 1990). Amphipathic α-helices can mediate protein-protein interactions through the formation of coiled-coil structures.

Both AP2 domains have some conserved amino acids such as (**YRGVT**), (**RWE**), (**YLGLF**), and α-helices (**AAARAYDRAAIR**). Moreover, the 3 β-sheets rich with Arg (R), and Leu (L) and Try (W) residues which play the important role to form homo-dimer by contacting the DNA (Jofuku *et al.,* 1994; Simon *et al.,* 2006).

Another region of AP2 domain is α-helices region, which has several conserved of the hydrophobic amino acids such as Ala (A), Val (V), Phe (F) and Asp (D). Jofuku

et al., (1994) confirmed that the importance of the conserved amino acids in AP2 domain by comparing between two mutant lines in Arabidopsis, which have substitutions at glycine residues in the same positions in the two AP2 domain repeats. This substitution of glycine residues with either Ser or Glu may disturb the formation or the position of these α-helical structures within the AP2 protein structure. This indicates the conserved amino acids are likely to be necessary for its proper function.

Generally, the alignment sequences show there are various replacements but each euAP2 clade share some conserved amino acids. For example, Q protein and its orthologous have conserved amino acids, which were substituted in another group of RSR1 and its orthologous **(Fig.2.23)**

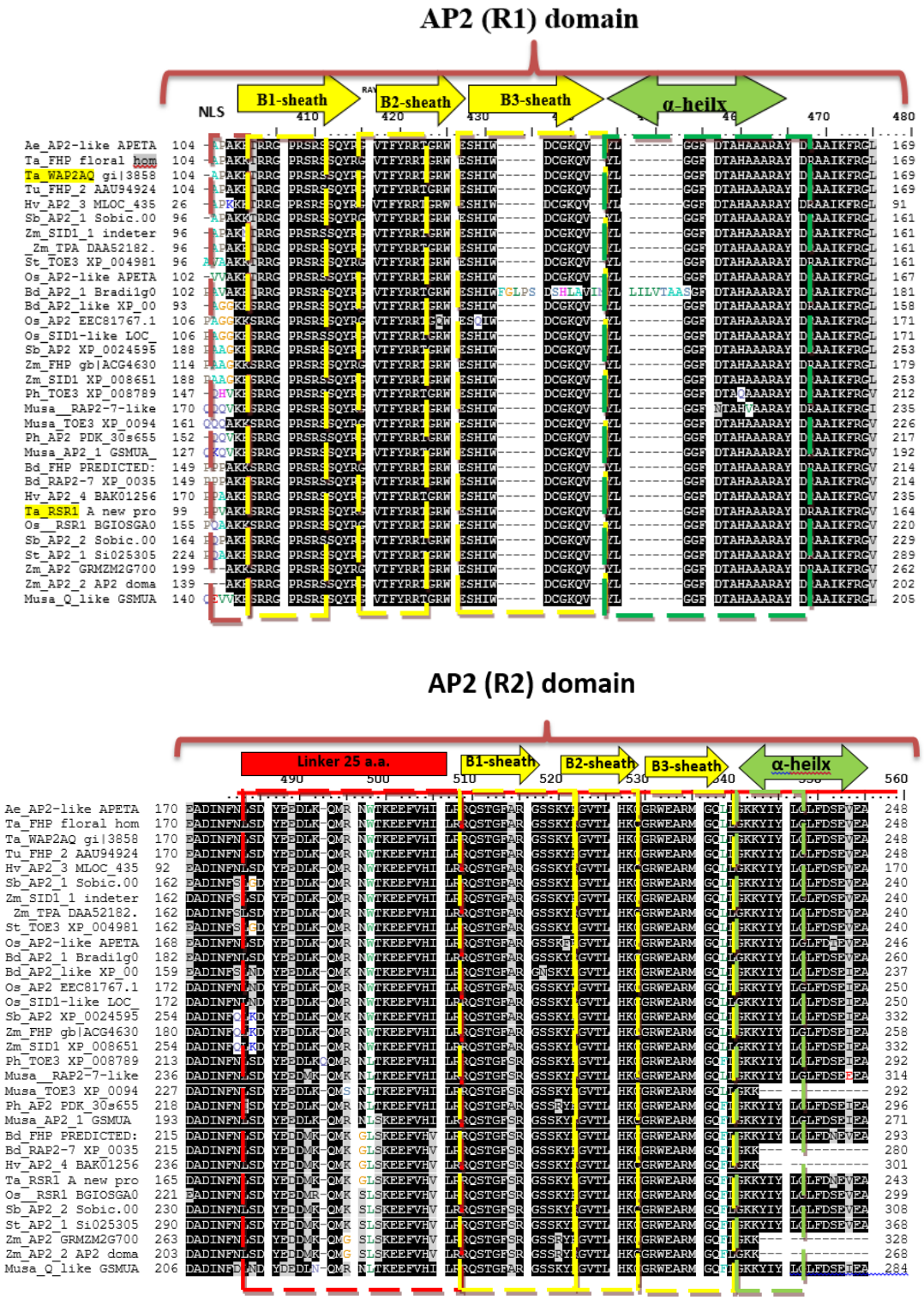


Figure 2.23 The euAP2 protein sequences were aligned, showing the structure of AP2 domain.

In addition, another feature in euAP2 members is NLS presents at the N-terminus of the AP2-R1 domain, its length around 10 amino acids (KKSR …). The function of NLS has been confirmed by Jofuku *et al.,* (1994) suggested that AP2 may have a function in the nucleus because of the sequence of NLS which appears in the most members of AP2. NLS is enriched with arginine and lysine, with a difference of only one amino acid between the paralogous, where the third amino acid in the NLS is threonine (in the TaQ) instead of serine (in TaRSR1). That means that TaQ and its orthologous have KKT…, while TaRSR1 and its orthologous have KKS…, with the expectation of Zmsidi, although Gil-Humanes *et al.,* (2009) confirmed that the nuclear localization is conserved in the Triticeae of AP2 in this study.

TaQ AKKTRRGPRS

TaRSR1 AKKSRRGPRS

However, Liu *et al.*, (2018) have confirmed that TaQ protein plays a role as a nucleus-localized transcriptional regulator. In this study, it was confirmed that the subcellular of TaRSR1 in wheat and barley, HvRSR1 shows signals in the nucleus and in the cytoplasm, were shown in **(Fig. 2.24).**

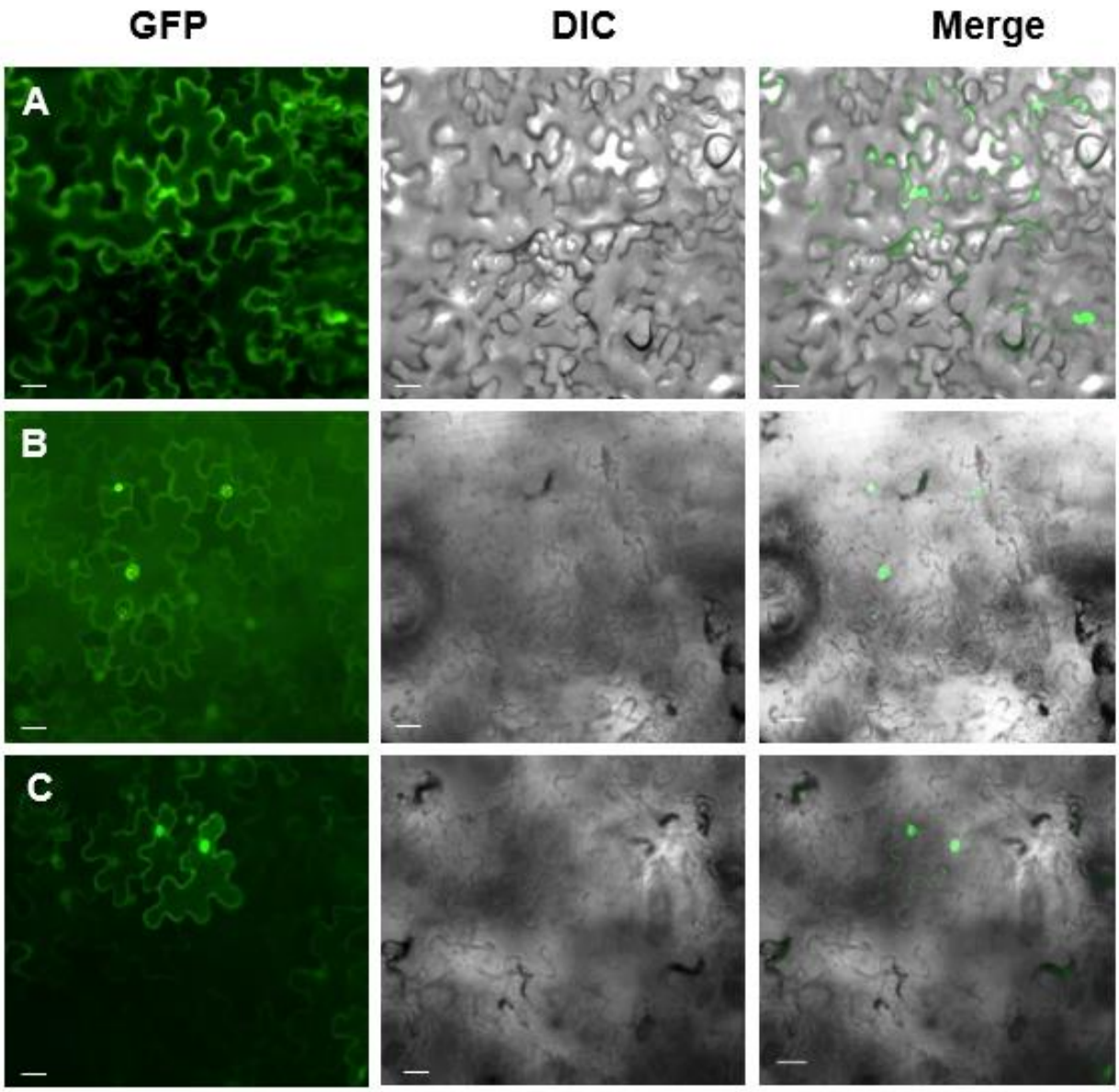


Figure 2.24 Subcellular localisation of TaRSR1/ HvRSR1 variants.

Fluorescence microscopy analysis shows nuclear protein localization *N. benthamiana* leaves. **A:** promCaMV 35S:2XGFP (control), **B:** promCaMV 35S:TaRSR1-GFP, **C:** promCaMV 35S:HvRSR1-GFP.

2.3.3.4 *Conserved motifs outside AP2 domain and MEME analysis*

Aligned sequences of euAP2 proteins including the dicot, such as Arapidopsis and Brassica appear the main structure of euAP2, and the conserved regions throughout these proteins. There is a conserved motif in the N-terminus, which is the EAR1 motif, then two AP2 domains, another conserved motif known as the EAR2 motif, and a miRNA binding site. In addition, in the N-terminus of the AP2-R1 domain there is a small conserved region, is nuclear localization. Generally, the members that share the same conserved motifs could have similar functions (Wu *et al.,* 2015; Rashid *et al.,* 2012). Also, the conserved motifs are correlated with the topology of the phylogenetic tree **(Fig. 2.25).**

It has been confirmed that the motif M (V/E) LDL) in the N-terminal plays an essential role in the transcriptional repression activity for euAP2, for example, this motif is essential for TaQ to interact with the transcriptional co-repressor, such as TaTPL (Liu *et al.,* 2018). The mutation in this motif of TaQ protein by replacing the tandem Alanine (AAAAAA) led to prevent QmEAR1 to interact with TaTPL, which means the N-terminus EAR1 motif of TaQ is required in order to interact with TaTPL (Liu *et al.,* 2018). In addition, there is a motif which has been shown to function as a repression domain (Fujimoto *et al.,* 2000; Ohta *et al.,* 2001), it has a consensus (L/V) DL (N/D) L(R/S) (I/M) SQP, in the C-terminus region and which is known as EAR2 motif. A mutation in this motif could reduce the repression activity of the protein (Tsukagoshi *et al.,* 2005). The last motif at the C-terminus is (AAASSGFS), which is identified as miRNA172, and it presents in all euAP2 members **(Table 2.4).**

Table 2.4 The main conserved motifs between the members of euAP2 family and the substitution amino acids within these motifs.

No. of motif	All euAP2	euAP2 in Poaceae	Putative function
M1	M.LDLNV	M[V/E]LDL	Transcriptional repression activity/protein protein interaction
M2	SSV.N	SSVLN	Transcriptionactivation +phosphoralyation.
M3	____	TR[Q/E]LFP	Transcription activation
M4	____	TaQ (DDDDCS) TaRSR1 (G/A)DD X E	
M5	(L/V)DL(N/D)L	(L/V)DL(N/D)L	Repression domain
M6	____	PKRI exclusive (TaQ cluster), (TaRSR1 cluster)	
M7	AASSGF-	AAASSGFS	MiRNA172

The distribution most the motifs among the specific clades of the phylogenetic tree could reflect the structural similarities among proteins within the same group as (**PKRI**) motif. Also, (**MLDL….)** in the N-terminus is a conserved amino acid sequence, which may be involved in various different biological functions such as transcriptional activity and protein-protein interactions. Motifs are rich with acidic amino acids such as Gln (Q), Pro (P), Ser (S), or Thr (T) are often described as transcriptional activation domains (Liu *et al.,* 1999) such as GTS, SSVLN and QLFP. Furthermore, most of the conserved motifs were characterized by their structures in terms of amino acid compositions, whereas the functions of these motifs have not been entirely demonstrated. In addition, there are several motifs related to putative phosphorylation sites, which are conserved in proteins in euAP2, whilst some are distributed across all members of euAP2, such as SGTS in the N-terminus, and GSP at 250 amino acid position. Also, the motif SSVLN presents in the N-terminus which could for phosphorylating (Rashid *et al.,* 2012; Nakano *et al.,* 2006).

There are many phosphorylation sites; however, comparing between two main groups (Q and RSR1) of euAP2 can be noted that there are many replacements between them. For example, the motif (SPE) at the C-terminus for RSR1, while [S(F/Y)E,SS] motif was found in (Q) group, but GS motif was in (RSR1) group.

Also, SPL motif was found in (Q) group, while RSR1 group has (MPL) motif. PPP motif was detected in TaQ group, while, RSR1 group has (PSS) motif. In addition, (SSSS) motif is exclusive to the RSR1 group such as BdFHP, BdRAP2-7, StAP2/1, HvAP2-4, TaRSR1, OsRSR1, and Sb-AP2.

MEME was used to detect more motifs within the euAP2 family, which were around 25 common motifs **(Appendix 7. Table 7.2.7).** All members of the euAP2 share a large number of conserved motifs, which were distributed among the members. The result showed that most of these motifs are selectively distributed among specific clades; for example, M1, M2, M3, M4, M5, which correspond to the AP2 domain region, whilst M8 M9, M13, M14, M6, and M7 were found in all members at the C-terminus. M7 is the motif for miRNA172, is very highly conserved **(Table 2.5).**

However, there are some specific motifs were found in the number of proteins such as M10 in TaQ and its orthologous. However, members of TaQ group share motifs as M13, M9, M14, M8, and M17 in the C-terminus, and M11, M10, M12 and M22. M22 can be found in MusaQ-like and MusaAP2-1. However, the RSR1 group has many conserved motifs such as M13 and M9, with the exception of TaRSR1 has particular motifs which are M14 and M8 in the N-terminus. However, M16 was distributed within the TaRSR1 group with the exceptions of TaRSR1 and OsRSR1, and it was found in some members of euAP2 such as NtTOE3-2, NtTOE3-3, SoSIAP2, SoAP2, SoRAPD2-like, VvAP2 and VvRAPD2-7. M9 is rich with Serine (S), identified as (SGTSXSS), which may function as an activation domain (Jofuku *et al.,* 1994).

TaRSR1 shows some slight differences from the other members of euAP2 in terms of the positions of some of the motifs. For example, TaRSR1 has (M22) in the N-terminal region with consensus sequences [PPPP], but the position of this motif in the C-terminal of some proteins such as AeAP2, TaFHP, TaQ, TuFHP2, SbAP2-1, OsAP2, OsSIDI-like, MusaQ like and MusaTOE3. In addition, all euAP2 members have M13 at the N-terminal, while this motif has a different position in TaRSR1, namely at the C-terminus. Furthermore, TaRSR1 and its orthologous

have a specific motif, which is M17. This motif was found in TaQ and its orthologous. As a result, M17 is specific to TaRSR1 and TaQ. Generally, both TaQ and TaRSR1 groups share some specific motifs, such as M13, M9, and M14 except TaRSR1, M8, M11, M12 and M7. That means this lineage of the monocots within euAP2 is more conserved than ather lineage of euAP2. In addition, M15 was found in BdRAPD2-7, ZmAP2, ZmAP2-2, NtAP2, HvAP2-4, MusaTOE3 and BrRAPD2-7-like. In terms of euAP2 of dicots share, the monocots in several conserved motifs such as M1, M5, M4, M2 M3 and M6 were considered as AP2 domain. In terms of M6 has consensus [L/Y] DL [N/S] LXXSQP was found in the entire AP2 family, could be a repressive motif (Nakano *et al.,* 2006). In addition, these proteins of dicots have various motifs, namely M13, M9, M14, M8, M11 and M7. Also, they share TaQ group with M10, while they share TaRSR1 with M16. This group of proteins has specific motifs such as M23 and M19.

In the previous study on the AP2/ERF family members in rice, there is a group-I in rice which has a conserved M5. This motif is similar to the M5 consensus sequences [R/S, R/S, G, P/F, S/R, S/R, R/G, S/A, S] in our group, namely those rice proteins involved in seed (embryonic growth, seed development, seed germination) and flower traits (petal cell identity, flowering time, and flower meristem identity (Rashid *et al.,* 2012).

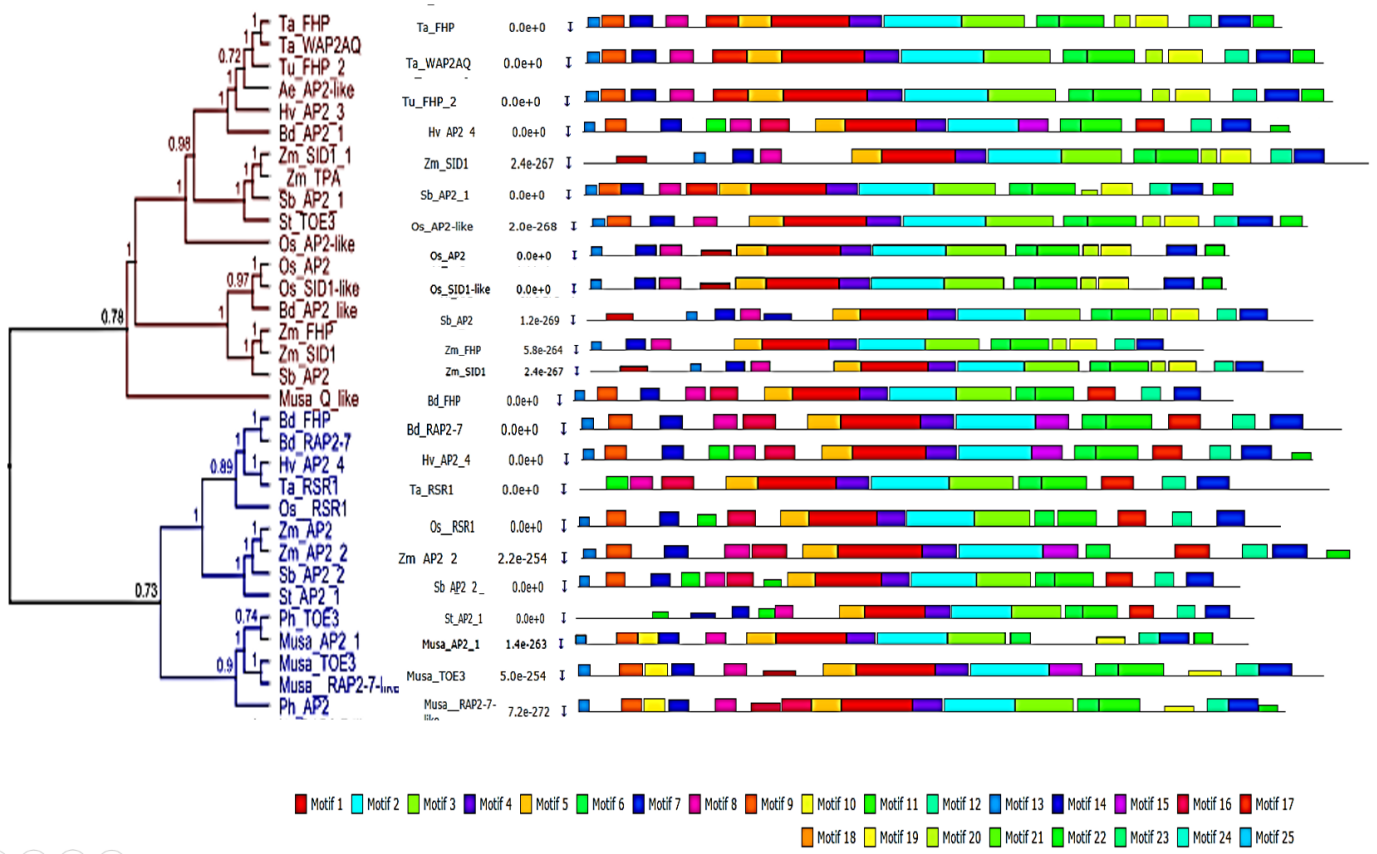


Figure 2.25 Phylogenetic analysis of euAP2 and distribution and position of conserved motifs along the euAP2 proteins by using MEME.

Table 2.5 25 conserved Motifs of euAP2 proteins were identified by using MEME.

	Sequence logo	E-value	Size	Width	Annotation
M1		2.9 - 3.274	61	50	AP2 domain
M2		1.0 - 3090	62	50	AP2 domain
M3		8.9-1520	53	41	AP2 domain
M4		4.3e - 348	59	15	LxLxLxLPP, conserved motif in the C-terminal regions.
M5		3.9e-335	59	21	MiRNA binding site
M6		1.1e-186	36	15	function as an activation domain (Jofuku *et al*. 1994). It is found in Q and some dicots.

2.3.4 Identification of plant microRNAs in euAP2

The method for identifying miRNA was based on the major characteristics of those miRNAs, they usually consisting of ~20–22 nucleotides for animals and ~20–24 nt for plants (Reinhart *et al.*, 2002; Bartel, 2004). The miRNA represented in different species of plant with different families of miRNA. In addition, most plant miRNAs are primarily found in the protein coding genes in the 3′ untranslated regions of the target mRNAs (Reinhart *et al.,* 2002). By analysing the sequences of *euAP2* members, can recognize a conserved miRNA in the similar position in the 3' UTR, exon 10 in different euAP2 genes of Arabidopsis such as *TOE1*, *TOE2*, and *TOE3* (*At2g28550*, *At5g60120* and *At5g67180*, respectively) and in maize *AP2* genes such as Zm-ISID1 (*INDETERMINATE SPIKELET1)* **(Appendix2.Table7.2.5)** (Chuck *et al.,*1998) and *GLOSSY15* (Moose and Sisco, 1996).

To identify the type of miRNA in *euAP2* genes, their sequence were extracted from the data for subsequent alignment. The results showed that all euAP2 genes have a conserved miRNA and only a few nucleotides change between the miRNA binding sites**.** For example, the aligned sequences of miR172 in two paralogous (TaRSR1, TaQ) show variation between them **(Fig.2.26)**. This result was similar to the observations of Floyd and Bowman, (2004). Also, by blasting the *euAP2* sequences in the website miRbase (www.mirbase.org/), can identify the type of miRNA in these genes. The result showed that all *euAP2* genes have one type of miRNA which is miR172, there is only a single miRNA-complementary site, and most corresponding miRNAs perfectly complement these sites and cleave the target mRNAs (Kidner and Martienssen, 2005) **(Fig.2.27)**.

According to Sunkar *et al.* (2008), all four members of the miR172 family were highly expressed in seedlings, though miR172c was not expressed in grains. Consequently, most euAP2 members have 172a, 172e and 172d.

miRNA172 binding site in the *euAP2* genes

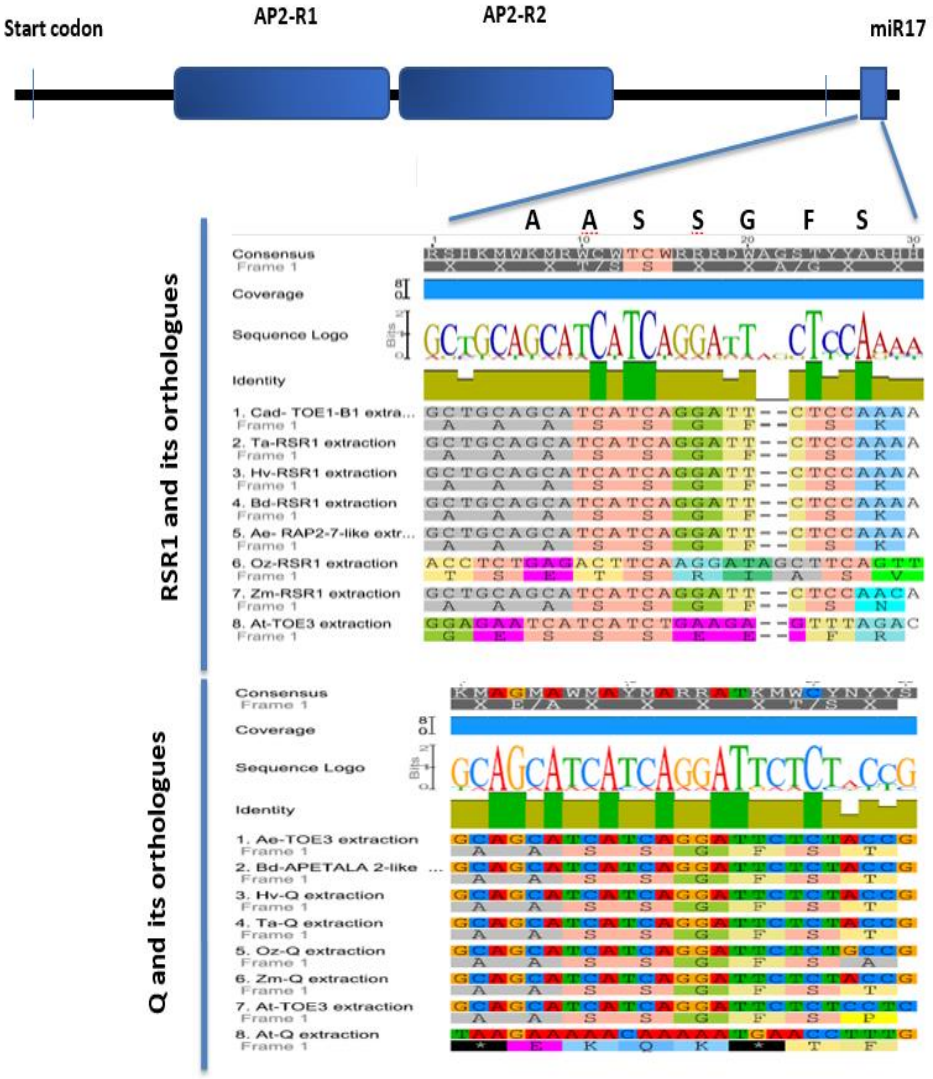
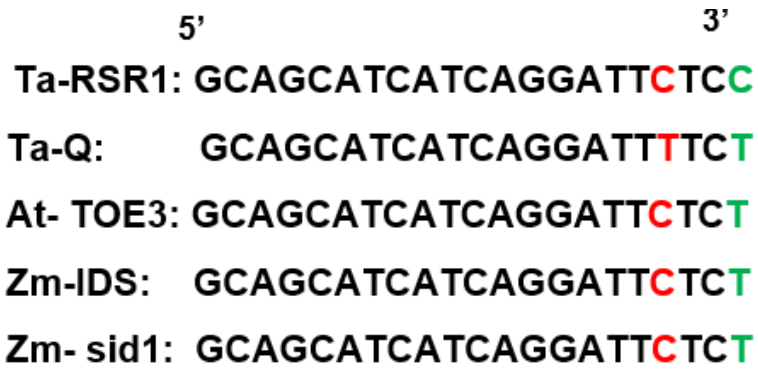


Figure 2.26 Alignment sequences of euAP2 genes (TaRSR1 and its orthologous and TaQ and its orthologous), showing miRNA172 binding site.

Comparing euAP2 proteins between the monocots such as, INDETERMINATE *SPIKELET1* maize (Chuck *et al.,* 1998) and *GLOSSY15* (Moose and Sisco, 1996), and dicots such as *TOE1*, *TOE2*, and *TOE3* in Arabidopsis, all of them to have the *miR172* target sites in similar locations, and all the *euAP2* genes have the same type of miRNAs. Aukerman and Sakai, (2003) reported members of *AP2* family have *miR172* target sites from various other plant species including soybean, rice,

tomato, and pea. Moreover, our results confirmed that all the varieties of wheat such as *Chinese spring*, cadenza durum and *Hordeum vulgare* have the same miRNA 172. By comparing between *TaRSR1* and *TaQ* genes, both have highly conserved miRNA172d binding sites except for two nucleotides in the 3' end. However, it is clear that 5' region of miRNA172 is highly conserved in all *euAP2* genes. Generally, the conservation in the sequence of miRNAs in different species is likely due to the regulation of the expression of these particular *AP2* family members in response to *miR172*-like miRNAs. However, the sequence of mi172 in the *euAP2* genes showed the similarities between them, which confirmed that mi172 is conserved.



1-CTGCAGCATCATCAGGATTCTC *Chinese spring* RSR1

2-CTGCAGCATCATCAGGATTCTC *Hordeum vulgare* subsp. vulgare RSR1

3-CTGCAGCATCATCAGGATTCTC ***Triticum aestivum cultivar Cadenza*** **target of EAT1-B1 (TOE1-B1) gene**

4-CTGCAGCATCATCAGGATTCTC ***Brachypodium distachyon*** **floral homeotic protein APETALA 2**

5-CTGCAGCATCATCAGGATTCTC **Oryza sativa Japonica Group AP2-like - TOE3c**

6-CTGCAGCATCATCAGGATTCTC ***Zea mays*** **floral homeotic protein APETALA 2**

7-CTGCAGCATCATCAGGATTTTC ***Triticum aestivum*** **Q mRNA for transcription factor WAP2AQ**

8-CTGCAGCATCATCAGGATTCTC ***Triticum spelt var.*** **duhamelianum q mRNA WAP2Aq**

9-CTGCAGCATCATCAGGATTCTC ***Triticum turgidum*** **subsp. dicoccoides floral homeotic protein (Q) gene**

10-CTGCAGCATCATCAGGATTCTC ***Hordeum vulgar*** **subsp. vulgar mRNA for Floral homeotic APETALA 2**

11-CTGCAGCATCATCAGGATTCTC ***Brach podium distachyon*** **floral homeotic protein APETALA 2-like**

12-CTGCAGCATCATCAGGATTCTC ***Triticum aestival cultivar*** **Charger target of EAT1-B1 (TOE1-B1) gene**

13-CTGCAGCATCATCAGGATTCTC ***Zeal mays*** **indeterminate spikelet 1 (ids1)**

14-CTGCAGCATCATCAGGATTCTC ***Zeal mays*** **- AP2-EREBP transcription factor (EREB11)**

15-CTCAGCATCATCAGGATTCTCG ***Orzo saliva*** **Japonica Group AP2-like - TOE3**

16-CTGCAGCATCATCAGGATTCCC ***Zea mays*** **GLOSSY15 (gl15) genec**

17-CTGCAGCATCATCAGGATTCCC ***Arabidopsis thaliana*** **- TOE3 (TOE3) gene**

18-TGGCAGCATCATCAGGATTCTC ***Arabidopsis thaliana*** **- TOE1 (TOE1) gene**

19-CTGCAGCATCATCAGGATTCGC ***Musa acuminata subsp. malaccensis*** **AP2-like TOE3**

20-CTGCAGCATCATCAGGATTCTC ***Zea mays*** **sister of indeterminate spikelet 1 (sid1)**

21-CTGCAGCATCATCAGGATTCTC ***Zea mays*** **floral homeotic protein mRNA**

22-CTGCAGCATCATCAGGATTCTC ***Setaria italica*** **floral homeotic protein APETALA 2**

23-CTGCAGCATCATCAGGATTCTC ***Hordeum vulgar*** **APETALA2-like protein (AP2L1) mRNA**

24-CTGCAGCATCATCAGGATTCTC ***Zea mays*** **floral homeotic**

25-CTGCAGCATCATCAGGATTCTC ***Musa acuminata subsp. malaccensis*** **floral homeotic protein**

26-CTGCAGCATCATCAGGATTCTC **Sorghum bicolor floral homeotic protein APETALA 2**

27-CTGCAGCATCATCAGGATTCTC ***Oryza sativa*** **Japonica TOE3**

28-CTGCAGCATCATCAGGATTCTC ***Oryza brachyantha*** **floral homeotic protein APETALA 2**

Figure 2.27 miRNA172d binding site in all members of euAP2 genes, showing the alignment between the sequence of miRNA172d from miRbase (www.mirbase.org/) and the query sequence of euAP2.

2.4 Discussion

The phylogenetic tree of bZIP TFs presents the common O2 members, which belong to the bZIP (C-group) in different species of monocots, and dicots such as At BZO2H3 in Arabidopsis, which is close to the O2 members of the monocots. However, there are two paralogous in wheat and barley, which are TaSPA/HvBLZ2 and TaBLZ1/HvBLZ1. The phylogenetic analysis showed that Poaceae species is presented exclusively in a specific lineage, which is named the **A1** group**.** Notably, TaSPA and HvBLZ2 show 77.48% complete sequence and domain identities, and 26.46% with TaBLZ1. Similarly, HvBLZ2 shows identity with HvBLZ1 of around 29.23%. However, the identities between TaSPA and HvBLZ2 to ZmO2 are 77.4% and 28.97%, respectively **(Fig.2.28).** Interestingly, the topology of proteins reflected the fact that these proteins were structurally closer to each other. It was further characterized the O2 homologous proteins by comparing between them in terms of the structure, such as the bZIP domain which is known the longest bZIP domain in the bZIP superfamily.

As mentioned previously, our result confirmed that O2 members included members of monocots and dicots such as Arabidopsis. Lara *et al.,* (2003) confirmed that AtBZIP10 and AtBZIP25 are close to O2 in cereal, and are expressed in the seed of Arabidopsis in which they participate in regulating Seed Storage Protein genes (SSPs), namely albumins and cruciferins. However, those genes are not specific to the seeds, and are expressed everywhere **(Table 2.6).**

A number of similarities and differences were found in the O2 homologous surveyed in this study. One of the similarities between them is they share the longest bZIP domain, but also the fact that they shared conserved motifs. In addition, comparing HvBLZ1 to its orthologous such as OsREB in rice and ZmOHP1 in maize, it can be seen that around 19-20 residues are identical in the basic domain (Schmidt *et al.,* 1992; Pysh *et al.,* 1993; Pirovano *et al.,* 1994). Furthermore, TaBLZ1 is related to TaSPA, parsley CPRF-2 and rice OsRITA-1 (Izawa *et al.,* 1994; Albani *et al.,* 1997). However, from the topology of the phylogenetic tree,

	Hv_BLZ2 extr...	Ta_SPA extra...	Bd_opaque-2-...	Os__O2_like	Hv_Blz1	Ta_BLZ1	Bd_BLZ1	Zm_O2	At_bZIP10_lik...	At_bZIP10 ex...	At_BZO2H3 e...
Hv_BLZ2 extraction		77.481840%	40.235294%	35.827664%	29.230769%	26.956522%	28.907923%	29.978118%	17.621145%	17.937220%	22.058824%
Ta_SPA extraction	77.481840%		41.176471%	34.772727%	28.884026%	26.464208%	27.931770%	28.976035%	17.543860%	18.303571%	22.493888%
Bd_opaque-2-like extraction	40.235294%	41.176471%		30.227273%	26.106195%	23.210412%	24.891775%	26.651982%	17.488789%	17.727273%	17.690418%
Os__O2_like	35.827664%	34.772727%	30.227273%		38.425926%	35.388128%	36.343115%	38.018433%	24.361949%	23.953488%	25.063939%
Hv_Blz1	29.230769%	28.884026%	26.106195%	38.425926%		76.513317%	75.118483%	67.241379%	25.740319%	24.657534%	28.039702%
Ta_BLZ1	26.956522%	26.464208%	23.210412%	35.388128%	76.513317%		65.532880%	60.328638%	21.640091%	20.319635%	25.440806%
Bd_BLZ1	28.907923%	27.931770%	24.891775%	36.343115%	75.118483%	65.532880%		68.396226%	25.512528%	24.657534%	26.555024%
Zm_O2	29.978118%	28.976035%	26.651982%	38.018433%	67.241379%	60.328638%	68.396226%		25.284738%	24.200913%	27.804878%
At_bZIP10_like_1 extraction	17.621145%	17.543860%	17.488789%	24.361949%	25.740319%	21.640091%	25.512528%	25.284738%		87.440758%	28.960396%
At_bZIP10 extraction	17.937220%	18.303571%	17.727273%	23.953488%	24.657534%	20.319635%	24.657534%	24.200913%	87.440758%		30.325815%
At_BZO2H3 extraction	22.058824%	22.493888%	17.690418%	25.063939%	28.039702%	25.440806%	26.555024%	27.804878%	28.960396%	30.325815%	

Figure 2.28 The percentage of identity between some members of bZIP.

TaBLZ1 is close to ZmO2 maize than its paralogous (TaSPA and HvBLZ2), respectively. However, TaBLZ1 has not previously been studied in wheat. In addition, bZIP domain was confirmed that it has dual functions in terms of the DNA binding and nuclear localization (Varagona and Raikle, 1994). It has been reported that some residues of the DBD of AtbZIP transcription factors especially at position 11, 15 and 19 revealed the presence of all serines which are likely to have potential phosphorylation sites in all C-group members, allowing to control activation of the bZIP factor at the DNA–protein interaction level (Kirchler *et al.*, 2010). A previous evidence suggested that most members of the O2 family in Arabidopsis share certain features with known bZIP proteins that include maize (O2) and parsley (CRF2), such as the length of the leucine zipper with up to nine heptad repeats and potential target sites for protein modification such as phosphorylation sites (kircher, 2010). Another significant feature is the nuclear localization signal (NLS) plays an important role because some proteins, which could be regulated in the nucleus at specific developmental times and in specific tissues. A possible explanation for the importance of NLS is that some regulatory proteins remain in the cytoplasm until their function in the nucleus becomes necessary (Raikhel, 1992). The observation showed the basic region has conserved NLS in all the O2 proteins. It was confirmed the function of NLS of some members of bZIP, for example, Raikhel, (1992) reported that NLS was a part of the basic DNA-binding domain of O2, which was

Table 2.6 Members of bZIP (C- group) of monocots and dicots (Adapted from Wang *et al.,* 2013). The highlighted genes were added in this study.

bZIP C-group	Gene	Species	Binding site	Target gene	Expression	Phenotypes	Reference
MONOCOTS	SPA	Wheat	GCN4 like motif (TGASTCA)	LMWG-1D1	Seed-specific	NA	Albani *et al.*, 1997
	BLZ1	Wheat				NA	-
	BLZ2	Barley		Hor-2	Endosperm-specific	NA	Onate *et al.* , 1999
	BLZ1	Barley		Itr1	Endosperm, roots and leaves	NA	Vicente-Carbajosa *et al.,*1997
	Opeque2 (O2)	maize		a-zein, 32 kDa albumin(b-32)	Endosperm-specific	Soft and chalky endosperm with high lysine and tryptophan	Lohmer *et al.*, 1991; Schmidt *et al.,* 1992.
	BZO2H2	Zea mays					
	OHP1	Zea mays		a-zein	General expression	No obvious phenotype	Pysh *et a*l.,1993; Zhang *et al.,*2015
	OHP2	Zea mays		a-zein	General expression	No obvious phenotype	Pysh *et al.*1993; Albani *et al.,*1997
	OsbZIP88/RIZB1	Rice		OsLKR/SDH	Endosperm-specific	Higher lysine content	Kawakatsu and Takaiwa, 2010; Kawakatsu *et al.*,2009, Onodera *et al.,* 2001
	OsbZIP15/RISBZ4	Oryza sativa Indica Group		SSPs	Endosperm-specific	NA	Takaiwa *et al.,* 2007
	OsbZIP20/RITA-1/RISBZ3	Oryza sativa Indica Group		SSPs	Endosperm-specific	NA	Izawa *et al.,* 1994
	OsbZIP52/RISBZ5	Oryza sativa Indica Group			Universal, increase in the seeds	involved in stress response, OsbZIP52 overexpression in transgenic lines enhanced seedling sensitivity to cold and drought stress	(Liu *et al.,* 2012)
	OsbZIP33/ /RISBZ2/REB	Oryza sativa Indica Group		a-globulin Wx and SBE1 and is involved in starch synthesis	Endosperm-specific	NA	(Nakase *et al.* 1997) (Cai *et al.,* 2002)
	OsbZIP58/RISBZ1	Oryza sativa Indica Group		six starch-synthesizing genes, OsAGPL3, Wx, OsSSIIa, SBE1, OsBEIIb, and ISA2,	Endosperm-specific	abnormal seed morphology with altered starch accumulation in the white belly region and decreased amounts of total starch and amylose.	Wang *et al.,* 2013
DICOTS	AtbZIP9/BZO2H2	Arabidopsis *thaliana*		-	Universal	overexpressing AtbZIP9changes in the phloem developmental process	Weltmeier *et al*, 2009 Silveira *et al.* 2007
	AtBZIP10/BZO2H1	*Arabidopsis thaliana*		At 2S albumins	Universal	NA	(Lara *et al*, 2003)
	AtbZIP25/BZO2H4	*Arabidopsis thaliana*		At 2S albuminsN	Universal	NA	(Lara *et al*, 2003)
	AtbZIP63/BZO2H3	*Arabidopsis thaliana*		-	Universal	NV	(Satoh *et al*, 2004)

confirmed by identifying NLSs in two other b-ZIP tobacco proteins, TGA-1A and TGA-1B, which are likely to have a bipartite structure (van der Krol and Chua, 1991).This would be suggested that NLS in bZIP has two roles which are import to the nucleus, and another role is DNA binding. Until now, NLSs had been described in only a few studies of other members of the bZIP protein family.

By drawing on the extensive studies of bZIP and sequence analysis of O2, the result in this chapter showed that two monocot regulatory proteins, Ta/HvBLZ1 were confirmed to be in both the nucleus and cytoplasm, indicating that the basic region might be responsible for localizing TaBLZ1 proteins in the nucleus. However, the previous study conducted by (Varagona and Raikle, 1994), it was shown that the basic region was highly conserved in the most bZIPs in the same class, suggesting that it has a bi-functional role (DNA binding and translocation). Also, this study confirmed the NLS of maize OPAQUE-2 in the basic domain showed that positions 210 (R) and the 222 (R), 223 (K) are required for an efficient translocation process. Also, it has been reported that one single conservative change in the amino acid of the basic region does not have any effect on localization.

The C-terminal of the bZIP domain contains a leucine zipper, which contains variously repeated heptads (hydrophobic residues) that may play an important role in dimerization. The presence of attractive or repulsive interhelical g↔e electrostatic interactions and the presence of polar or charged amino acids in a and d positions of the hydrophobic interface of the leucine zipper region can be conserved within O2 members especially in the monocots. It is noteworthy that the position and the types of heptads are highly conserved between the members of O2 in monocots. Most of them have charged amino acid residues at position g, while leucine are aligned in d position. All these arrangements of charged amino acid residues are important for dimerization (Vinson and Boyd, 1993).

Another point is detection the conserved motifs which may have important functional roles in these factors. Some of these motifs may participate in the transcriptional activation (Vicente-Carbajosa *et al.,* 1997; Onodera *et al.,* 2001). Furthermore, these motifs are often conserved between members of groups or

subgroups of the large family of TFs that have the same functions (Nakano *et al.,* 2006) which are correlated to the topology of these proteins in the phylogenetic tree.

On the other hand, phylogenetic analysis was based on euAP2 proteins confirmed the relationship between the RSR1 and Q as paralogous, which can be considered a specific lineage in the monocot. In addition, it illustrates the relationship between these proteins and the other euAP2 in the monocots which have been identified, such as putative orthologous of Q have been identified in maize, rice, and barley (Faris *et al.,* 2003). In maize, indeterminant spikelet1 (ids1), which plays a role to determine the number of floral meristems produced (Chuck *et al.,* 1998), but functions have not yet been assigned to the orthologous in rice and barley. In dicots, euAP2/TOE3 in Arabidopsis which were found similarities between them in terms of protein sequences and their functions (Zumajo-Cardona, and Pabón-Mora, 2016). All these genes share a highly conserved AP2 domain. Jofuku *et al.,* (1994) confirmed the functional importance of the AP2 domain involved in processes such as the specification of the floral meristem identity (Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993), the identity of the sepal and petal organs (Haughn and Somerville, 1988; Kunst *et al.,* 1989; Bowman *et al.,* 1989, 1993). Some *euAP2* genes involve in different aspects of seed development such as embryonic growth and seed germination, and flower traits, cell identity, flowering time and floral meristem identity (Krizek *et al.,* 2000; Boutilier *et al.,* 2002; Aukerman and Sakai, 2003; Cernac and Benning, 2004).

Furthermore, one of our goals has been identified protein domains in details, particularly (DNA binding site, linker, and the N-terminus and C-terminus of AP2 domain), conserved motifs and the functional domain within the *euAP2* genes clade. The results showed the euAP2 members are highly conserved in different ploidy wheats and other of monocots than in the dicots, also they share highly conserved motifs such as the AP2 domain, and (AASSGF) motif. Additionally, several conserved motifs were identified and described; most of these motifs were widely distributed among euAP2 members. Several motifs were found across

monocots and dicots are likely to have roles such as repression domains (LxLxL) motif is present in all euAP2 proteins examined. It has been suggested that this motif is a repression domain in *euAP2*/TOE3 genes (Zumajo-Cardona and Pabón-Mora, 2016). This motif in *euAP2*/TOE3 genes is possible that involves the recruitment of TOPLESS (TPL) and TPL-related co-repressors (Ohta *et al.,* 2001; Causier *et al.,* 2012). Another motif is (MLDLNV) was found in the N-terminal of *euAP2* proteins, exhibits as transcriptional activity especially has a role in protein-protein interaction. Another conserved motif in all euAP2 sequences which was identified as miR172 binding site. Aligned sequences of *euAP2* genes present conserved miRNA172 binding sites in these genes with a single mismatched nucleotide in the 3' end. So that, The miR172 target site within euAP2 can be categorized into two groups, the first being the genes which are perfectly complementary to miR172d such as *RSR1*, while the other has a single nucleotide mismatch at position 20 of the miRNA binding site, such as *TaQ*, with the latter expected to be effective as cleavage agents. The miR172 target site within the cereals is completely conserved. Also, it was confirmed that a single nucleotide mutation at position 20 (3' end) of the microRNA binding site in mRNA disrupts the functionality (Brown and Bregitzer, 2011).

All these features confirmed that there is a change in the protein sequences which occur after the duplication event, which may explain the fact that the role of the duplication event for functional diversification. In addition, it has been reported that those gene duplicates that result from polyploidy, redundancy followed by sub-functionalization has proven to be a regular trend (Conant *et al.,* 2014).

Chapter 3

Gene expression analyses

Abstract

Background: To achieve a better understanding of the role of TFs and their regulating mechanism, it is better to focus on the expression profiles of O2 members (TaSPA/TaBLZ1) ;(HvBLZ2/HvBLZ1) during the grain development stages and other tissues. Expression patterns were assayed in different temperate species with known have differences in grain quality profiles to see if their expression could be correlated with grain composition. The aim of this chapter is to investigate the spatial and temporal expression patterns of these genes which could reflect their function.

Results: Detailed spatial and temporal expression patterns were confirmed by RT-PCR, qRT-PCR and mRNA-ISH approaches. I examined the differential expression of genes either during the grain development for TaSPA and HvBLZ2 which confirm these genes are endosperm-specific TFs, while their paralogous (TaBLZ1, HvBLZ2) were expressed in different tissues of species. All these genes show conserved spatial expression with a slight difference in their expression level in examined species, which positively correlated with the protein accumulation. The second group of euAP2 genes (*RSR1*, *Q*) were detected in their expression during the life cycle of the plant. In addition, it is known euAP2 genes were regulated negatively by miR172. For this reason, I compared the expression level of miR172 and (*RSR1*, *Q* genes) in different species during the life cycle of the plant which reflects the reverse correlation between them.

Conclusion: I summarized the temporal and spatial expression of TFs which have been shown to play important roles in controlling the accumulation of starch and protein in the endosperm during the development of grain in different species. The overlap of the expression patterns of TaSPA/HvBLZ2 and TaBLZ1/HvBLZ1 confirmed their role in controlling the accumulation, also TaBLZ1/HvBLZ1 could involve in the formation of the protein bodies at an early stage. While the euAP2 paralogous although share several expression patterns, they have specific patterns which reflect their role. The results obtained are of interest to understand the roles of the bZIP and *euAP2* genes during the development of grain.

Chapter 3 Gene expression analyses

3.1 Introduction

During cereal grain's development, the storage protein and starch accumulate in the grain. Seed storage proteins (SSPs) and starch are synthesized in the endosperm of cereal grains or embryo of dicots. Seed storage proteins are known to be specific in the cereal grain endosperm and accumulate highly during the development of grain in the endosperm. The concentration and composition of seed storage proteins (SSPs) and starch in wheat flour are the most important determinants of its end-use value. The comparison of the quality of varieties of grains shows significant variation among the cultivars species of wheat and barley which are grown in the same environment. This is at least partially due to variation in the composition of storage protein and starch in the different grain types.TFs are important candidates for studying the mechanism behind the regulation of specific genes in the endosperm which affect grain quality.

This chapter aims to describe and compare the key regulators controlling the accumulation of starch and storage proteins in grain in different species of Poaceae.The first is bZIP members TaSPA/HvBLZ2 which have been characterized as specific TFs in the endosperm, also TaBLZ1/ HvBLZ1 (Albani *et al.,* 1997; Vicente-Carbajosa, *et al*., 1997; Vettore *et al*., 1998).

The second is RSR1 (Rice Starch Regulator1), it belongs to euAP2 clade of the APETAL2/ERF protein family. It was identified in rice, where it was determined to down-regulate the expression of some starch synthesis genes in the endosperm (Fu *et al.,* 2010). Furthermore, as described previously RSR1 is paralogous to *Q* that is known the main domestication gene in wheat regulating several aspects of development including grain shape and size (Simon *et al.,* 2005).

To gain a deeper insight into the expression patterns, the analysis was extended from RT-PCR analysis to qRT-PCR survey in the different grain development stages in cultivated hexaploid, tetraploid wheat species, barley, Rosner and rye. Also, the similarities and differences were identified between

them in terms of the spatial and temporal expression by using mRNA in situ hybridization in different sections of the grains and spikelet.

Another point is that, since the identification of miRNAs in plants it has become clear that they have the ability to regulate many aspects of plant development and responses to the environment. One of the small RNAs is a miR172 was one of the earliest plant miRNA genes to be identified by small RNA cloning and sequencing in Arabidopsis (Park *et al.,* 2002). It is approximately a 21-nucleotide mature miR172 that regulates a small group of the plant-specific transcription factor gene APETALA2 (AP2) which is *euAP2* (Chen, 2004). This a group of AP2-like genes include different genes of monocots such as *RSR1* and *Q*, in dicots such as TARGET OF EAT1 (*TOE1*), *TOE2, TOE3*, SCHLAFMUTZE (*SMZ*), and SCHNARCHZAPFEN (*SNZ*) in Arabidopsis (Aukerman and Sakai, 2003; Schmidt *et al.,* 2003; Schwab *et al.,* 2005). The role of miR172 was investigated in many species such as Arabidopsis, maize, barley, and soybean, which can control flowering time (Jung *et al.,* 2007; Mathieu *et al.,* 2009; Nair *et al.,* 2010; Yoshikawa *et al.,* 2013). In Arabidopsis and maize, the expression of miR172 increases steadily during the life of plant corresponding the age of plant (Aukerman and Sakai 2003; Lauter *et al.,* 2005). However, miR172 in rice is highly expressed at the late vegetative stage and during the development panicles as well (Zhu *et al.,* 2009; Lee and An, 2012). miR172 accumulates in the leaves and flower buds over time (Fahlgren *et al.*, 2006; Wu *et al.*, 2009; Nodine and Bartel, 2010; Zhu and Helliwell, 2010). The temporal expression of miRNAs in specific patterns during the life cycle of plant reflect the ability of miRNAs to control the transition from juvenile to adult phase and following by reproductive phase *(Huijser and Schmid, 2011;* Yamaguchi and Abe, 2012*).* Comparing between miR172 and other of miRNAs, it can recognize the level of the expression of miR172 increases after flowering stage in both dicotyledons and monocotyledons (Aukerman and Sakai, 2003; Lauter *et al.,* 2005; Chuck *et al.,*2007; Jung *et al.,* 2007; Zhu *et al.,* 2009).
Consequently, the aim of this chapter is focusing on the expression analyses of bZIP and *euAP2* genes, by introducing the comprehensive survey of the expression of TFs in different tissues in different species which could reflect the function of these TFs during the development the grain or in other tissues.

Moreover, the expression levels of miRNA172 and its targets (euAP2) in different development stages were compared to investigate the relation between them.

3.2 Material and Methods

3.2.1 Plant material

Various species of cereals such as different varieties of cultivated hexaploid wheat (*Triticum aestivum*), (*Triticum aestivum cadenza*), (*Triticum turgidum durum*), Barley (*Hordeum vulgare*), rye (*Secale cereale*), and Rosner (*Triticale hexaploidiL att*.), using different samplels such as grain, vegetative tissues and reproductive tissues **(Appendix 7. Fig.7.3.1).**

3.2.2 Seed sowing and growth conditions

All grains were germinated on moist filter paper at 4°C for 48h. After Five days the seedlings were placed into 9 cm square pots with 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.

3.2.3 Tissue collection, disruption and storage

Leaves (11-days after sowing) were collected for DNA extraction, then flash frozen in liquid nitrogen and stored at - 80°C until required. Other materials were used for RNA extraction (**Appendix 7.Fig.7.3.1)**, were snap frozen in liquid nitrogen and stored at -80°C. For mRNA in situ hybridization, some tissues were immediately immerged in FAA (3.7% formaldehyde, 5% acetic acid, 50% ethanol). Samples were vacuum infiltrated in 3 x for 5 min. and leave them overnight at 4°C with continuous agitation. Tissues were washed with 70% EtOH and stored at 4°C.

3.2.4 Genomic DNA isolation

To extract gDNA by using 300 mg of a young leaf tissue using a typical phenol: chloroform extraction protocol (Kang and Yang, 2004). The leaves were placed into a 1.5 micro-centrifuge tube containing ~200 µm glass beads (Sigma-Aldrich). Leaf samples were flash frozen in liquid nitrogen and were grinded for 12 seconds in a Silamat amalgam mixer (Ivocar Vivdent, UK) at room temperature. 400 µl of DNA extraction buffer was added and samples were incubated in 50 °C for 20-30 min, followed by a phenol: chloroform extraction

and precipitation using isopropanol. DNA pellets were washed with 70% ethanol, re-suspended in sterile water, and stored at -20ºC.

3.2.5 Isolation of total RNA.

Total RNA was extracted from frozen grinded tissues using TriSure (Bioline) according to manufacturer's instructions. For grain samples that contain more starch, the Isolation II RNA plant kit (Bioline) was used according to manufacturer's instructions. Then, RNA samples were treated with DNase enzyme was performed using DNase I (NEB), with a 10 min incubation at 37°C and enzyme deactivation at 75°C for 10 min. Finally, samples were stored at -80ºC

3.2.6 RNA quantification and cDNA synthesis

RNA samples were quantified using a NanoDrop™ 2000 UV-Vis spectrophotometer (Thermo Scientific) and RNA integrity was checked in a 1XTAE agarose gel by agarose gel electrophoresis. Samples were stored at -80ºC until required. For cDNA synthesis, 700 ng RNA was used in 20 µl cDNA synthesis reactions using Tetro Reverse Transcriptase (Bioline) and a poly (dT) primer, according to the manufacturer's instructions. 0.7µg RNA sample, 2µl Oligo (dT) primers (0.5µg/µl, Promega), 1µl dNTP (10mM, Invitrogen), and make up with nuclease-free water to 20 µl, then Incubated in PCR thermal cycler T100 (BIO-RAD) at 65°C for 5 minutes to denature the secondary structures of RNA potentially formed in samples. The second step is to prepare master mix 2: 8µl 5X First Strand Buffer (Invitrogen), 4µl Dithiothreitol (DTT, 0.1M, Invitrogen), 1µl RNasin Plus RNase inhibitor (40U/µl, Promega) and 1µl SuperScript III reverse transcriptase (200U/µl, Invitrogen). The mixture of master mix 1 and Master Mix 2 was incubated at 50°C for 50 minutes (for reverse transcription) and 70°C for 15 minutes (for inactivation of reverse transcriptase).

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

Specific primers of three homologous genes of wheat, H genome of barley and R genome of rye were designed for Reverse transcription-polymerase chain reaction (RT-PCR) survey to detect the expression patterns of TaSPA/HvBLZ2, and Ta BLZ1/HvBLZ1 in different stages of grain development as shown in **(Fig 3.1).**

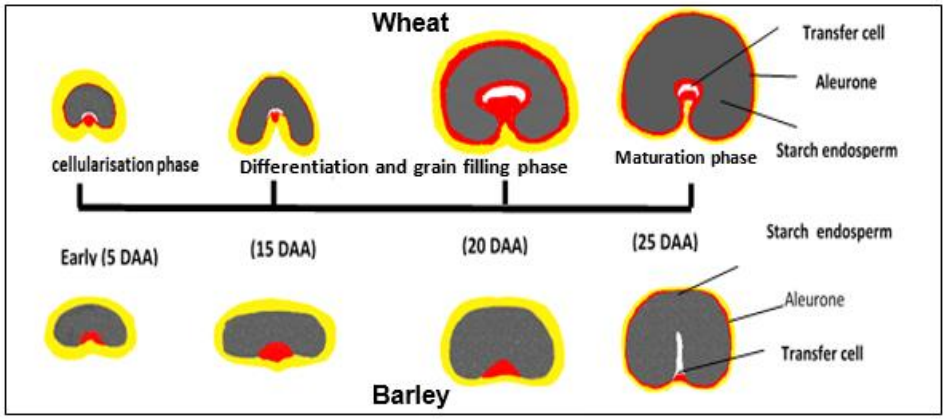


Figure 3.1 Diagram shows the main stages of grain development comparing between the wheat and barley.

The synthesized cDNA from various tissue samples were used with gene-specific primers, alongside the reference genes such as GAPDH, or ACTIN. The sequences of the primers used can be found in (**Appendix.3; Table 7.3.2).** (Table 3.1) shows details of the PCR reaction and programme.

Table 3.1 RT-PCR reaction and thermo-cycling program to investigate gene expression.

1x PCR reaction	Volume
DreamTaq Green PCR Master 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 programme	Temperature	Time (min)	
Initial denaturation	94	05:00	
Denaturation	94	00:30	25 cycles
Annealing	Variable	00:30	
Extension	72	00:30	
Final extension	72	05:00	
Hold	4	00:00	

3.2.8 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 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.

3.2.9 Quantitative (qRT-PCR) analysis

3.2.9.1 *Selection of Reference Genes*

I selected the known housekeeping genes were chosen based on previous gene expression studies and/or by the database such as (NCBI) (http://www.ncbi.nlm.nih.gov.). To identify barley homologous a “blastn” (NCBI) was run with reference genes sequences on NCBI database using the default settings of the on line program and the full-length sequence selected.

3.2.9.2 *Optimization of Primers for Target Genes and Reference Genes*

Oligonucleotides for cDNA amplification were designed based on predicted sequences of wheat and barley, which are available from Ensemble plant (www.ensembleplant.org), or NCBI (http://www.ncbi.nlm.nih.gov/gene) databases. PCR primers of SYBR Green qPCR assays were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/) (Rozen and Skaletsky, 2000). For design primers, we follow some parameters as follows: amplicons were designed to be 80 to 200 bp; primer melting temperature (Tm) ranged from 57°C to 60°C; the length of primer around 18 to 20 bp; and primer GC content around 50%. Primer efficiencies of both target genes and reference genes were calculated by using five or 10-fold cDNA dilutions for example, (1:1, 1:5, 1:10, 1:100 and 1:1000) in duplicate as well as checking for amplification in a negative control without template. Standard curves were constructed by the Ct (y-axis) versus log cDNA dilution (x-axis). The primer efficiency (E) of one cycle in the exponential phase was around 80% - 100%. In terms of the wheat primers, they were designed to get the transcription profiles which represent the integration of the transcript levels of three homologous of each gene. (**Appendix.3; Table 7.3.2).**

3.2.9.3 *SYBR Green method*

Each target gene along with the reference genes (GAPDH, ACTIN) and negative control (no template) were analysed in one plate. In this assay, SYBR green was used to quantity the genes expression in different tissues. The qPCR Master Mix reaction was performed in 20 µL by following reaction components:

containing serial dilation of cDNA, 10µl QuantiFast SYBR Green qPCR SuperMix (QIAGEN), 1.25µl forward primer (10µM, Sigma aldrich), 1.25µl reverse primer (10µM, Sigma aldrich), and 6 µl nuclease-free water reactions were performed in triplicate with the test cDNA sample(s), and separate triplicate reactions were performed using a reference genomic DNA. Master Mix (10µl) was filled in 96 well plate (Applied Biosystem), followed by adding of 0.7µg cDNA as PCR templates. Wells were closed with strip caps, centrifuged briefly to bring contents to bottom of wells and placed into PCR instrument (Applied Biosystem).

The following PCR programme was used: denaturation cycle (95°C for 10 minutes), amplification programme repeated 40 cycles (denaturation: 95°C for 15 seconds, annealing: 60°C for 20 seconds, 72°C for 60 seconds with a single fluorescence measurement), dissociation programme (95°C for 60 seconds, 55°C for 30 seconds, 95°C for 30 seconds with a continuous fluorescence measurement) and a cooling programme to 40°C.

3.2.10 mRNA in situ Hybridization (ISH) method.

3.2.10.1 *Tissues fixation and processing*

The plant tissues were processed as follows:

Serial dilution of ethanol (70%, 85%, 95%, 100%) 1 hour at 4 °C.

100% EtOH for 1 hour at room temperature (RT).

Serial dilution of [Histoclear (Agar Scientific) (H/C) in EtOH] (25%, 50%, 75%) 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 overnight at 60°C; 100% fresh paraffin for 8 hours at 60°C and a final change with 100% paraffin for overnight at 60°C.

3.2.10.2 *Sectioning and embedding*

All plant tissues were embedded in paraffin using Tissue Embedding Molds (Polyscience, USA). A Bright 5030 (Bright instrument Co., England) microtome was used for the section. The sections were normally cut in a thickness of 10-14 µm, organized on poly-lysine-coated slides (Grace Biolab, Supplied by Stratech Scientific, Soham, UK).

3.2.10.3 *RNA probe synthesis*

The probe template was produced by PCR using cDNA as template and gene-specific primers. Primers were designed amplifying the specific region of each gene. The reverse primers incorporated a T7 promoter region for subsequent in-vitro transcription **(Appendix3; Table 7.3.2).** 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 thermo-cycling conditions for mRNA probe template synthesis

PCR reaction (50µl)	Volume
H2O	21.5 µl
5x Q5 Reaction Buffer	10 µl
GC Enhancer	10 µl
Forward primer (10 µM)	2 µl
Reverse primer with enzyme attachment (10 µM)	2 µl
dNTPs (10 µM)	1 µl
cDNA template	3 µl
Q5 High-Fidelity DNA Polymerase	0.5 µl

PCR Condition	Temp.	Time	
Initial denaturation	98	00:30	
Denaturation	98	00:10	25 cycles
Annealing	55	00:20	
Extension	72	00:30	
Final Extension	72	02:00	
Hold	4	∞	

3.2.10.4 *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 mixture 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 reaction was 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.

3.2.10.5 *Slide Pre-treatment*

Slide pre-treatment was performed following the protocol of Drea *et al.,* (2005).

3.2.10.6 *Hybridization and Washing*

These steps were performed using the method was described by Drea *et al.,* (2005).

3.2.10.7 *Detection*

The slides were soaked in AP buffer for environment equilibration for 2-3 min at RT. The slides were then soaked in developing the solution (15 µl NBT-supplier by Sigma Aldrich (0.2 mM) and 15 µl BCIP (0.2 mM) in 10ml AP Buffer and were left in darkness.

3.2.11 Expression Analysis of miR172

3.2.11.1 *miRNA isolation*

The mirVana miRNA Isolation Kit (Thermo Scientific) was used to isolate the small RNA separately from the total RNA from different plant tissues following to manufacturer's instructions.

3.2.11.2 *RNA quantitation and quality assessment*

To assess the quantity of small RNA, using NanoDrop 2000/2000c (Thermo Scientific). The ratio of A260 to A280 provides an indication of RNA purity. An alternative way to estimate RNA quality and quantity is Bioanalyzer (Agilent 2100) was used, and the small RNA can be visualized.

3.2.11.3 *cDNA synthesis*

To synthesize cDNA of small RNA, using QuantiMir Kit (System Biosciences) according to the manufacturer's instructions.

3.2.11.4 *Primer Design for cDNA*

To perform end-point or qPCR reactions, specific primers were designed for miRNA. The forward "sense" orientation primer was designed only. The size of microRNAs typically ranges from 19 – 24 nt. The sequence for forward primer is exactly for the sequence of the target of miRNA. To confirm the sequence of the typical miRNA, the miRBase database can be used

(http://microrna.sanger.ac.uk/sequences/search.shtml) (Kozomara and Griffiths-Jones, 2013). The mature miRNA sequence can be used simply after converted to a DNA sequence. However, the reverse primer was provided with the kit.

3.2.11.5 *End-point PCR Reactions*

The following PCR reaction conditions as shown below:

1 µl of cDNA
0.5 µl Universal Reverse Primer (10 µM)
1 µl miRNA-specific Forward Primer
2.5 µl 10X PCR Buffer with 2.5 mM MgCl2
1 µl 10 mM dNTPs mix
20 µl RNase-free water

The total volume = 25 µl

PCR Programme:

Heat denature at 95°C, 10 min.
Heat denature at 95°C, 15 sec.
Anneal Primers at 60°C, 1 min. Hold at 15°C (optional)

20 cycles

Prepare and pour a 3.5% agarose (1X TAE) gel with (Ethidium Bromide). Add 2.5 µl of 10X Loading dye, mix and then load 10 µl of each sample into a well of the gel. Also, run a suitable DNA size marker (50 bp) along with the samples.

3.2.11.6 *qPCR Reaction Setup*

To quantify the expression profile for miRNA, qPCR (Applied Biosystems) was used, following this reaction:

15 µl 2X SYBR Green qPCR Master mix buffer
1 µl User-designed Forward Primer (10 µM)
0.5 µl Universal Reverse Primer (10 µM)
1 µl Diluted cDNA
12.5 µl RNase-free water

30 µl Total / well

Load all the samples into the wells of the plate (Applied Biosystems), cover the plate with the optical adhesive cover.

3.2.12 Statistical Analysis

For the normalization, the relative expression of target genes, the cycle of threshold (Ct) for each gene transcript was used for quantification of gene expression. Ct value was obtained when the fluorescence rises appropriately above the background fluorescence. The baseline signal and threshold signal of fluorescence were determined automatically by the PCR machine. The calculation was conducted in Microsoft Excel 2007. According to the delta delta-Ct method (Livak and Schmittgen, 2001), the relative expression was calculated in relation to the expression of reference genes and was defined as relative expression of a target gene in different samples by using the following equation:

RE = 2 ^ (-(ΔCt sample (target - reference))

ΔCt sample is the Ct difference of target-reference genes.

It was used to normalize and calibrate transcript values the relative to the endogenous GAPDH/ ACTIN controls which have efficiency around (95-100%). All the statistical analyses were performed using the graphing and statistical software package GraphPad Prism 7.0. The data was normal distributed, therefore, the parametric of two-way analysis of variance analysis (ANOVA).

3.3 Results

3.3.1 Expression profiles of candidate genes survey using RT-PCR

3.3.1.1 *(TaSPA, TaBLZ1/HvBLZ2, HvBLZ1); O2 expression.*

The results of the RT-PCR profiling for *TaSPA* and its orthologous Hv*BLZ2* confirms these genes have similar spatial expressions, both are expressed in the late stages of grain development in wheat and barley. There is no expression in the early stage **(Fig 3.2).** This result is in agreement with Albani *et al*., (1997) findings, which showed that TaSPA is a specific transcription factor in the endosperm which was highly expressed during the filling stage. However, there was slight variation in the expression of *TaSPA* was detected during the grain development stages in different genotypes. By comparing between different varieties of wheat and other species related to wheat such as Rye and Rosner, *TaSPA* was expressed in the mid-length in the rye and decreased sharply at the end of the filling stage, while *TaSPA* in Rosner was expressed in the full length only.

TaBLZ1/HvBLZ1 is known as paralogous *TaSPA* wheat and *HvBLZ2* barley. It was expressed in the grain but the expression of TaBLZ1 started at the early stage during grain development until the grain fully mature **(Fig.3.2).** It was noticed the spatial expression of *BLZ1* has similar patterns in all different varieties of wheat, barley and other species. In addition, *TaBLZ1* expression was revealed in different tissues (vegetative tissues such as root, seedling and flag leaf), or reproductive tissues (ovary, anther, palea and lemma). It was found that TaBLZ1 was expressed at the consistent level **(Fig3.2).**

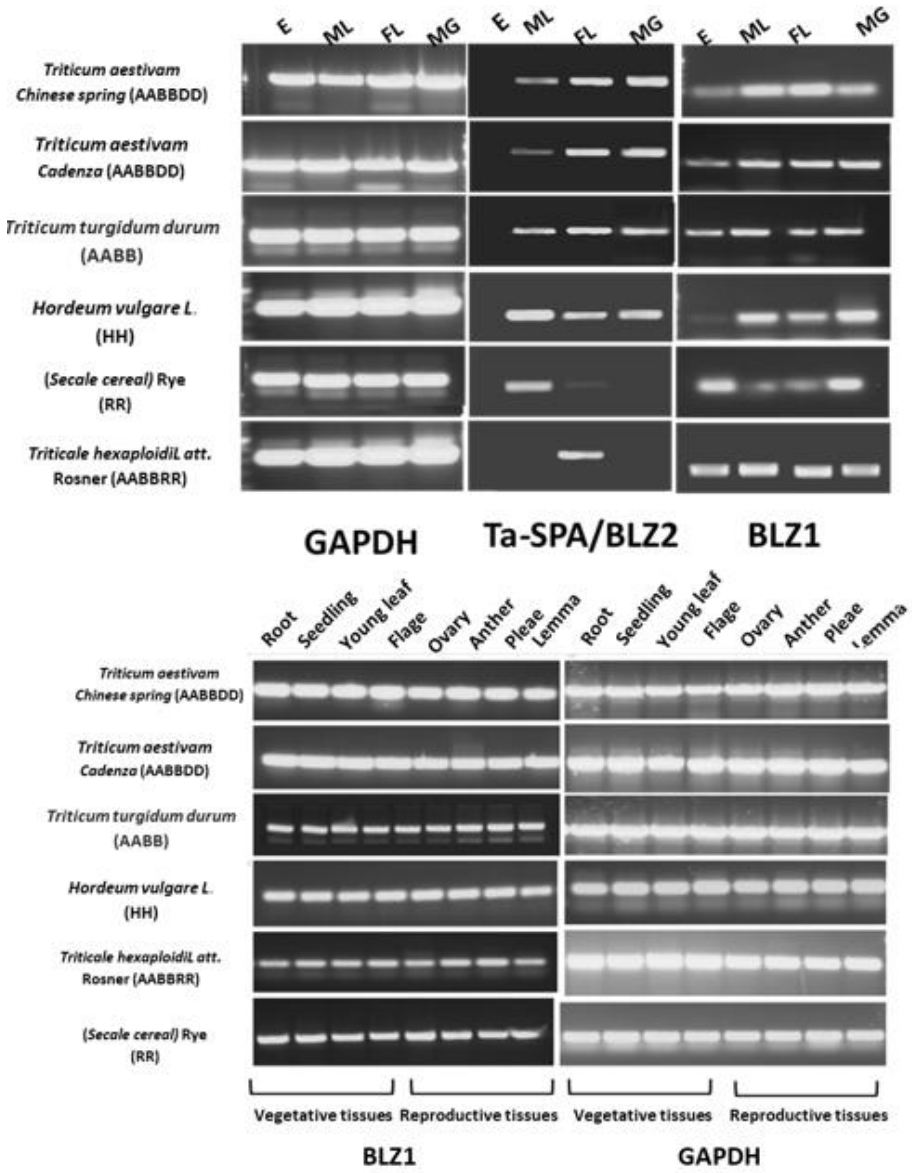


Figure 3.2 RT-PCR analysis of *TaSPA/HvBLZ2* and *TaBLZ1/HvBLZ1* expression patterns in the grain development stages in different species. (E, early stage; ML, mid length; FL, full length; MG, mature grain).

3.3.1.2 *(RSR1, Q); euAP2 clade expression*

The results showed that *RSR1* and *Q* were expressed in vegetative tissues and in different reproductive tissues **(Fig3.5).** A comparison between the expressions of the two-paralogous revealed that both have constant expression in different vegetative and reproductive tissues. However, both showed specific expression patterns such as *RSR1* was expressed during the development of grain **(Fig3.3),** while*TaQ* was expressed in different the development stages of spikelet meristem **(Fig3.4).**

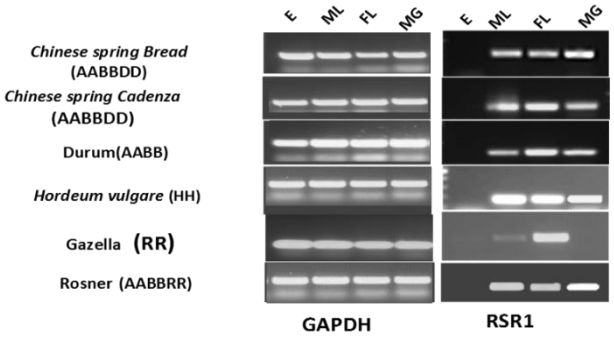


Figure 3.3 RT-PCR analysis of *TaRSR1/HvRSR1* expression patterns in the grain development stages in different species. (E, early stage; ML, mid length; FL, full length; MG, mature grain).

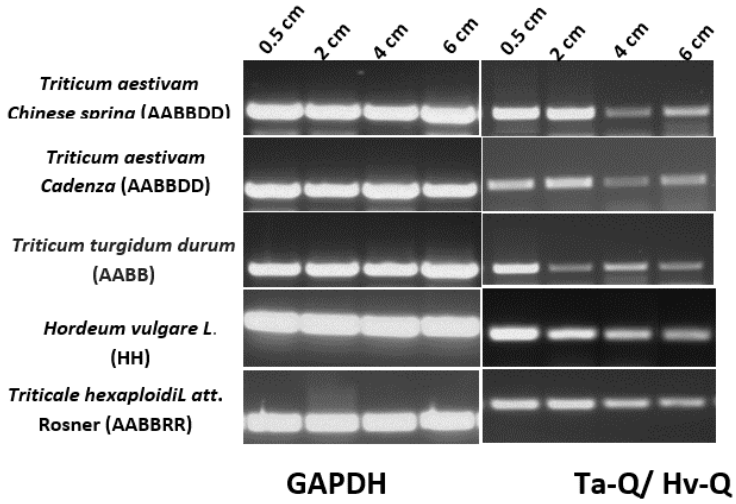


Figure 3.4 RT-PCR analysis of *TaQ/HvQ* expression patterns in different stages of development of the immature spikelet depending on the length, starting with (0.5 cm, 2cm, 4 cm and 6cm).

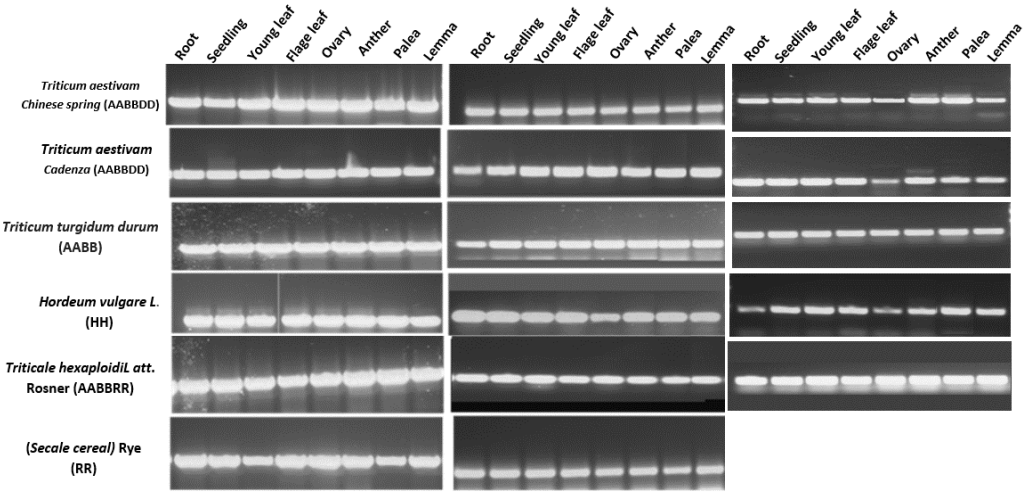


Figure 3.5 RT-PCR analysis of *TaRSR1/HvRSR1* and *TaQ/HvQ* expression patterns in the vegetative and reproductive tissues (Root, Seedling, Young leaf, Flage, Ovary, Anther, Palea and Lemma).

3.3.2 quantitative expression analysis of candidate genes using (qRT-PCR)

3.3.2.1 *(TaSPA/ TaBLZ1)(HvBLZ2, HvBLZ1); O2 family*

Further analysis was performed using quantitative Reverse transcription-polymerase chain reaction (qRT-PCR) (Applied biosystem).

To assess relative gene expression in the different species, the amount of each transcript of interest in a cDNA sample was expressed relative to reference controls transcripts. These reference controls were used, as they are likely to be expressed constitutively in all varities of wheat and barley. Ta/HvGAPDH and Ta/HvActin were used for wheat and barley to normalize the target genes and calculate their relative expression levels, which include an efficiency of the individual transcripts. The amplification efficiency was established for each of the targets from a 10 fold serial dilutions. Mean values and SDs were obtained from three biological replicates.

The result showed that no transcripts from *TaSPA/HvBLZ2* were detected by qRT-PCR in early stage (0-5 D.A.A) grain. In addition, *TaSPA* and *HvBLZ2* have similar expression level during the filling stage of grain (20 DAA) which were expressed highly in this stage **(Fig.3.2)**. Nevertheless, there is a variation in the expression level between TaSPA and HvBLZ2 in different species and *HvBLZ2* barley. By comparing the hexaploid and tetraploid wheat, it emerged that the expression level in the tetraploid is higher than in the hexaploid wheat which correlated with the amount of protein in each one. However, there is also variation in the expression between the hexaploid themselves, though they are more similar compared to others. However, the expression in the Rosner was performed by using the same primers which were designed for the three homologous which are present in all species of wheat. The results showed that the expression of TaSPA in Rosner was the lowest in the full length compared to others. All these variations in the expression level of *TaSPA* or *HvBLZ2* reflect the variations in the level of storage protein in different grains.

However,*TaBLZ1/HvBLZ1* are expressed in all grain development stages **(Fig.3.3)**. The expression starts at early stages of development, peaks at this stage and declines slightly during the filling stage, while *TaBLZ1* expression

drops sharply at the maturation stage especially in the tetraploid and barley. However, by comparing between the species, it was found that nearly all the species have similar expression levels in the mid-length and full-length stages, especially between the hexaploid. However, the expression in the barley shows difference during the full-length which is very high compared to others. The strong expression of *TaBLZ1* across grain development correlated with the period of most intense maternal tissue growth and endosperm cellularisation and continues during the filling stage and maturation as it is reported so far as functioning primarily in grain development stages (Diaz and Carbonero, 1998). It is apparent from the expression rate of TaSPA/HvBLZ2 during the development of grain in examined species are correlated with the temporal and spatial expression of seed storage protein genes which increased gradually during the grain development and drop in the maturation.

To conclude, the paralogous either *TaSPA/HvBLZ2* were expressed higher than *TaBLZ1* in the full length, while both have low expression level in the mature grain in the hexaploid species. However, *TaSPA/ HvBLZ2* showed lower expression in the mid-length comparing with *TaBLZ1*. Consequently, the expression of *TaSPA/HvBLZ2* confirmed that they could positively regulate storage protein genes in the endosperm, correlating with the increase of the accumulation rate of storage protein genes in the same stages. In addition, these specific expression patterns for *TaSPA/ HvBLZ2* and *TaBLZ1* either in a specific stage or tissue confirmed that the expression of our target is not just related with the accumulation of protein in the endosperm during the filling stages but it is also associated with the structure of grain which will bw address by mRNA in situ hybridization in the next part.

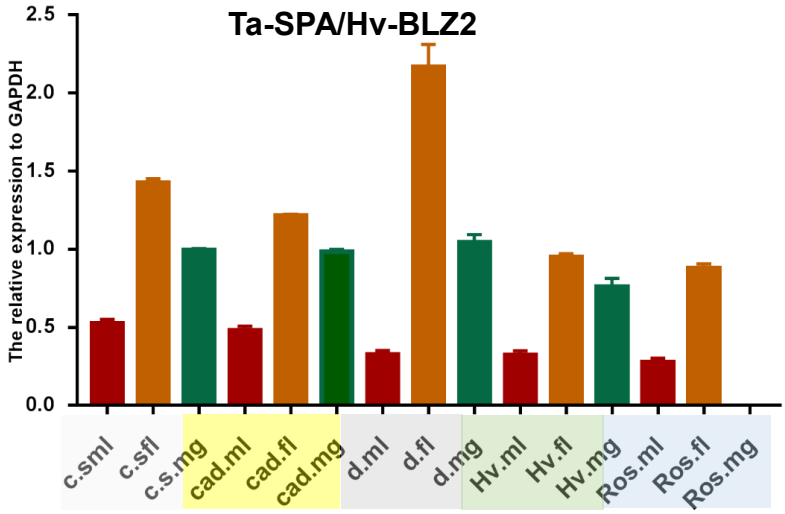


Figure 3.6 Relative expression levels of Ta-SPA /Hv-BLZ2 genes.

Both are expressed during the grain of development of grains in different species (c.s, Chinese spring; cad, cadenza; D, Durum; Hv, barley; and Rosner). (ML, mid length; FL, full length, Mg, mature grain). Two way ANOVA was run to determine the effect of time and varieties of genotypes on gene expression. The main effect of the time was significant on gene expression, F (2, 20) = 2383, P<0.0001. Moreover, the effect of varieties of genotype on gene expression was always significant, F (4,10)= 332, P<0.0001. The interaction between the time and varieties of genotypes was statistically significant, F(8,20)=150.7, P<0.0001. All pairwise comparisons were conducted to determine the difference between the combination of the time and genotypes on the gene expression.

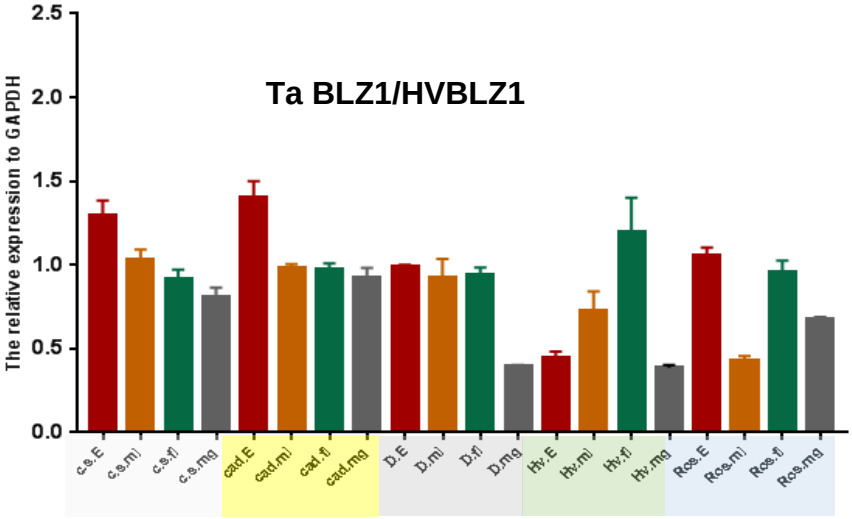


Figure 3.7 Relative expression levels of TaBLZ1/HvBLZ1 genes.

Both are expressed during the grain of development of grains in different species. (ML, mid length; FL, full length; MG, mature grain). Two way ANOVA was run to determine the effect of time and varieties of genotype on gene expression. The main effect of the time was significant on gene expression, F (3, 40) = 89.69, P<0.0001. Moreover, the effect of varieties of genotype on gene expression was always significant, F (4, 40) = 55.95, P<0.0001. The interaction between the time and varieties of genotype was statistically significant, F (12, 40) = 29.30, P<0.0001. All pairwise comparison was conducted to determine the difference between the combination of the time and species on the gene expression.

3.3.2.2 *Expression levels of three homologous of TaSPA during the filling stage.*

In the wheat genome, there are three homologous of TaSPA were identified in the genome of the hexaploid wheat (A, B, and D), while two homologous in the tetraploid wheat (A and B) by using Ensembl Plant (http://plants.ensembl.org/Triticum_aestivum/Info/Index) (Kersey *et al.*, 2017).

From the aligned sequences of TaSPA homologous, it can be noticed that there was an average of SNPs between three homologous which was high particularly in TaSPA (A genome), **(Fig.3.4)**. The similarity between TaSPA (A genome), and TaSPA (B and D) is around 87, 03%. To detect the relation between three homologous of TaSPA, the phylogenetic tree was constructed of aligned sequences of proteins by using Clustal W. The phylogenetic analysis shows TaSPA (B genome) and TaSPA (D genome) are very close compare with TaSPA (A genome). Specific pairs of primers were used for each copy of homologous to recognize the specific region for each TaSPA homologous. The qRT-PCR results show that there is variation in the expression level between these homologous **(Fig.3.5).** The reason for that it can be from the polymorphism between three homologous resulted in the polyploidy event. TaSPA (B genome) is strongly expressed than TaSPA (A genome) and TaSPA (D genome) in hexaploid and tetraploid species during the full length but it expresses higher in the tetraploid. Although TaSPA (B genome) and TaSPA (D genome) share similar sequences, TaSPA (D genome) was expressed lower than others. Generally, there are variations in the expression level in each copy in different species of wheat. The expression of TaSPA (B genome) is the highest in the tetraploid (Durum) comparing with hexaploid, while TaSPA (A genome) has the similar level of expression in tetraploid and hexaploid. Recently, it has been confirmed that the variation in the copies of genes and its expression results from early polyploidization induced regulatory changes or from the parent donor of the B genome (Ravel *et al.,* 2009).

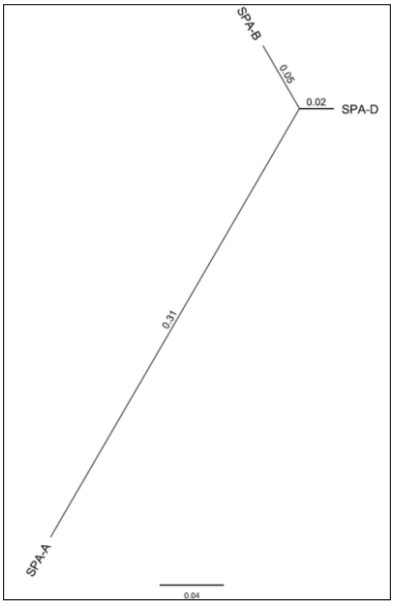


Figure 3.8 The phylogenetic tree illustrates the relation between three homologous of TaSPA gene which are (TaSPA-A, TaSPA-B, and TaSPA-D) in Chinese spring.

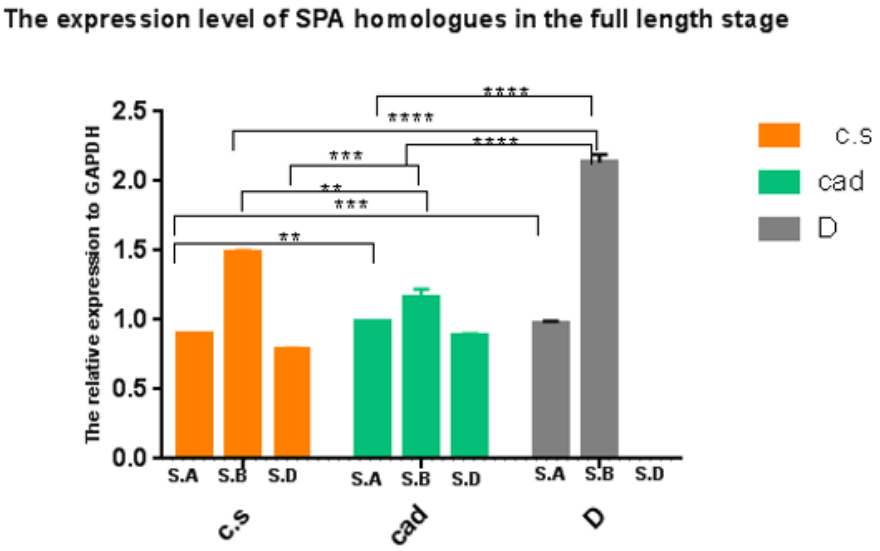


Figure 3.9 The expression levels of three homologous of TaSPA.

It was detected the expression of three homologous of TaSPA (TaSPA.A) A genome, (TaSPA.B) B genome, and (TaSPA.D) D genome in two cultivated hexaploid wheat (c.s, Chinese spring and cad, cadenza), and tetraploid wheat (D, Durum) in the full length stage (20DAA).

Two way ANOVA was run to determine the effect of time and varieties of genotypes on gene expression. The main effect of the full length stage (fl) was significant on gene expression, $F(2, 4) = 19643$, $P < 0.0001$. Moreover, the effect of varieties of genotype on gene expression was always significant, $F(2, 4) = 3.576$, $P= 0.1287$. The interaction between the full length stage (fl) and varieties of genotype was statistically significant, $F(4, 8) = 613.1$, $P<0.0001$. All pairwise comparison was conducted to determine the difference between the combination of the time and genotype on the gene expression; the significant results of the pairwise comparison are shown as the asterisk.

3.3.2.3 *(RSR1, Q); euAP2 clade*

Although *euAP2* genes were expressed in all examined tissues such as their expressions were detected in different vegetative tissues (root, seedling and flag leaf) and reproductive tissues (ovary, anther, palea and lemma), they have specific expression patterns during the life cycle of the plant. The expression of *RSR1* and *Q* were explored in different grain development stages and the different development stages of meristems spikelets, respectively (**Appendix 7. Fig.7.3.1).** In addition, This analysis of the expression was carried out in order to understand the euAP2 expression patterns which will in turn increase understanding the function of this family.

With regards to RSR1, it has specific expression patterns during the grain development, which was high at the beginning of the filling stage and decreased slightly in the full length, particularly in the hexaploid species. In contrast, the expression of RSR1 in the tetraploid and barley decreased gradually during the grain development stages **(Fig. 3.6)**.

Analysis of expression patterns by quantitative real-time reverse transcription (qRT-PCR) showed that the expression profile of RSR1 was oppositely corollated to the accumulation rate of starch in the endosperm which increases gradually during the early and mid-stages of grain development and decreases slightly at the maturation stage (Kang *et al.,* 2012; Fu *et al.,*2010).

In terms of *Q* expresses, the primordia of spikelets were chosen in different stages which were around (0.2-0.5) cm in the length, 2 cm the differentiation stage and (6-7) cm when the most of the floral organs are complete. The transcription level peaked in the early stages of spike development and gradually decreased with spike maturation particularly when the spikelet meristem transitioned to floral meristem **(Fig. 3.7)**. Similar results in all different species of wheat and barley were obtained, similarly to those reported by Simons *et al.,*(2006) in wheat. Both paralogous (*RSR1, Q*) showed expression in different tissues, suggesting that euAP2 may play a broader role during the development of the plant, **(Figs. 3.8;3.9)**. The results showed that both expressed highly in the vegetative stage in different tissues such as root and seedling and in all reproductive tissues. Additionally, both were expressed

predominantly in the ovary and anther before the anthesis which raises a question concerning whether these TFs play roles in controlling the grain quality which this process of grain development starts in the early stage of spikelet meristem and during the development of the floral organs before the anthesis. So that, they could play an important role in improving the quality of grain, although their functions are poorly understood. However, it is possible that RSR1 have functional conservation during the flowering period which has been confirmed by the floral organs of *rsr1* mutant or RSR1 overexpression lines (Fu *et al.,*2010). Furthermore, the high expression level of *Q, RSR1* at the early stage of development of the spikelet meristem and the early stage of the development of the grain, rspectively indicate that both genes could play important roles in the development of plant and are likely to be involved in the transition between the various stages of development.

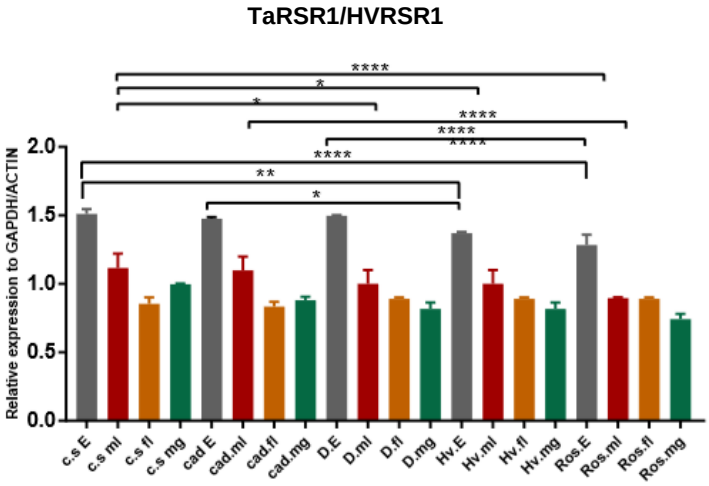


Figure 3.10 Relative expression levels of TaRSR1/HvRSR1 genes.

The expression was detected during the grain of development of grains in different species (c.s, Chinese spring; cad, Cadenza; D, durum; Hv, barley; and Rosner). (ML, mid length; FL, full length, MG, mature grain). Two way ANOVA was run to determine the effect of time and of varieties of genotype on gene expression. The main effect of the time was significant in gene expression, F (2,20)=4.539, P < 0.0001.Moreover, the effect of varieties of genotype on gene expression was always significant, F(4,10)= 6.03, P = 0.0098. The interaction between the time and varieties of species was statistically significant, F(8,20)= 4.539, P = 0.0029. All pairwise comparison was conducted to determine the difference between the combination of the time and genotype on the gene expression.

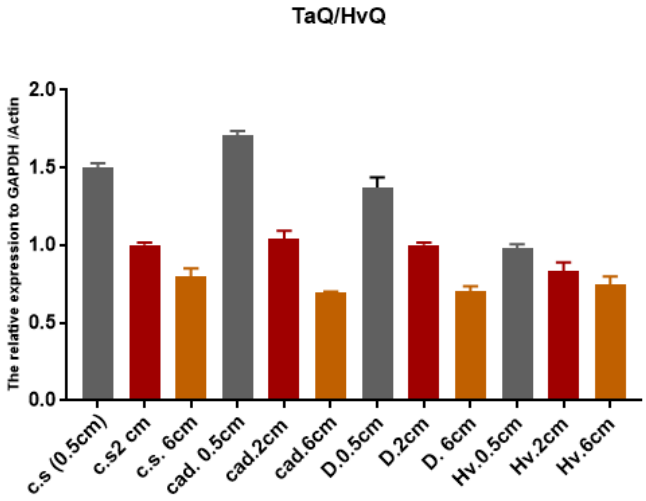


Figure 3.11 Relative expression levels of Ta-Q/ Hv-Q gene.

It was expressed in three different of the development of meristem (0.5 cm; 2cm; and 6 cm) in different species (C.S, Chinese spring; cad, cadenza, D, Durum; Hv, Barley).Two way ANOVA was run to determine the effect of time and varieties of species on gene expression. The main effect of the time was significant on gene expression, F (2, 16) = 1602, P < 0.0001.Moreover, the effect of varieties of genotype on gene expression was always significant, F(3,8) = 69.11, P<0.0001. The interaction between the time and varieties of species was statistically significant, F (6, 16) = 114, P<0.0001. All pairwise comparison was conducted to determine the difference between the combination of the time and species on the gene expression.

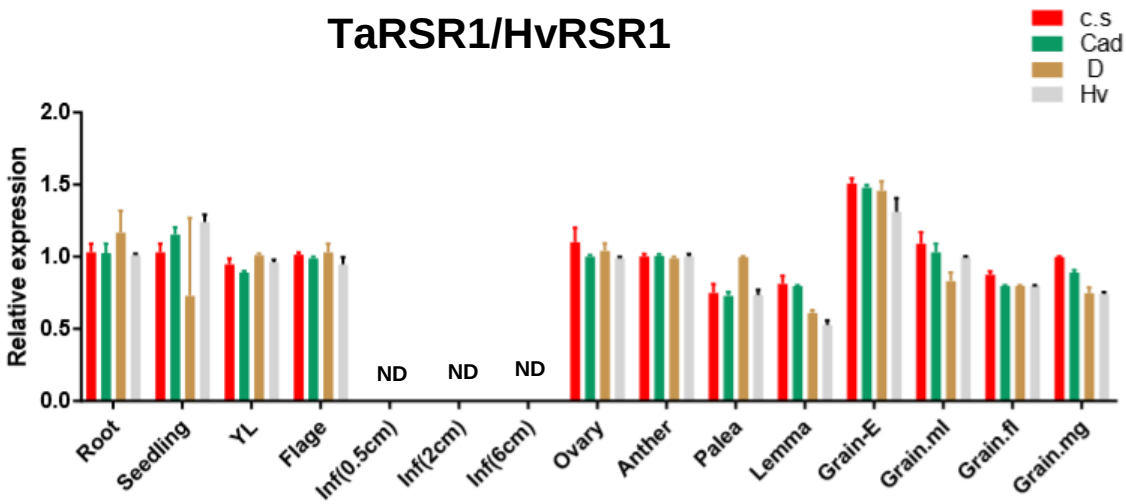


Figure 3.12 Relative expression levels of TaRSR1/HvRSR1.

They are generally expressed across vegetative tissues, floral organs and grain development stages (ML, mid length; FL, Full length; and MG, mature grain) in different species (c.s, Chinese spring; cad, Cadenza; D, Durum; and Hv, barley). Two-way ANOVA was run to determine the effect of time and varieties of genotypes on gene expression. The main effect of the time was significant in gene expression, $F(3, 6) = 236.1$, $P<0.0001$. Moreover, the effect of the verities of genotype on gene expression was always significant, $F(4, 8) = 9.362$, $P<0.0041$. The interaction between the time and varieties of genotype was statistically significant, $F(12, 24) = 6.067$, $P<0.0001$. All pairwise comparison was conducted to determine the difference between the combination of the time and genotype on the gene expression

.

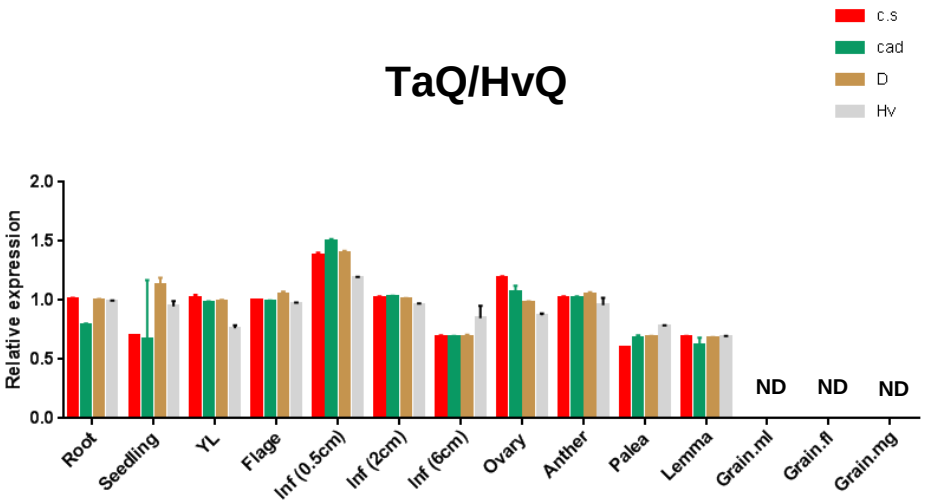


Figure 3.13 Relative expression levels of Ta-Q/Hv-Q.

They are generally expressed across vegetative tissues, floral organs and spikelet meristem in different species (c.s, Chinese spring; cad, Cadenza; D, Durum; and Hv, barley). Two way ANOVA was run to determine the effect of time and verities of species on gene expression. The main effect of the time was significant on gene expression, $F(13, 104) = 442.3$, $P<0.0001$. Moreover, the effect of verities of genotype on gene expression was always significant, $F(3, 8) = 4.91$, $P<0.0001$. The interaction between the time and verities of species was statistically significant, $F(39, 104) = 5.353$, $P=0.0320$. All pairwise comparison was conducted to determine the difference between the combination of the time and species on the gene expression.

To conclude, comparing between (TaSPA /TaBLZ1), (HvBLZ2/ HvBLZ1) and (TaRSR1/HvRSR1) during the development of grain showed that the expression of TaSPA/HvBLZ2, at the beginning of the filling stage (5-13D.A.A) was lowest comparing to Ta/HvBLZ1 and RSR1. During this stage the accumulation rates of starch and storage protein are low. Also, TaSPA/HvBLZ2 showed high expression in the filling stage (13-20 DAA-full length) which correlates to the rate of accumulation of starch and storage protein genes, while the expression of TaBLZ1 and TaRSR1 were low during the filling stage. Nevertheless, there was a significant variation in TaSPA expression between the species even between the hexaploid species such as (*Chinese spring*, Cadenza and Rosner), which was the lowest in Rosner. However, the highest expression of TaSPA was in the Durum. To clarify that, the storage protein is the higher in the durum comparing to hexploid cultivar species. However, RSR1 had low expression during the filling stage in all species which suggesting previously mentioned by Fu *et al.,*(2010), that RSR1 negatively regulates starch synthesis genes in this stage. Generally, at the maturation stage, the expression level for all our targets of TFs are low compared to that at mid-length and full length.

3.3.3 mRNA in situ hybridization

3.3.3.1 *(O2 members); bZIP family*

Messenger RNA *in situ* hybridization (ISH) is a powerful technique that can be used for detecting the temporal and spatial expression patterns of genes in different tissues from a great variety of organisms.

RNA probes are derived from sequences within one of the bZIP cDNA PCR products by using specific primers. bZIP members have a putative DNA-binding region (bZIP domain), which is present in all members of the bZIP family. Therefore, to avoid the possibility of cross-hybridization with other RNAs have been selected in the *in situ* hybridization experiment, 3' untranslated region of target genes. The Control used in this experiment were Histone three and four (H3, H4) (**Appendix 7. Table 7.3.2).**

The comparison results between different species showed that (*TaSPA* RNA) is localized in the center of the cross-section of grains which is the starchy endosperm that contains starch and protein. These signals were detected in the mid-length, full-length and mature grain in all different varieties wheat. Furthermore, *HvBLZ2* has a similar temporal and spatial expression in the grain of barley. However, it was confirmed that no signal in an early stage and in the mature grain of rye **(Figs.3.10; 3.11)**. In more details, comparing the spatial expression of *TaSPA/HvBLZ2* in two stages which are the full-length stage and mature stage in specific layers of endosperm tissues, it can be seen that the *TaSPA/HvBLZ2* were expressed in the peripheral aleurone layer and modified aleurone layer during the full-length or the maturation **(Fig.3.12).** Furthermore, although the protein and starch accumulate in the early and mid-length in the pericarp which appears clearly, also in the endosperm, the signals of *TaSPA/HvBLZ2* were restricted in the endosperm only. The mRNA hybridization confirmed all the previous results of RT-PCR for these genes.

In addition, the expressions of *TaBLZ1/HvBLZ1* were detected in four main stages of grain development in different species. TaBLZ1 showed heavy signal during the filling stages (full length and mature grain) in all examined species. Also, they were detected at the early stage and mid-length. **(Figs.3.13;3.14)**. However, the temporal and spatial expression of TaBLZ1/HvBLZ1 genes which

were not correlated with the accumulation phase of starch and protein in the endosperm but it could be associated with the changing in the structure of the grain during the development. In addition, the mRNA-ISH results of the spatial expression *TaBLZ1* in full-length and mature grain revealed that the *TaBLZ1* in different species were expressed in all endosperm tissues, which are: the central starchy endosperm, the peripheral aleurone layer, and modified aleurone layer in both stages, although there is variation in the structure of aleurone layers between the examined grains **(Fig.3.15).**

The grain development of temperate cereal has been studied in details particularly in barley, which shows similar developmental patterns between examined species, but with the variance in the developmental time between them (Olsen *et al.,* 1992). However, the endosperm development has four main stages; early syncytial stage, the cellularization phase, the differentiation and grain filling stage and finally, cell death and drying down. Additionally, the accumulation of starch and storage proteins occurs in the central endosperm and peaks during mid-late development.

Therefore, The differential expression patterns of either *HvBLZ1* and *TaSPA* wheat, or *TaBLZ1* and *HvBLZ2* barley may be related to the timing of protein formation and accumulation. This means *TaBLZ1* could involve in protein body formation at the early stage, while *TaSPA* and *TaSPA/HvBLZ2* expression starts when the cellulation and the differentiation are complete, and the accumulation of storage protein and starch occur. During this stage, both TaSPA and HvBLZ2 could control the accumulation of protein (Haseneyer *et al.,* 2010).

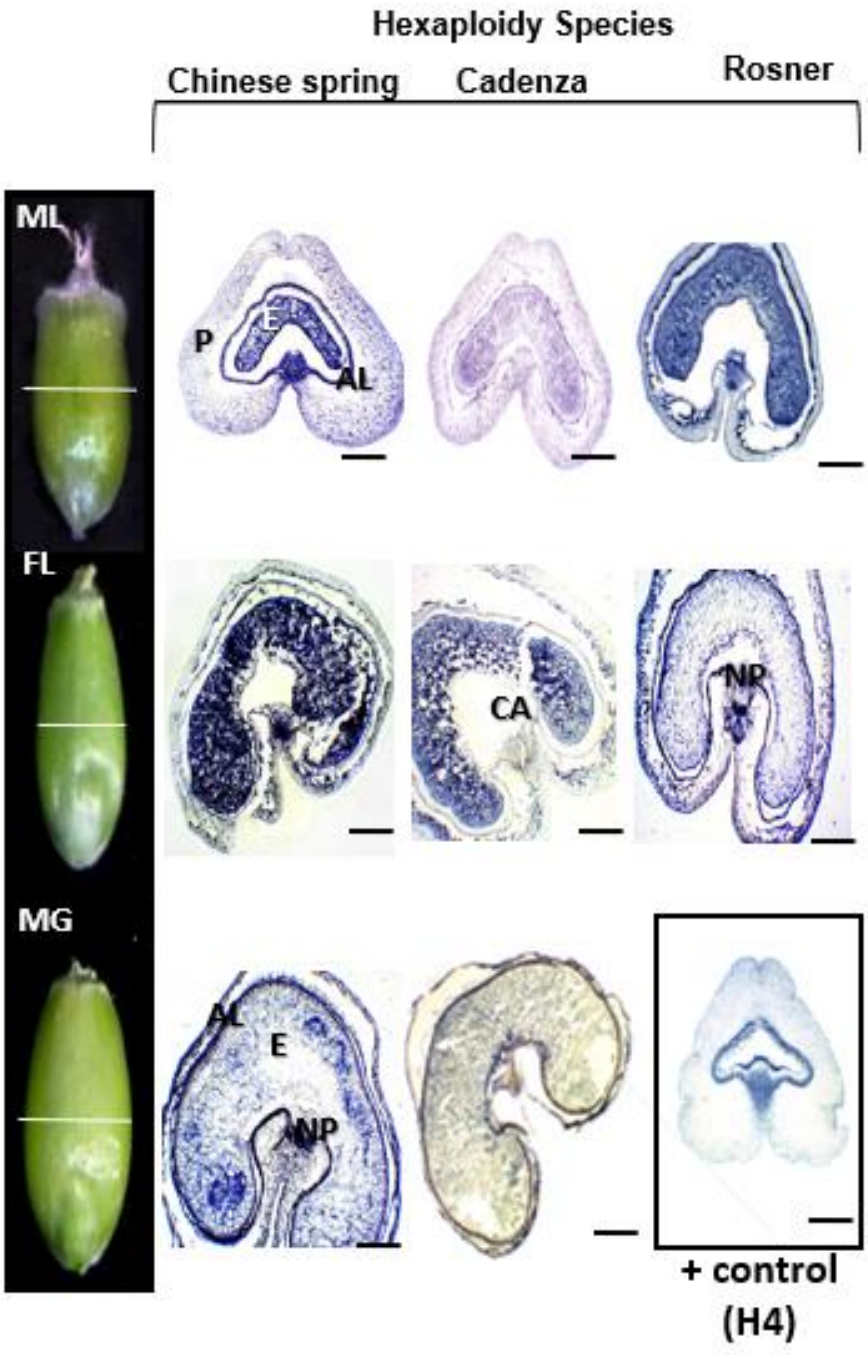


Figure 3.14 Specific expression patterns of endosperm gene (TaSPA).

As determined by in situ hybridization in the Hexaploid varieties of wheat during the grain development stages (ML, mid-length; FL, full-length; MG, mature grain).

ML) Transverse sections of 15 DAA (mid-length). **FL)** Transverse sections of 20 DAA (Full-length)**. MG)** Transverse sections of 25 DAA (Mature grain). **P)** Pericarp**, AL)** aleurone layer, **E)** Central endosperm, **NP)** nucellar projection. Histone 4 was used as Positive control. Scale bars: 200 µm.

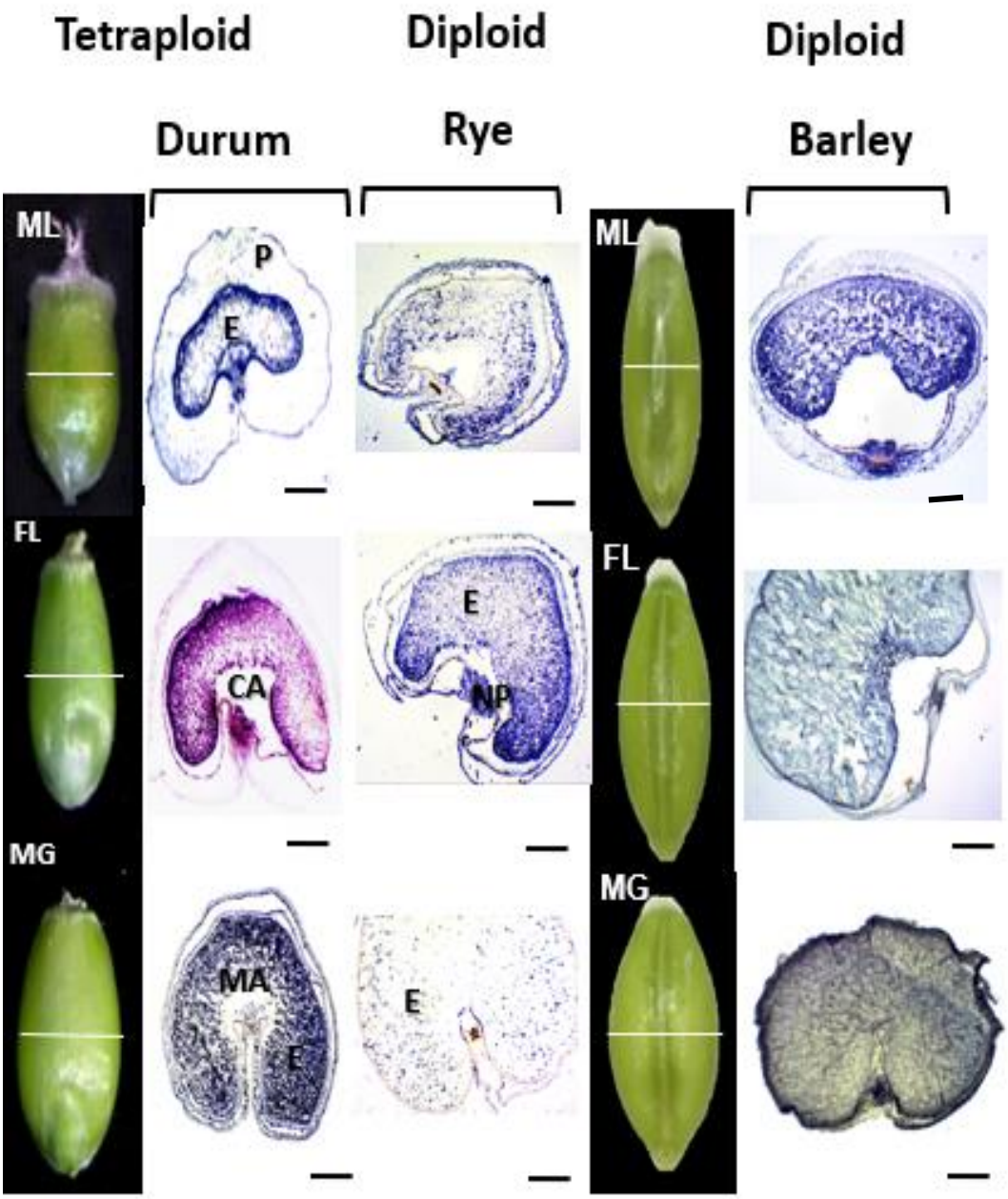


Figure 3.15 Specific expression patterns of endosperm genes (TaSPA/HvBLZ2).

As determined by in situ hybridization in the tetraploid wheat (Durum) and diploid species barley and rye during the grain development stages (ML, mid-length; FL, full-length; MG, mature grain).

ML) Transverse sections of 15 DAA (Mid-length). **ML)** Transverse sections of 20 DAA. **MG)** Transverse sections of 25 DAA (Mature grain). **P)** Pericarp, **AL)** Aleurone layer, **E)** Central endosperm, **NP)** Nucellar projection, **MA)** Modified aleurone layer. Scale bars: 200 µm.

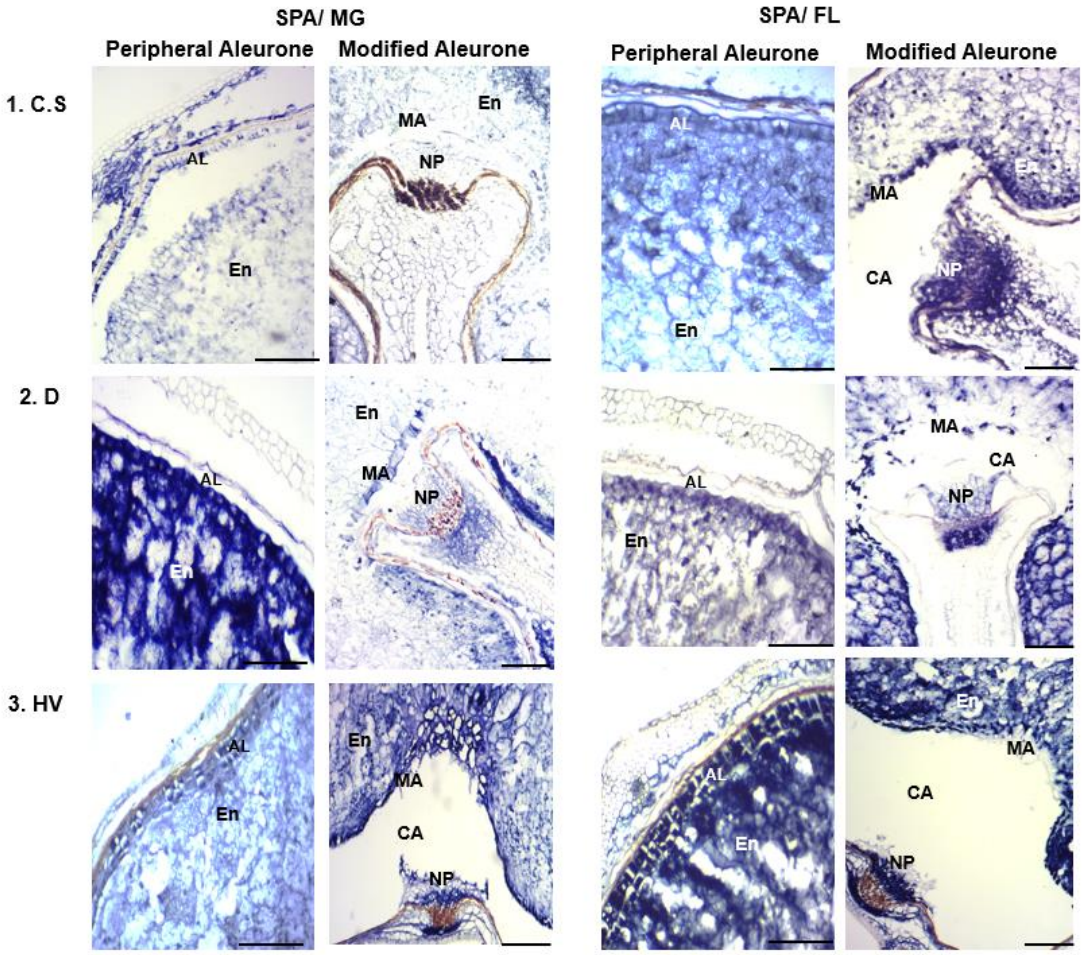


Figure 3.16 Transverse sections illustrate specific expression patterns of (TaSPA and HvBLZ2) in wheat varieties and barley comparing between the full-length and mature grain stage.

(1) Expression patterns of TaSPA in Chinese spring (C.S) in peripheral aleurone layer and modified aleurone mature grain (MG) and Full-length (FL) stages.
(2) Expression patterns of TaSPA in Durum (D) in mature grain (MG) and Full-length (FL).
(3) Expression patterns of TaBLZ2 in Barley (Hv) in mature grain (MG) and Full-length (FL) 200 µm. AL, alureone layer; En, endosperm; MA, modified alurone layer; NP, nuclear projection; CA, cavity.; Scale bars: 100 µm.

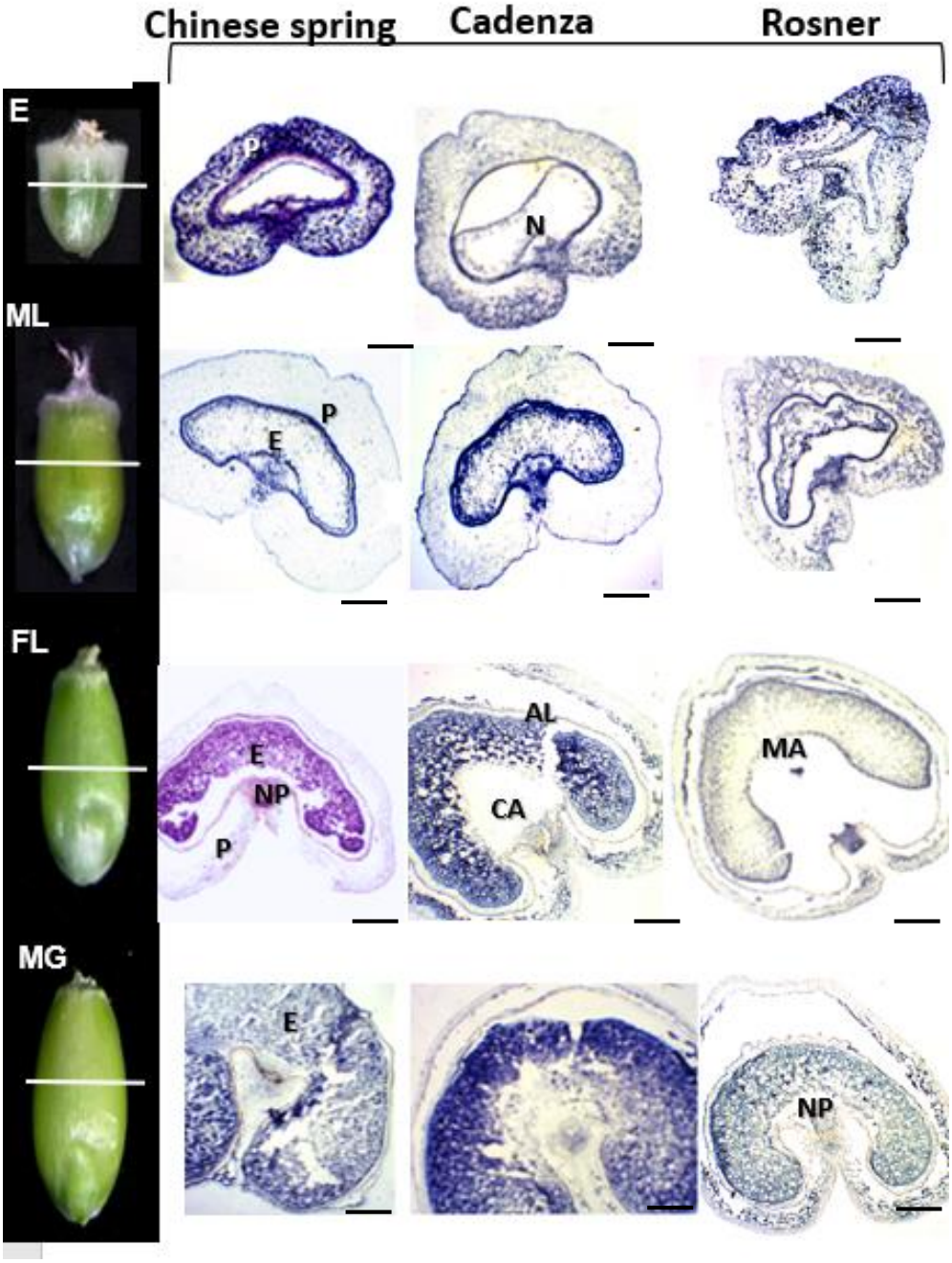


Figure 3.17 Specific expression patterns of (TaBLZ1).

As determined by in situ hybridization in the hexaploid varieties of wheat during the grain development stages (E, early stage; ML, mid-length; FL, full-length; MG, mature grain).

E) Transverse sections of 3-5 DAA (Early stage). **ML)** Transverse sections of 15 DAA (Mid-length stage. **FL)** Transverse sections of 20 DAA (Full-length). **MG)** Transverse sections of 25 DAA (Mature grain). **P)** Pericarp, **AL)** Aleurone layer**, NP)** Nucellar projection, **E)** Central endosperm **MA)** Modified aleurone layer. Scale bars; 200 µm

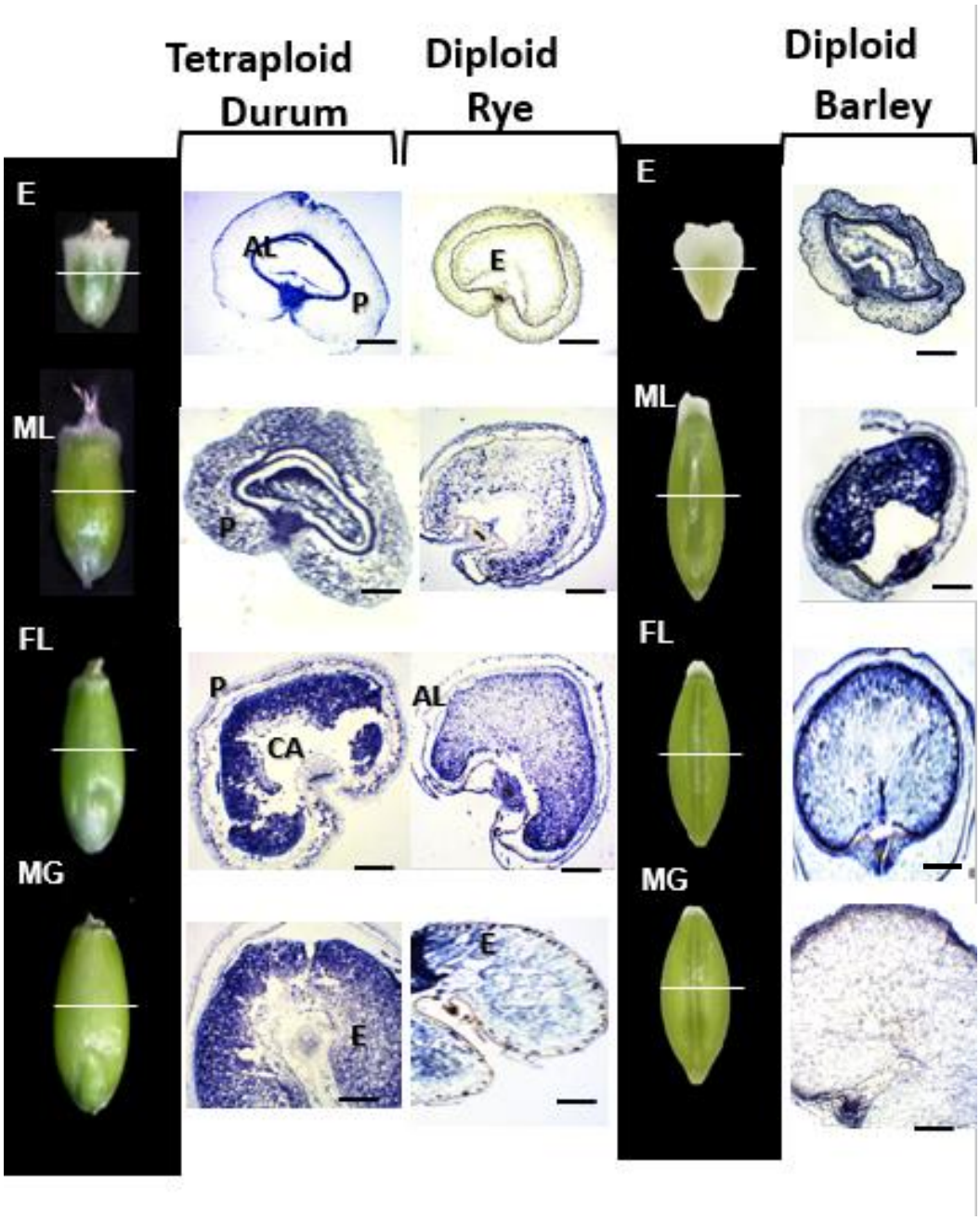


Figure 3.18 Specific expression patterns of (TaBLZ1/HvBLZ1).

As determined by in situ hybridization in the tetraploid wheat (Durum), and diploid species such as barley, and rye during the grain development stages (E, early; ML, mid- length; FL, full-length; MG, mature grain).

E) Transverse sections of 3-5 DAA (Early stage). **ML)** Transverse sections of 15 DAA (Mid-length stage). **FL)** Transverse sections of 20 DAA (Full-length) stage. **MG)** Transverse sections of 25 DAA (Mature grain stage). **P)** Pericarp, **AL)** Aleurone layer, **NP)** Nucellar projection, **E)** Endosperm, **CA)** cavity. Scale bars: 200 µm.

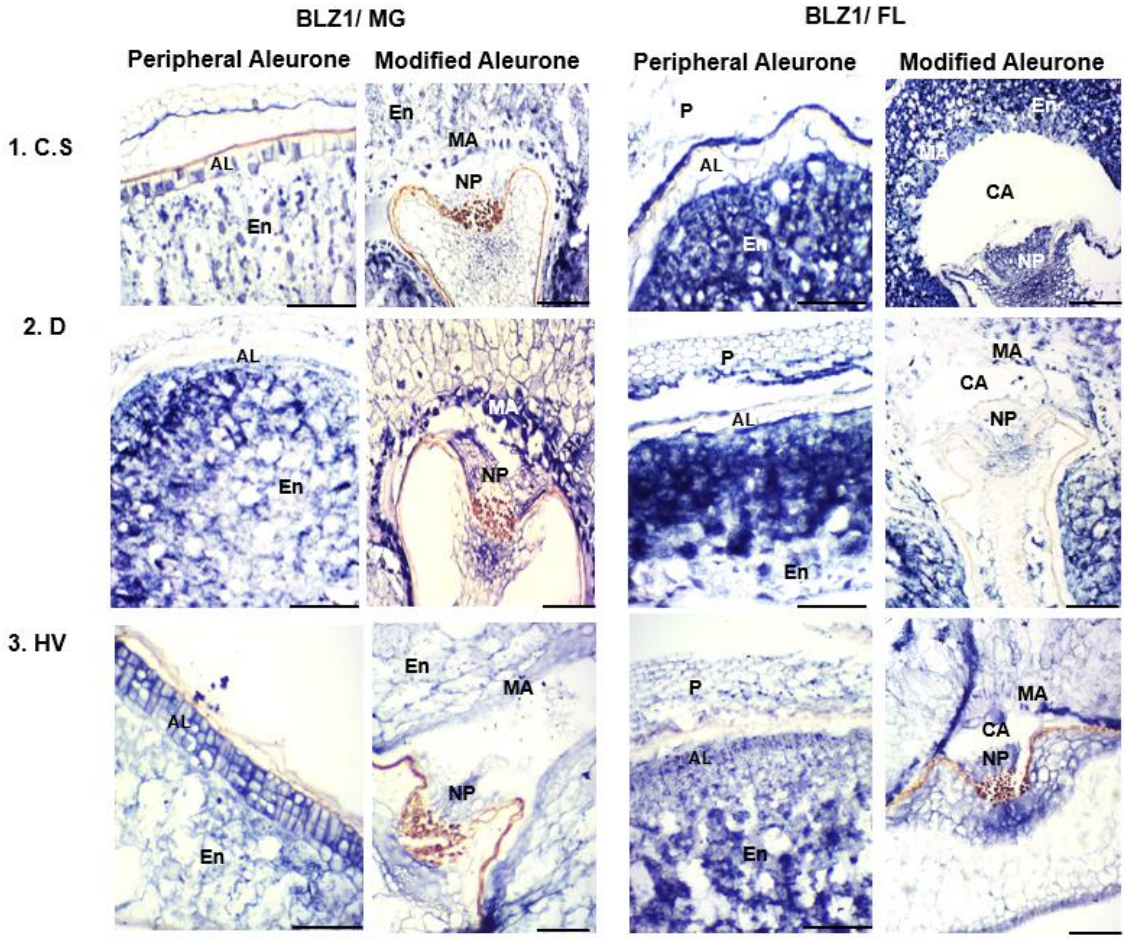


Figure 3.19 Transverse sections illustrate specific expression patterns of (TaBLZ1 and HvBLZ1) in wheat varieties and barley comparing between the full-length and mature grain stage.

(1) Expression patterns of TaBLZ1 in Chinese spring (C.S) in peripheral aleurone layer and modified aleurone in mature grain (MG) and Full-length (FL) stages.
(2) Expression patterns of *TaSPA* in Durum (D) in mature grain (MG) and Full-length (FL).
(3) Expression patterns of *TaBLZ2* in Barley (Hv) in mature grain (MG) and Full-length (FL) 200 µm. **AL**, alureone layer; **En**, endosperm; **MA**, modified alurone layer; **NP**, nuclear projection; **CA**, cavity. Scale bars: 100 µm.

3.3.3.2 *(RSR1, Q); euAP2 clade*

As it is known the RSR1 has low expression during the grain development that make the detection the spatial expression patterns it was not easy. So that, in this section, only *Q* expression patterns were examined by using mRNA-ISH, including different development stages of different species in primordia of spikelet tissue, spikelet meristem tissues and the floral organs after the emergence as an ovary, lodicules and anthers. Additionally, Inflorescence development in varieties of wheat and barley was comparatively studied with scanning electron and ISH. It was performed in the spikelet initiation stage (0.5 cm), the floral differentiation stage (2-3 cm), and the young spike with immature florets (6-7 cm) **(Figs. 3.16; 3. 20**). The inflorescence meristem differentiates to florets meristem. It forms a sterile lemma and rudimentary glume primordial, three stamen primordia were observed in the spikelet on the lemma side and on the top of the spikelet meristem **(SEM photos).**

In terms of the *Q* was expressed highly in the primordia of meristem of spikelet especially at the apical meristem itself. In addition, the first heavy labeling occurs in the newly arising floret primordia. *Q* expression was gradually localized to regions that will give rise to floral organs primordia. For example, the meristem spikelet shows a little evidence of the formation of ridges which can be seen in different levels. Each level indicates one of the floral organs. The first structure of floral can be seen is lemma, after that the palea. Later, there are three little papillae can be recognized on the top of meristem of spikelet which are the primordia of the stamens and in the middle which is the pistil. Furthermore, the ovary differentiates to follow with the style and stigma at the end. Therefore, *Q* expression appeared signal in all the initial floral organs such as stamen, ovule, palea and lemma, *Q* expression also presented in the floral organs after the emergence. In addition, the expression of *Q* appeared in the procambium cells in the rachis in the early stage of the development of spikelet, which ultimately gives rise to the vasculature. Comparing between all the different species of wheat and barley, the result showed that *Q* expression patterns are similar in different species either hexaploid species, tetraploid or diploid species. To conclude that, most these sections have strong signals which appeared everywhere in the primordia of

spikelet but during the differentiation of spikelet the position of the *Q* expression localized in the specific position of primordia of floral. Furthermore, the expression of *Q* still continues after emergence to be specific in the floral organs such as it has heavy labeling in the anther, ovule and lodicules of floret after the emergence. Consequently, from this results *Q* is likely to involve in the development of the floral meristem, the identity of the floral organs and controlling the flowering time, while, the role of *RSR1* in the flowering stage is unknown, whether it could involve in the development of the floral organs and control of flowering time.

1

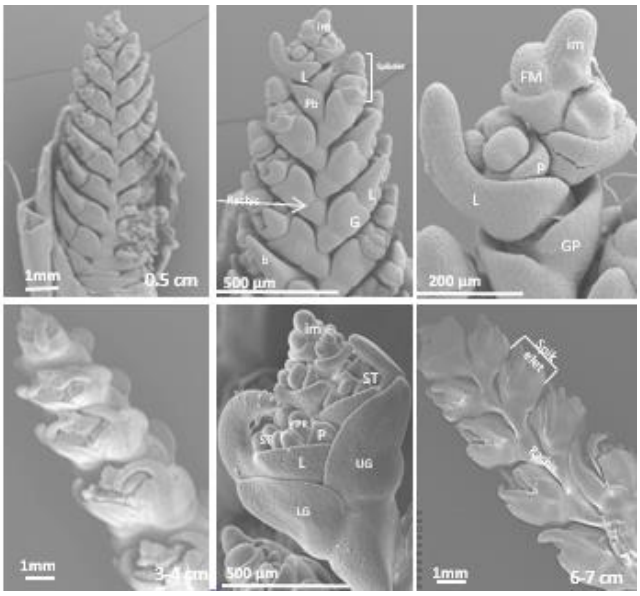
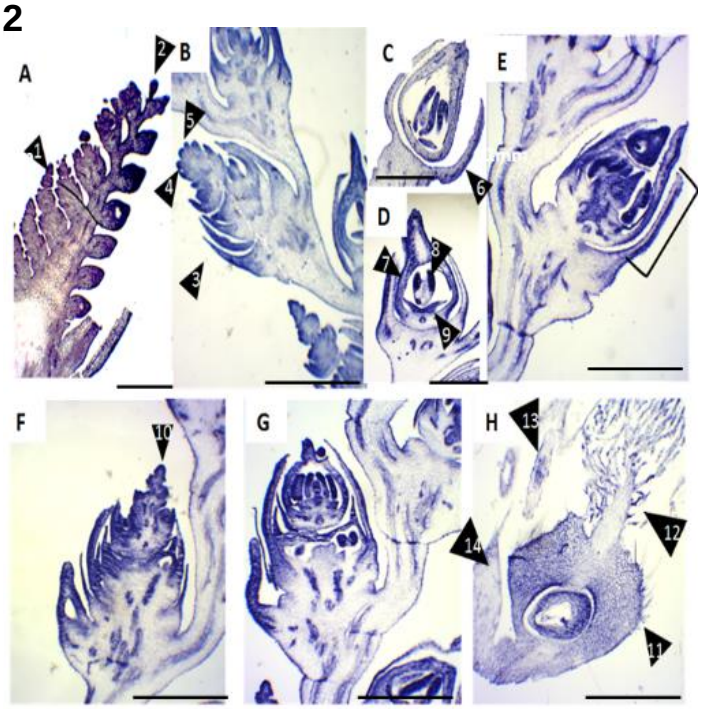


Figure 3.20.1. SEM for the spikelet meristem of *Chinese spring* and mRNA-ISH of spikelet meristem sections with a Q mRNA probe.

L, Lemma; **P,** Paleae; **G,** Glume; **GP,** Glume primordia; **LG,** lower glume; **UG**,Upper glume; **b,** bracts; **Pb**, primary branch; **im,** inflorescence meristem; **ST,** stamens.

2



2. The results of mRNA-ISH show that regionalization of label is occurring within spikelet primordia, floral organ primordia and ovary.

A) Spikelet meristem primordia (0.5 cm). **A.1** floral primordia, **A.2** apical meristem.
B) Floral primordia (2cm). **B.3** Initial of glumes. **B.4** Initial of carpel. **B.5** Initial of stamens.
C/D) Floral organs. **C.6** glumes, **D.7** Palea and lemma. **D.8** stamens, **D.9** Lodicules
E) Floral, some labelling on the procambium cells.
F) Floral primordia. **F.10** ovule primordia.
G) Floral primordia 5 cm.
H) Carpel, Ovule and embryo are labelled internally. **H.11** Ovule, **H.12** stigma, **H.13** anthers. **H.14** Lodicules. Scale Bar 200 µm, 100 µm.

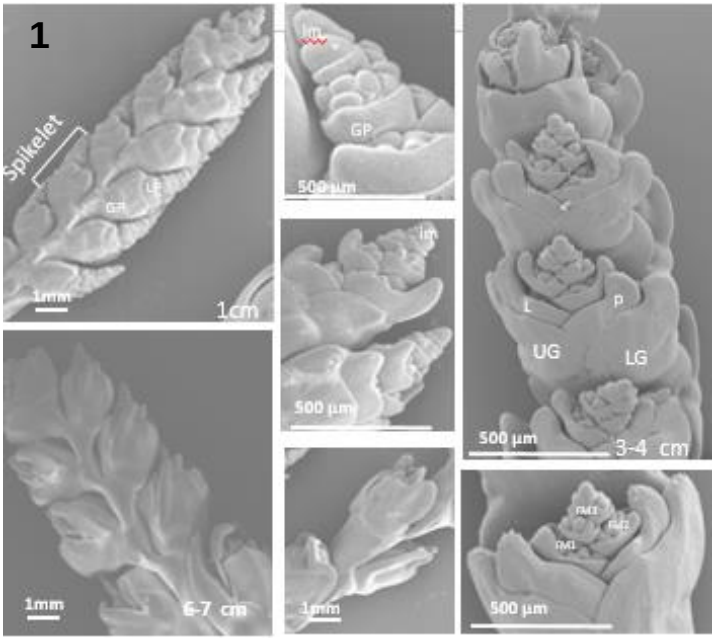
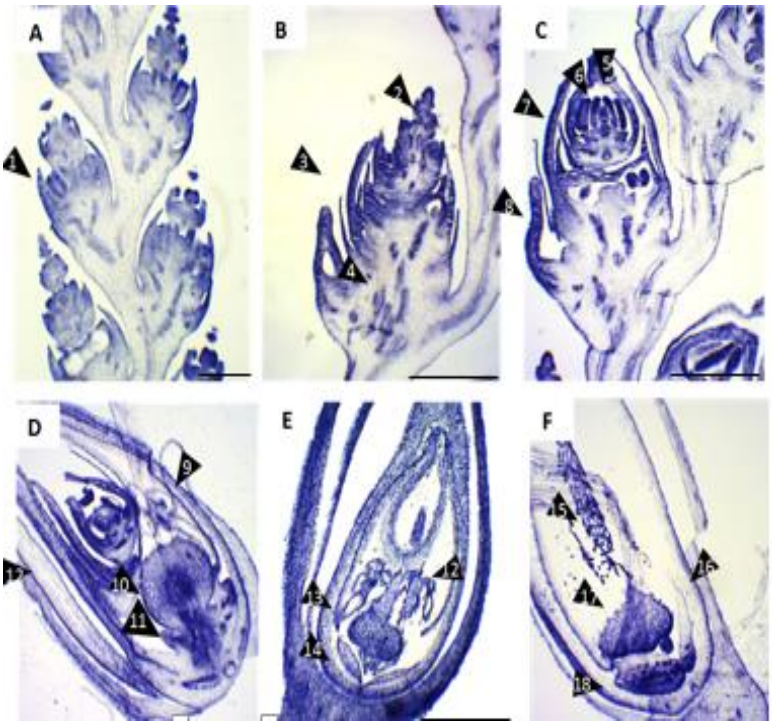


Figure 3.21.1. SEM for the spikelet meristem of Cadenza, and mRNA-ISH of spikelet meristem sections with a Q mRNA probe.

L, Lemma; **P,** Paleae; **G,** Glume; **GP,** Glume primordia; **LG,** lower glume; **UG,**Upper glume; **b,** bracts **Pb,** primary branch; **im,** inflorescence meristem; **ST,** stamens.

2



2. The results of mRNA-ISH show that regionalization of label is occurring within spikelet primordia, floral organ primordia and ovary.

A) Spikelet meristem (2 cm). **A.1** floral primordia.
B) Floral primordia (2cm). **B.2** stamens, **B.3** initial of glumes, **B.4** some labelling on the procambium cells.
C) Floral primordia 5 cm. **C. 5** ovule primordia, **C.6** stamens, **C.7** glumes, **C.8** Glumes primordial.
D) Floret before the emergence, **D.9** anther, **D.10** ovary, **D.11** lodicule, **D.12** glumes.
E) Floret after the emergence, **E.12** Stigma, **E.13** Lemma, **E.14** Lodicules.
F) Floret after the emergence, **F.15** Anther, **F.16** Palea, **F.17** Ovary, **F.18** Lodicules. Bar 200 µm, 100 µm.

1

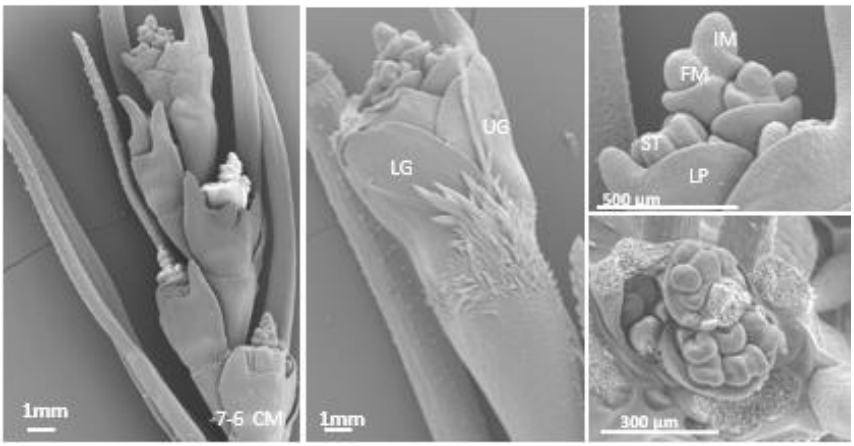
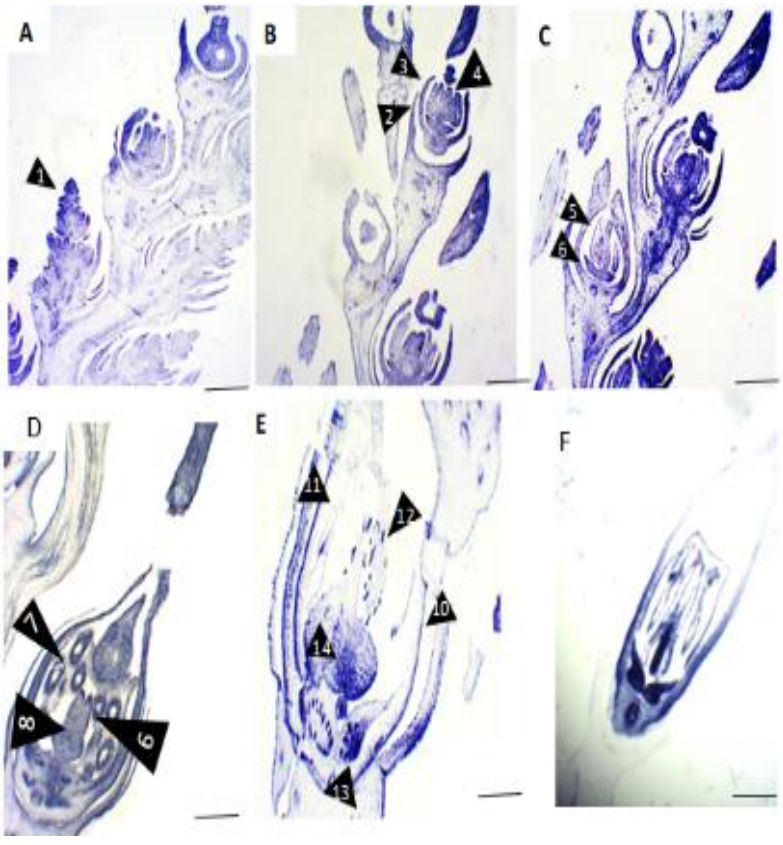


Figure 3.22.1. SEM for the spikelet meristem of Durum,

LG, lower glume; **UG**, Upper glume; **im,** inflorescence meristem; **ST,** stamens; **FM,** floret meristem.

2



2. The results of mRNA-ISH show that regionalization of label is occurring within spikelet primordia, floral organ primordia and ovary.

A) Spikelet meristem (2 cm). **A.1** Floral primordia.
B) Floral primordia (5cm). **B.2** Glumes, **B.3** Lemma, **B.4** Stamens.
C) Floral primordia (5cm). **C.5** Lemma, **C.6** Lodicule.
D) Floret before the emergence. **D.7** Anther, **D.8** Lodicules, **D.9** ovule.
E) Floret after the emergence, **E.10** Palea, **E.11.** Anther, **E.12** Stigma, **E.13** Lodicules, **E.14** Ovule. Scale Bar 200 µm, 100 µm.

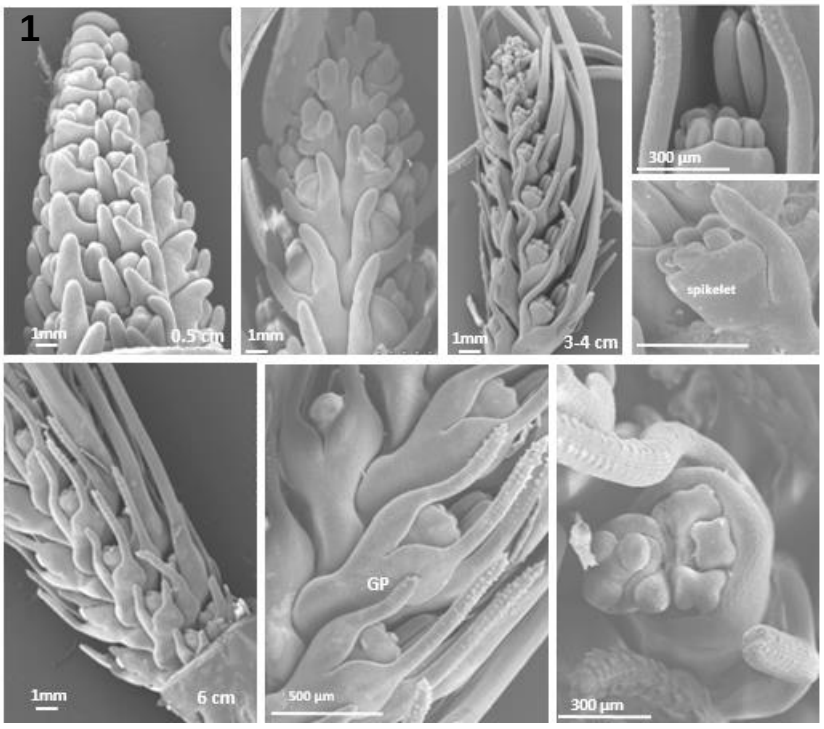
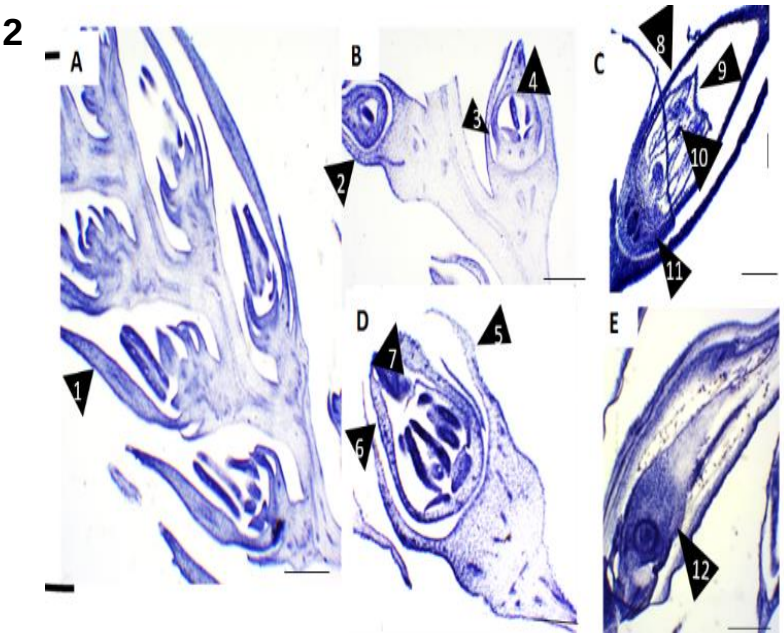


Figure 3.23.1. SEM for the spikelet meristem of Barley (Hordeum vulagare), and mRNA-ISH of spikelet meristem sections with a Q mRNA probe.

L, Lemma; P, Paleae; **G**, Glume; **GP**, Glume primordia; **LG**, lower glume; **UG**,Upper glume; **b,** bracts **Pb**, primary branch; **im**, inflorescence meristem; **ST,** stamens. and ovary. Scale Bar 200 µm, 100 µm.



2. The results of mRNA-ISH show that regionalization of label is occurring within spikelet primordia, floral organ primordia.

A) Spikelet meristem (2 cm). **A.1** Floral primordia, Bristles
B) Floral primordia (5cm). **B.2** glumes, **B.3** lemma**, B.4** stamens.
C) Floral primordia (5cm). **C.8** glumes, **C.9** Stamens, **C.10** ovule; **C.11** Lodicule.
D/E) Floret after the emergence, **D.5** palea, **D.6** stigma; **D.7** Anther; **E.12** ovule. Scale Bar 200 µm, 100 µm.

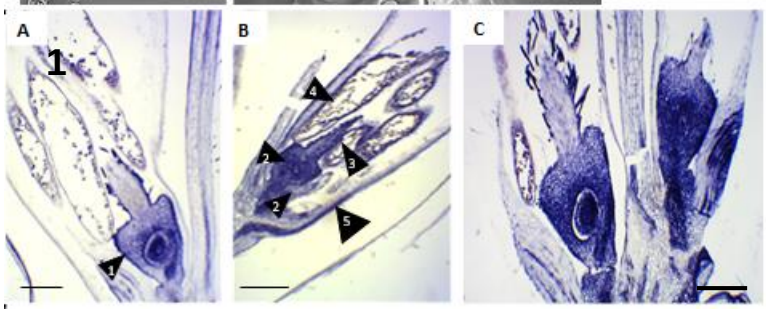
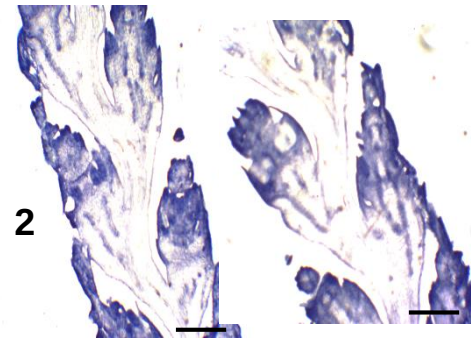


Figure 3.24. 1. SEM for the spikelet meristem of Rosner, and mRNA-ISH of spikelet meristem sections with a Q mRNA probe.

L, Lemma; P, Paleae; **G**, Glume; **LG**, lower glume; **UG**,Upper glume; **im**, inflorescence meristem;**ST,** stamens.



2. The results of mRNA-ISH show that regionalization of label is occurring within spikelet primordia, floral organ primordia and ovary.

A, B, C) Spikelet shows the floret organs after the emergence. **A.1** Ovary, **B.2** Ovule, **B.3** Stigma, **B.4** Anther, **B.5** Lemma, **B.6** Lodicules. Scale Bar 200 µm, 100 µm.

3.3.4 **Expression profile of miR172 and its role**

This study focused primarily on the miR172 regulation of *euAP2* and compared the expression levels of miR172 and its targets *euAP2* genes during the development of varieties of wheat and barley. The expression of a miR172 was detected via endpoint PCR ((**Fig.3.22**), and qPCR analysis by using a specific forward primer and a universal reverse primer for the amplification of miR172 from different tissues. The relative quantity of individual transcripts was calculated by using a mathematical model, including the efficiency, which was 85-100%, and the internal control. The result shows that miR172 was not expressed in the root (2 DAG, days after germination). Following this, a progressively steady transcript accumulation was noticed during the vegetative stage in the seedling (**Fig.3.21**). The miR172 transcripts also accumulated highly during the reproductive stage particularly in the development of meristem of spikelet, where the result confirmed that miR172 had accumulated in both the leaves and flower buds over time (Nodine and Bartel, 2010; Zhu *et al.,* 2010). The expression level of miR172 increased highly during the development of spikelet meristems, which could indicate the importance of miR172 to the transition of spikelet meristem to floral meristem and to determine the floral organ identity. However, miR172 was expressed highly during the early grain development stage and decreases gradually during the development of the grain. It is possible that the function of the miR172 pathway is controlling cell fate through the regulation of euAP2 genes. Thus, the higher expression of miR172 in the later stage of vegetative tissues and during the developing spikelet meristem indicates that miR172 could regulate the timing of floret initiation. Comparing the level of miR172 expression and (TaRSR1 and TaQ) showed inverse correlation, but this relation was conserved within the examined species. This result was confirmed in the Arabidopsis (Aukerman and Sakai, 2003), that indicated that miR172 downregulated its target during the flowering time (**Fig. 3.22 A**). Therefore, it can be concluded that, both genes targeted by miR172 normally function as floral repressors, resulting in the promotion of flowering. Consequently, a model was suggested that assumes the patterns of downregulation of *euAP2* target genes by miR172 which occur during

the development of wheat (**Fig. 3.22 B**). This result is different from the data was obtained from Arabidopsis which showed that the temporal expression of miR172 causes temporal down regulation of the *AP2*-like target genes at the level of translation during the vegetative and flowering time, which in turn triggers flowering. This means that the expression of miR172 is inversely correlated with its target *AP2-like* genes in Arabidopsis during the vegetative and reproductive stages but there is no result in the seed or fruit development stage in Arabidopsis. From this, it was concluded that miR172 controls the transition from the vegetative to reproductive stage, flowering time and floral organs identity, as well as controls the development of grain by down regulating a *euAP2* in specific positions. (**Fig.22B).**

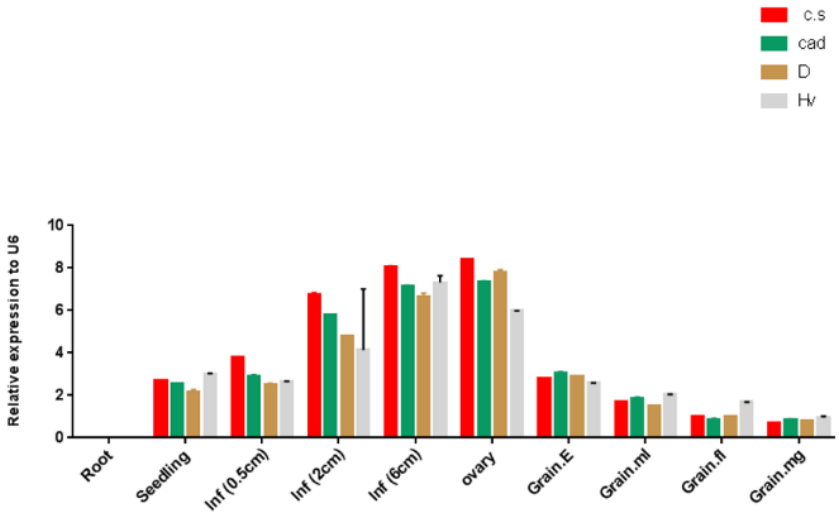


Figure 3.25 Relative expression levels of miR172 that are generally expressed across vegetative tissues, spikelet meristem and grain development stage.

Two-way ANOVA was run to determine the effect of time and varieties of genotypes on gene expression. The main effect of the time was significant on gene expression, F (9, 80) = 385.8, ****=P<0.0001. Moreover, the effect of varieties of species on gene expression was always significant, F (3, 80) = 9.793, ****= P<0.0001. The interaction between the time and varieties of species was statistically significant, F (27, 80) = 4.215, ****=P<0.0001. All pairwise comparison was conducted to determine the difference between the combination of the time and species on the gene expression.

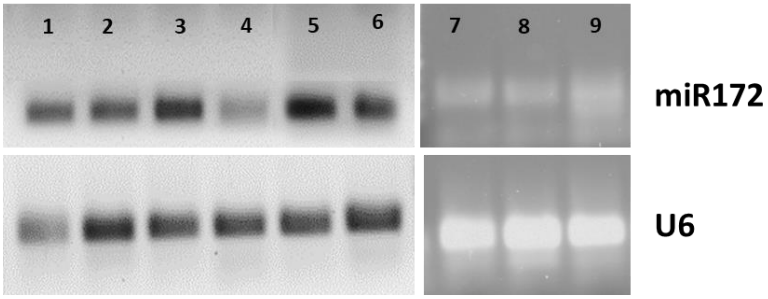


Figure 3.26 End–point PCR results for miRNA172 of some tissues, comparing between the reference U6 and the target of miR172.

1. Seedling; 2. Inflorescence (0.5 cm); 3. Inflorescence (1cm); 4. Grain (E); 5. Ovary; 6. Inflorescence (6cm); 7. Grain (ml); 8. Grain (fl); 9. Grain (mg).

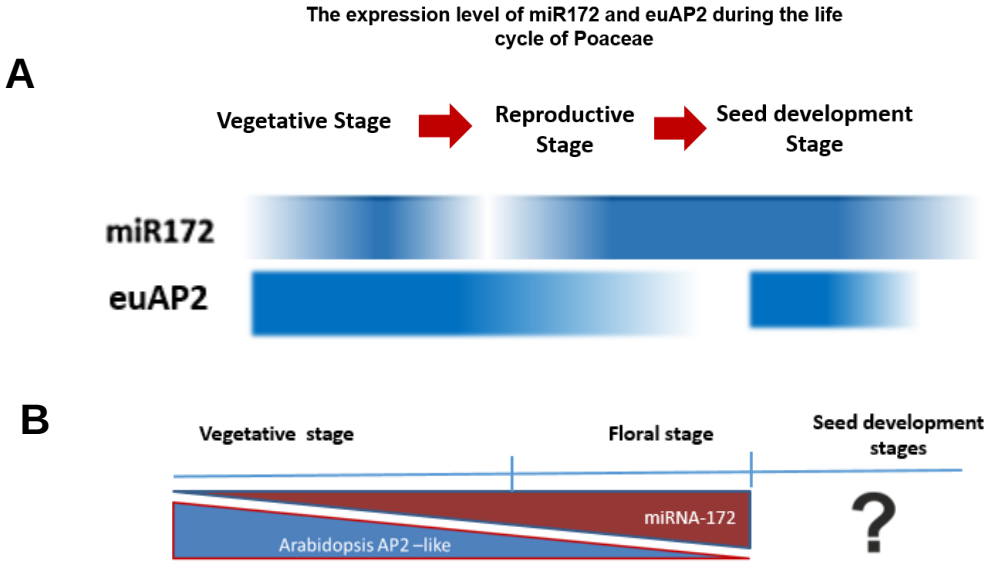


Figure 3.27 A. A predicted model in this project to represent the expressions level of miRNA172 comparing to its targets (TaRSR1, and TaQ) during different development stages (vegetative, reproductive and grain development.

B. The predicted model of miR172 and *AP2*-like in Arabidopsis during the vegetative and flowering time (Aukerman and Sakai, 2003).

The expression for both are illustrated by the shaded bars; time increases from left to right. I propose that miR172 repress the expression of *euAP2* genes that act in these pathways.

3.4 Discussion

Our target genes are a minor part of a network of many TFs, which involved in the regulation of seed storage proteins and starch synthesis genes, although they are known to be important factors influencing seed quality traits. Furthermore, wheat yield is largely determined during the period prior to flowering, when the final numbers of fertile florets and grains per spike are established. The number of fertile florets is determined when the spike meristem elongates inside the culm before the emergence (Miralles *et al.,* 2000; Ferrante *et al.,* 2012). Therefore, it can be assumed that the grain quality and yield are influenced by flowering time and inflorescence architecture, which affect both the number and the size of seeds (Kellogg *et al.,* 2013). Also, during the development of grain the yield of grain can be controlled by protein and starch which were regulated by examined TFs either directly or indirectly.

The TFs expression profiles during seed development were determined at four developmental stages and these were chosen based on earlier studies of candidate O2-like. A detailed comparative expression profile of O2-like genes has not been performed using mRNA ISH previously.The mRNA abundances of putative TFs were determined at each stage using RT-PCR, and (q RT-PCR) approches. From the expression profile of (O2 members) bZIP family, it can be divided them into two groups, the first includes *TaSPA* /*HvBLZ2* were both specifically expressed in the starchy endosperm during the filling stage. Additionally, it was observed that they were expressed in the three tissues, which are peripheral aleurone, the starchy endosperm and modified aleurone layer during the filling and maturation stages. For that, they were described as endosperm-specific TFs, which control the accumulation of storage protein. The second genes are*TaBLZ1/HvBLZ1* have a broader expression in different tissues. They were expressed constitutively throughout the plant, also, during the development stages of grain (early stage, mid-length, full-length and maturation). Both have expression patterns in different endosperm tissues such as peripheral aleurone layer, the starchy endosperm and modified aleurone layer.

This result supports earlier observations that these TFs are master regulators of storage protein genes expression during the filling stage. Albani *et al.,* 1997 reported that *TaSPA* was specifically expressed in wheat grain. The *TaSPA* expression was detected around 10 DAA and increased relatively in the following stages, peaking at 18 DAA and then decreased slightly during the maturation. On˜ate *et al.,* (1999) examined *HvBLZ2* expression by using northern blot analysis, which was restricted to the endosperm.

Diaz and Carbonero, (1998) identified *HvBLZ1* as the member of bZIP, and mRNA was detected at the early stage in the endosperm, and it was expressed in leaves and root. Therefore, it can be concluded that there was no specified for the expression of TaBLZ1during development grain or the vegetative and the reproductive tissues. In addition, HvBLZ1 was the transcriptional activator in the yeast system.These results were correlated with our result but our survey was extended to detect the expression in reproductive tissues. In addition, the spatial expression of *TaBLZ1/HvBLZ1* raises a question as to whether *TaBLZ1/HvBLZ1* involve in determining the identity of floral organs or controlling flowering time, which plays an important role in grain quality (Haseneyer *et al.,* 2010).

In terms of the paralogous, *RSR1* and *Q* have been studied as paralogous within the euAP2 clade in the Poaceae. Most of the comparative studies in *RSR1* or *Q* genes have focused on the specific expression patterns during the development of the grain or the meristem spikelet, respectively. Consequently, I compared them with a broader scope in order to increase the understanding of these genes, this may help to predict their roles during the development of the plant. Our analysis indicates that both paralogous were expressed universally, and not restricted to specific tissues or stages during plant development. The expressions of both were detected in different tissues such as vegetative tissues (roots, seedling, young leaf and flag leaf). Furthermore, they were expressed in all reproductive tissues (ovary, anther, palea and lemma). Therefore, it can be deduced these genes have a broader functions.

RSR1 is not an endosperm-specific gene. Also, the expression level of *RSR1* was low during the filling stages and is lower than that of other genes which were expressed at the same stages such as SPA. One interesting point is that the *RSR1* is highly expressed in the ovary before the anthesis period, and expression continues in all the floral organs. It is unknown whether *RSR1* has any function during floral development. *RSR1* may be functionally conserved during the flowering period; this has been confirmed from the floral organs of *rsr1* mutant or *RSR1* overexpression lines in rice (Fu *et al.,* 2010). In 2016, Liu *et al.,* pointed out the new method to silence *RSR1* transcription factor in wheat grains by using the BSMV-VIGS method. The results indicated that the starch contents, kernel weights, grain length, and width obviously increased in mature grains. Furthermore, the transcription levels of 19 starch synthesis-related enzyme genes were upregulated markedly. That means *RSR1* could be as a negative regulator and temporally regulates the expression of some starch synthesis enzyme genes in bread wheat.

In terms of the *TaQ* gene has specific expression patterns in the spikelet meristem and floral organs. However, by RT-PCR gave no detectable transcript in any seed tissues examined. Generally, The *TaQ* gene expression was investigated in several studies such as (Simons *et al*., 2006) reported about Q in terms of the structural, transcriptional, and regulatory differences between Q and q alleles. In addition, (Fans *et al.,* 2005) explored the expression of Q in different tissues of wheat and discovered *Q* role in floral morphology and domestication. Also, Q was found to be able to control domestication traits such as the spike phenotype, rachis fragility, glume toughness and plant height and flowering time (Faris *et al.,* 2005). Recently, in 2017, Greenwood *et al.,* identified that the three wheat homologous of Q have similar expression patterns across tissues and developmental stages. They showed the highest transcript levels in young spikes (~3-5 mm length) at Zadoks stage 32, when the elongating stem has two nodes (Zadoks *et al.,* 1974). They determined the Q-5A allele is expressed at higher levels than the q-5A (Simons *et al.,* 2006). This result confirmed that *euAP2* genes are likely to be functionally divergent as they control fruit development in tomato, and regulate inflorescence meristematic activity in maize (Zumajo-Cardona and Pabón-Mora, 2016). To date, only a few

genes have been found to be associated with spike development in wheat (Simons *et al*., 2006). However, the challenge now is to identify genes and their regulatory networks may physically interact with *TaQ* during wheat spike development.

As described earlier, miRNAs are a class of regulatory sRNAs involved in gene regulation. In plants, some miRNAs play crucial roles in the various developmental processes, such as the control of root and shoot architecture, transitions from the vegetative to reproductive phase, and the leaf and flower morphogenes (Chen *et al*., 2009; Curaba, *et al*., 2012; Debernardi *et al.,* 2017). In crop species, miR172 has been found to be one of the conserved miRNA in the plant, which regulates *AP2-like* genes that are involved in controlling floral organ identity in rice, maize, and barley (Lauter *et al.,* 2005; Zhu *et al.,* 2011). Furthermore, miR172 is known to play multiple roles in plant development.

In this chapter, it was reported that the expression of miR172 was associated with the early flowering, the inflorescence and during the development of grain. It was detected in the different tissues which confirmed the conservation of the spatial and temporal expression of miRNA172 in the species of the Poaceae, which reflect the conserved role as well. This result consents with the previous study reported that miR172 has a highly conserved role in the system in the regulation of flowering time in the plant (Zhu and Helliwell, 2010).

Conversely, Aukerman and Sakai (2003) reported that it is not clear if miR172 regulates its targets in response to intrinsic developmental age as what suggested by (Simpson and Dean, 2002) or the ambient condition as a response to temperature (Blazquez *et al.*, 2003). However, they confirmed the pathway regulation during the flowering time is controlled by *AP2*- like genes and mi172. For that, they predicted a model temporal up-regulation of *miR172* during seedling stage which down-regulated to *TOE1* and *TOE2* in Arabidopsis. This model further predicts that the role of mi172 during the flowering time by repressing its targets of *AP2*-like genes. The results of the current study showed that mi172 is expressed highly during the vegetative stage and increases sharply during the plant life cycle especially during flowering time. These results are in agreement with (Aukerman

and Sakai, 2003) finding which showed the same expression rate for mi172 in Arabidopsis. However, the single most striking observation to emerge from the data was the expression levels of (*RSR1, Q*) were constant during the development of plant during the vegetative stage and flowering time in different floral organs. However, both Q and *RSR1* have specific expression patterns in the spikelet meristem and seed development stages respectively, which showed a reverse relationship between mi172 and euAP2.

In this study was found that miR172 is significantly up-regulated and expressed at high levels in grains at the beginning of the filling stage. This has an opposite pattern to one of its predicted targets in the wheat/ barley grain, RSR1. This result is similar to the results obtained by (Meng *et al.*, 2013). Furthermore, it was known that the development of grain transition from cellularisation to endosperm cell division at approximately (3-5 DAA), flanking the nucellar projection at (5-14 DAA), and then grain filling of starch and protein in the endosperm at approximately (15 DAA to 25 DAA) (Olsen *et al.,* 1992; Olsen *et al.,* 2001). Therefore, miR172 expression increased highly during the early stage when the grain changes from cellularisation to endosperm cell division and the filling stage which is correlated to the rate of accumulation of protein and starch which increased highly. It is likely that the role of miR172 is a crucial role in the transitioning of the seed from the cellularisation to endosperm development. However, the accumulation of starch in the endosperm decrease during the maturation, this correlates with the level of miR172. Therefore, this data suggests that miRNAs play crucial regulatory roles during grain development in cereal.

Chapter 4

Functional analysis of candidates TFs

Abstract

Background: It is known that most of the regulators involve in the regulation of their target in a network of proteins. This regulation can be achieved directly by binding a specific cis-element of the promoter of genes or indirectly by interacting with other TFs to increase the activity of their targets. This variation in the seed storage proteins and starch synthesis genes regulation is due to the activity of specific transcription factors. So that the action of transcription factors allows for the unique expression of each gene in different cell types and during development. In this chapter, a series of experiments are described that have delivered insights into the function of some TFs in the endosperm. For example, confirming the function of *TaSPA* as the main activator of the storage protein gene by knocking down this gene. In addition, as explained previously the homologous of O2-like/ euAP2 TFs could interact together either *in vivo,* or *vitro* by using different techniques and how those effect on the activation of their target genes.

Results: TaSPA-RNAi lines showed the reduction in the expression of the *TaSPA* gene as expected which affect the quality and physical of grain. In addition, theTFs (TaSPA, TaBLZ1, TaRSR1, TaPBF, and TaMADS4) showed the ability to interact with different proteins by using Y2H (yeast two-hybrid system). These interactions play a significant role to increase the activity of the storage protein gene and starch synthesis genes, which have been confirmed by the transient assay.

Conclusion: The data summarized in this chapter allows us to estimate the main roles of TFs such as TaSPA, TaPBF, and TaRSR1 which are involved in the accumulation of starch and protein in the endosperm to activate their target either by binding a specific motif in the promoter or by interacting with other proteins to increase the activity of their GraphPad PRISM targets.

Chapter 4 Functional analyses of some TFs are involved in the regulation storage protein genes and starch synthesis genes.

4.1 Introduction

TFs have a significant role in the plants in different biological processes during their growth and development. However, several research efforts reported in the literature on gene function analysis by knockout or down techniques which are used to better understand the gene function. One of the genetic approaches can be used to deduce the functional roles of the target genes such as interference (RNAi) technology was used to down-regulate multi-genes (Travella *et al.,* 2006), so that it was used in this project to infer the function of *TaSPA* gene. (RNAi) induced gene silencing by hairpin RNA (hpRNA) vectors, which were composed of a promoter and a spacer in between an inversely repeated sequence of the target gene.

As it is known, there are several transcription factors regulating the expression SSP genes in the grain which are known as endosperm-specific TFs such as homologous of TaSPA which is ZmO2 and PBF. Most of these TFs have been silenced by using RNAi approach to prove their functions as ZmPbf (DOF protein) with (RNAi) only resulted in decreased expression of 22-kD a-zeins and 27-kD g-zeins (Wu and Messing, 2012). Another study was used two mutation in one line such as *Pbf RNAi* and *o2 RNAi* may influence on the content of protein in maize grain (Zhang et al., 2015). So that, compared between the previous studies and the role of SPA as transcriptional activators, we predicted the *spa* mutant can affect either the storage protein, or the starch content in the endosperm, which could explain that the possibility is that either the knockdown of *Taspa* is sufficient to influence on the protein, or other TFs are likely to be necessary for regulating protein genes in the endosperm.

In addition, yeast two-hybrid (Y2H) is an approach that can also be used to get a better understanding of the mechanisms regulating the expression of genes by the interaction between the proteins. Previous studies investigated the interaction between several TFs such as (bZIP family) which have the roles in

regulating various of developmental processes through the interactions between various proteins (Massari and Murre 2000; Gu *et al.,* 2000; Luscher 2001; Wagner 2001; Robinson-Rechavi *et al.,* 2003; Gronemeyer *et al.,* 2004). Yeast two hybrid analysis was used to identify those proteins that bind to proteins of interest. The dimerization is likely to affect the specificity of the DNA-binding and affinity and transactivation properties (Naar *et al.,* 2001). Generally, in this chapter, the possibility of the interactions were tested between candidate of bZIP proteins in wheat and other TFs which are thought to play role in the endosperm development.

Several TFs were identified as regulating genes expression by detecting their specific binding sites (TFBS) on the promoters. As discussed in Chapter 1, several TFs have been identified as regulating gene expression through binding to cis-regulatory specific sequences in the promoters of seed storage protein genes. For example, some TFs such as (ZmO2, ZmPBF and others) are known are responsible for the activation of particular classes of seed storage protein genes (Diaz and Carbonero, 1998), also, ZmO2 has been shown to regulate the 22-kD a-zein and 15-kD b-zein genes by recognizing the O2 box (TCCACGT) in their promoters (Schmidt *et al.,* 1992; Neto *et al.,* 1995). Also, several studies proved the regulation of the expression of seed storage proteins in maize were controlled by PBF (DOF family) which binds the endosperm box by recognizing a specific motif, which is P box (Ueda *et al.,* 1994; Zhang *et al.,* 2015; Yang *et al.,* 2016). In addition, B- and C-hordein promoters in barley have two motifs, while there is only one motif is present in D-hordeins, which is a P motif (TGTAAAG) (Müller and Knudsen, 1993). Consequently, most seed storage proteins promoters have a bifactorial motif that is composed of the closely linked P box and the O2. Zhang *et al.,* (2016) pointed out that both OPAQUE2 (ZmO2) and (ZmPBF) regulated the starch synthesis genes. Moreover, OsbZIP58 in rice regulated starch synthesis genes directly by binding a specific motif (ACGT) (Wang *et al.,* 2013). In addition, AP2/EREBP have been studied in detail, where it is known that some AP2 members regulate starch synthesis genes by binding specific motifs of the targets' promoters. All these proteins can recognize the GCC-box with some differences in instances of variant sequences of the canonical GCC box, with

changes at the 5th and 6th positions of the sequence GCCGCC (Wang *et al.,* 2013).

Consequently, I tried to examine the function of (*TaSPA*) gene, comparing to the *ZmO2* gene. Also, confirm the ability of these TFs to interact each other as network in a specific tissue, in a specific stage of development to control the accumulation of both starch and protein.

To do this, different approaches have been used which can be described as follows:

First, applying a reverse genetics approach to infer the functional roles of *TaSPA* gene in wheat using RNAi approach.

Second, investigating the potential of certain members of the bZIP to undergo either the dimerization via yeast two hybrids.

Third, identifying conserved motifs which are presented in certain promoters of starch synthesis genes and the seed storage protein gene, and then investigate the ability of the number of TFs of bZIP and euAP2 to activate these genes, using transient assay.

4.2 Material and Methods

4.2.1 The material of transgenic lines (TaSPA-RNAi)

This method was used to knock down the *TaSPA* gene (RNAi; RNA-interference) in several lines comparing to the control line (null segregates). The region of *TaSPA* that was chosen to effect the knockdown of this gene, was conserved between three homologous of A, B and D genomes of wheat, such as the 3' UTR region to silence all three homeologues in. Thenit was constructed into Entry construct flanked by attL sequences to be ready for transfer into a NIAB binary vector (pAct-IR-2) with a *nptII* selection cassette via Gateway™ cloning and the rice actin promoter. The rice actin promoter for RNAi silencing was present in the destination vector (pAct-IR-2) **(Fig.4.1).**

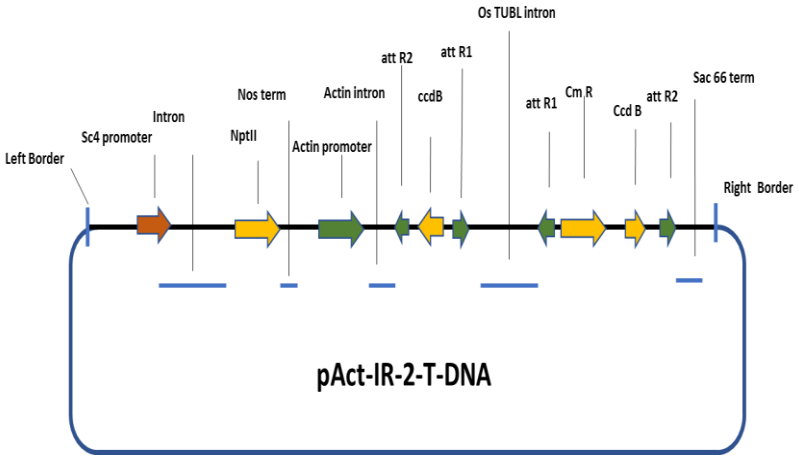


Figure 4.1 TaSPA-RNAi expression clone generated using GATEWAY® technology.

DNA fragment (TaSPA) was 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.

The TaSPA-RNAi construct used for plant transformation is shown in (**Fig. 4.1).** Fielder, a spring wheat cultivar, which was a silenced plant, carried an RNAi knock-down construct containing two short segments of *TaSPA* in an inverted orientation, separated by a spacer region which is (Os TUBL intron) between the TaSPA-RNAi fragments, inserted in sense and antisense orientation by recombination using the Gateway technology (Invitrogen). Sequences encoding the Inverted-repeat were cloned into (pAct-IR-2-T-DNA) intermediate vector,

which contains an enhanced rice actin promoter and a neomycin phosphor transferase gene forms the selectable marker.

4.2.2 Design of attB site primers

The primers were used to amplify a specific region of *TaSPA* gene (340 bp) of 3' UTR region. Those primers were flanking with partial of attB (attB1, attB2) recombination sites for Entry clone vector **(Appendix 7; Table 7.4.5).**

4.2.2.1 *Amplification and purification of PCR products*

cDNA was amplified by PCR using a high fidelity Taq polymerase (Q5, NEB). The PCRs were used for the first step of Gateway® recombination cloning. Specific primers containing attB recombination sites were used for PCR1. Then PCR product was used as a template with attB adapter primers in a 50 µl PCR reaction. All primers sequences were shown in **(Appendix 7; Table 7.4.5).** The reactions and PCR program details are shown in **(Table 4.1).**

4.2.2.2 *BP recombination reaction*

BP recombination reaction was performed to transfer an attB-PCR product (The target of interest gene) into an attP containing entry vector was the pDONOR221 vector (Invitrogen). The entry construct was ready for transfer into a NIAB binary vector with a *nptII* selection cassette via Gateway cloning.The seeds of transformation lines were germinated on moist filter paper in Petri dishes in the dark at 4 °C for 48 h. Then they were sown into pots containing peat and sand as described in chapter 3, were grown in a controlled environment room with 16 h light at 20 °C and 8 h dark at 15 °C with 70 % humidity. Light levels were 300 µmol m-2 s -1. For RNA extraction, the grains were immediately placed in Eppendorf tubes, frozen in liquid nitrogen and stored at -80 °C.

4.2.3 Extraction of starch and measurement the starch content.

The method of starch extraction of each line was used as described by Trafford *et al.,* (2013), while the starch content was measured using the Megazyme Total starch assay Kit (AA/AMG) (Megazyme International, Ireland).

4.2.4 Yeast two hybrid analysis

This expermint was performed following the protocol of as Matchmaker Gold Yeast Two-Hybrid System.

4.2.4.1 *Designing primers for PCR based cloning*

The primers used for each gene are shown in **(Appendix 7; Table 7.4.5).**

4.2.5 Transient expression assay.

Luciferase transient assay is known as a simple, rapid, and sensitive method for the study of promoter transactivation (Koncz *et al.,* 1990). Firefly luciferase is an enzyme has the ability to catalyse adenosine triphosphae (ATP) dependent oxidation of its substrate luciferin to produce light (Deluca and McElroy, 1978). Renilla luciferase is an enzyme that can catalyze a luminescent reaction by utilizing O2 and coelenterate luciferin (coelenterazine) (Matthews *et al.,* 1977). The combination of both reporter genes in a single assay, often referred to as a dual luciferase assay, provides a powerful method that can be internally controlled by normalizing the activity of firefly luciferase with that of Renilla luciferase (Sherf *et al.,* 1996). Then by using the Agrobacterium-mediated transformation of tobacco leaf (Sparkes *et al.,* 2006).

4.2.5.1 *Analysis of the Promoters of TaSSIII-1b, TaSSII-1b and TaSBE-1b genes, and TaLMW in wheat.*

4.2.5.1.1 *Genes set identification*

The starch synthesis genes (*TaSSIII-1b*, *TaSSII-1b* and *TaSBE-1b* genes), and *TaLMW* (Low molecular weight) were chosen. A region of 2 Kb upstream the start codon of the selected genes, then they were blasted in the Regulatory Sequence Analysis Tool (Plant PAN2.0; (http://PlantPAN2.itps.ncku.edu.tw) (Chow *et al.,* 2015),is available online and it can be used to predict the transcription factor binding sites (TFBS) in promoters of co-regulated genes.

4.2.5.1.2 *Polymerase Chain Reaction (PCR) and its applications*

4.2.5.1.2.1 *Oligonucleotide primer design.*

To amplify a fragment of the promoters and cDNAs, oligonucleotide primers were designed for amplifying either genomic DNA or cDNA and plasmid DNA. Primers were designed to be 18-25 bases in length and with a Tm of 50-65ºC and when possible, with a G or C at the 3' end. Pairs of primers were designed to have a Tm difference of ≤ 5ºC. Primers designed for use in the Gateway® recombination cloning of PCR amplification products into attP-containing pDONR vectors had partial attB adapters added at the 5' end **(Appendix 7.Table 7.4.5).**

4.2.5.1.2.2 *PCR conditions*

The fragments were amplified by PCR using a high fidelity Taq polymerase (Q5, NEB). The PCRs were used for the Gateway® recombination cloning. The reaction buffer was supplied with each enzyme. dNTP sets at a concentration of 100 mM were purchased from Invitrogen™ and diluted with nuclease-free water to a stock concentration of 10 µM. For each PCR reaction mixes. Plasmid DNA and gDNA/cDNA used in PCR were diluted 1:100 in nuclease-free water. The PCR products were used as a template for a 50 µl PCR reaction. The reactions and PCR program details are shown in TProfessonal Basic Gradient (Biometra) thermal cyclers. PCR thermocycling conditions for Q5 Taq program is given in section **(4.2.2.1)**

4.2.5.1.2.3 *Colony PCR (cPCR)*

cPCR was used to check colonies of *E. coli* for plasmid DNA and/or successful Gateway® recombination by using one vector backbone-specific primer and one insert-specific primer. One single colony was picked up and dipped into 10 µl of cPCR reaction mix. Cells remaining on the pipette tip were streaked onto antibiotic(s)-containing LB medium and grown as required. In terms of the PCR reaction mixes were placed in a PCR thermal cycler and thermo-cycling was performed as described in section 3.2.2.1.

4.2.5.2 *PCR for Gateway® recombination cloning*

Promoter fragments and coding sequence (CDS) were amplified for use in Gateway® BP recombination cloning (Invitrogen™) using a two-step PCR workflow.

4.2.5.3 *Transient transformation of Nicotiana tabacum*

According to Sparkes *et al.,* (2006), *N. tabacum* leaves were transiently transformed for use in the dual luciferase assay by using agro-infiltration. *A. tumefaciens* strains harbouring the appropriate report, effector or Renilla luciferase control constructs were inoculated in 5 ml of antibiotics-containing LB medium and grown overnight with shaking (200 rpm) at 28ºC **(Fig.4.2).**

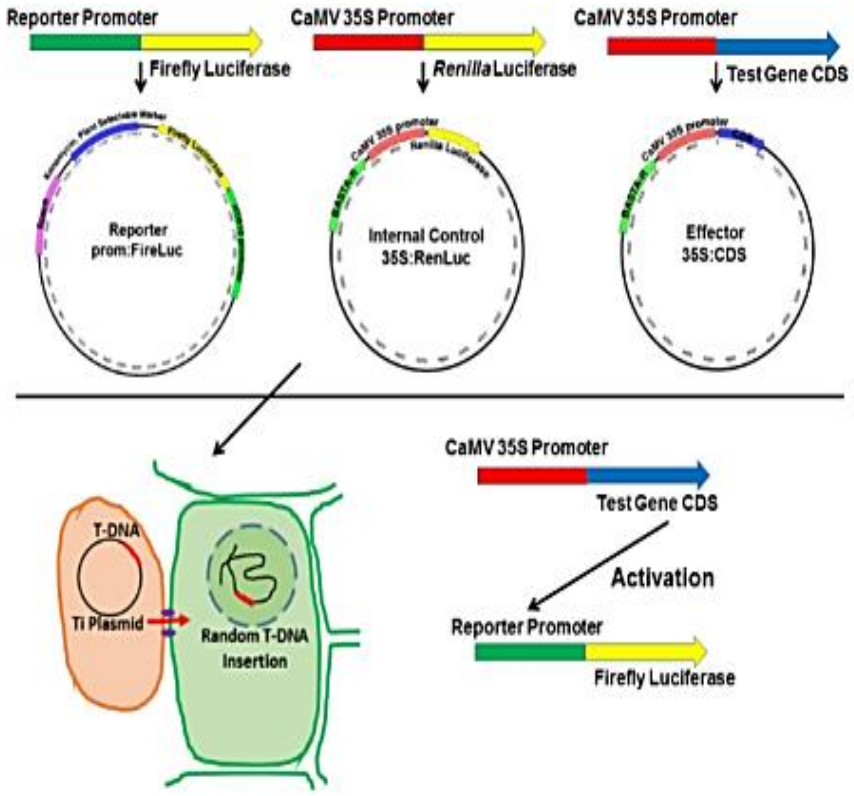


Figure 4.2 Diagram showing the three types of constructs are used for the dual-luciferase transient assay.

(Reporters, effectors, and an internal control). The blue part is insert of gene is driven by CaMV 35S promoter (Red); the green part is the target of promoter. The internal control is Renilla luciferase driven by 35S promoter. They are then delivered through Agrobacteria (From Păcurar *et al.,* 2011).

4.2.5.1 *Data analysis*

The transient assay data was performed on the means of technical replicates using the two-tailed t-test assuming equal variance. Normalised dual luciferase activity (FLuc/RLuc) was calculated by dividing Firefly luminescence (FLuc) with the Renilla luminescence (RLuc). Results from statistical tests were considered significant when $\alpha \leq 0.05$.

4.3 Results

4.4 TaSPA-RNAi transgenic plants

The plant transformations were carried out at NIAB organization. There were 37 lines confirmed the by PCR and/or qPCR copy number analysis for the presence of the nptII-selectable marker gene. 13 putative transgenic lines have been chosen with different numbers of copies in three replicates for each line. Genotyping showed that several numbers of TaSPA-RNAi lines were tested, had positive results for the neomycin phosphor-transferase II (nptII) selection marker gene **(Table 4.1)** shows the transgenic lines which have been generated.

Table 4.1 Summary of TaSPA-RNAi wheat lines generated.

No. of lines	qPCR copy number	Backbones	NptII insert
CW53.16	1	pEW280	+ (positive)
CW53.19	1	pEW280	+(positive)
CW53.25	1	pEW280	+(positive)
CW53.27	1	pEW280	+(positive)
CW53.29	1	pEW280	+(positive)
CW53. 1	4	pEW280	+(positive)
CW53.2	4	pEW280	-(Negative)
CW53.3	+4	pEW280	-(Negative)
CW53.18	+4	pEW280	+(positive)
CW53.11	2	pEW280	+(positive)
CW53.12	2	pEW280	+(positive)
CW53.con1	0	pEW280	-(Negative)
CW53.con2	0	pEW280	-(Negative)

4.4.1 Expression profile

All lines listed in table 4.8 were tested for expression of TaSPA and TaGAPDH control. Full-length grains (20 DAA) were collected for RT-PCR using specific primers for TaSPA in all lines, which have positive nptII result **(Figure 4.3).**

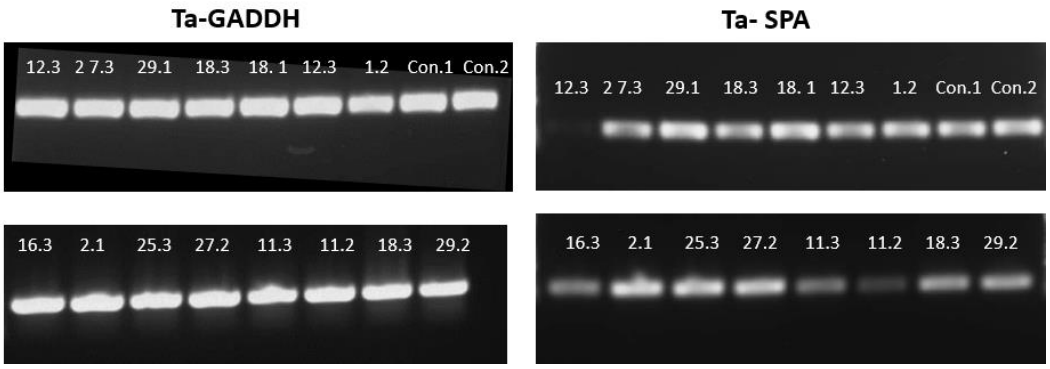


Figure 4.3 RT-PCR result for TaSPA-RNAi lines by using specific pair of TaSPA primers, and reference gene is Ta-GAPDH, comparing with control (WT).

The results of RT-PCR showed that there were some variations in expression *TaSPA* in some transgenic lines **(Figure 4.3).** Comparing the WT and transgenic lines, the TaSPA-RNAi lines showed the reduction in the expression of *TaSPA* in some lines compared to the controls. To investigate further, this result was confirmed by quantitative analysis of the expression levels of the *TaSPA* gene using quantitative real-time PCR (qPCR) at 20 days after the anthesis. The primers for the RT-PCR and qPCR were designed to amplify different regions of *TaSPA* to allow a more accurate quantification of the endogenous transcripts. Wheat the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene was used as an internal control. These results of RT-PCR were correlated with those of qPCR. In terms of the results for qPCR indicated that the reduction in the expression in certain lines such as line no. CW53.12, CW53.11, and CW53.16, compared with the reference gene (Ta-GAPDH) (**Figure 4.4**). **(Table 4.2)** shows a summary of the lines which have been grown, and the number of copies of genes in each line, compared to the expression level.

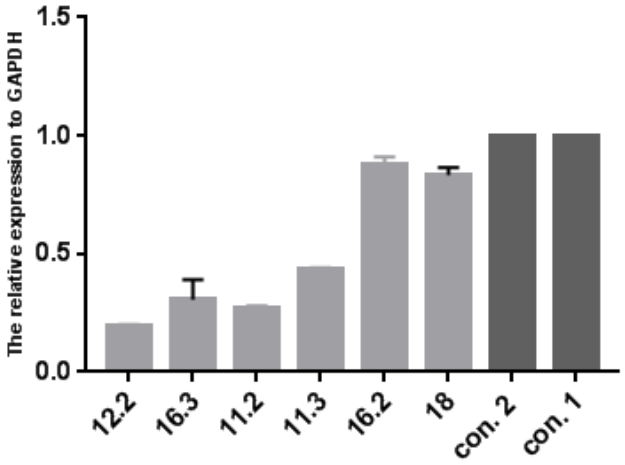


Figure 4.4 qRT-PCR result for TaSPA-RNAi lines of wheat comparing with control (WT).

Table 4.2 Summary for the result of qPCR for some TaSPA-RNAi lines.

(+) highly reduction; + slightly reduction; - no reduction.

No. SPA-RNAi line	No. of copy	Result of qPCR	No. SPA-RNAi line	No. of copy	Result of qPCR
16.2	1copy	**(+)**	29	1 copy	**-**
16.3	1 copy	**+**	18	4 copies	**+**
29	1 copy	**-**	27	1 copy	**-**
12	2 copies	**+**	11.3	2 copies	**(+)**
11.2	2 copies	**(+)**	2.1	4 copies	**-**

4.4.2 Phenotypic analysis of TaSPA-RNAi lines

The lines with low expression of *TaSPA* gene showed differences from the control (WT) in numerous features such as flowering time, which was approximately three weeks late in lines such as CW53.16, CW.53. 25, CW.53.27 and CW53.29. Furthermore, the second phenotype that was found between the lines the length of the tillers, where lines with mutant TaSPA-RNAi showed shorter than the control **(Fig.4.6).** In addition, the number of tillers varied between the mutant lines, which was two in some lines whilst others had three. This variation affected the total number of seeds at the end. The number of seeds were very low in most mutant lines, such as CW.53.11.2, CW.53.11.3 and CW.53.16.2, compared to the control **(Table 4.12; Fig.4.5).**

4.4.3 Physical criteria for grain

Comparing between TaSPA-RNAi lines and controls can be found the significant difference in physical characterization of grain**.** All the grains were collected around 35 DAA (Dry mature grain).

One of our observation which is the size of grain (The length and width) and yield. The size of the grains was measured with a Vernier caliper. The TaSPA-RNAi plants showed a slightly smaller in the size, which was around (2.60 -3.2 mm) in width compared to control plants (3.2 mm). Also, the grain length in mutant lines was around (5.5 -5.8 mm) **(Table.4.3).**

In conclusion, we estimated that the TaSPA-RNAi has a big impact on the physical characters of grains such as mass and size as shown in **(Table.4.3)** comparing to controls lines.

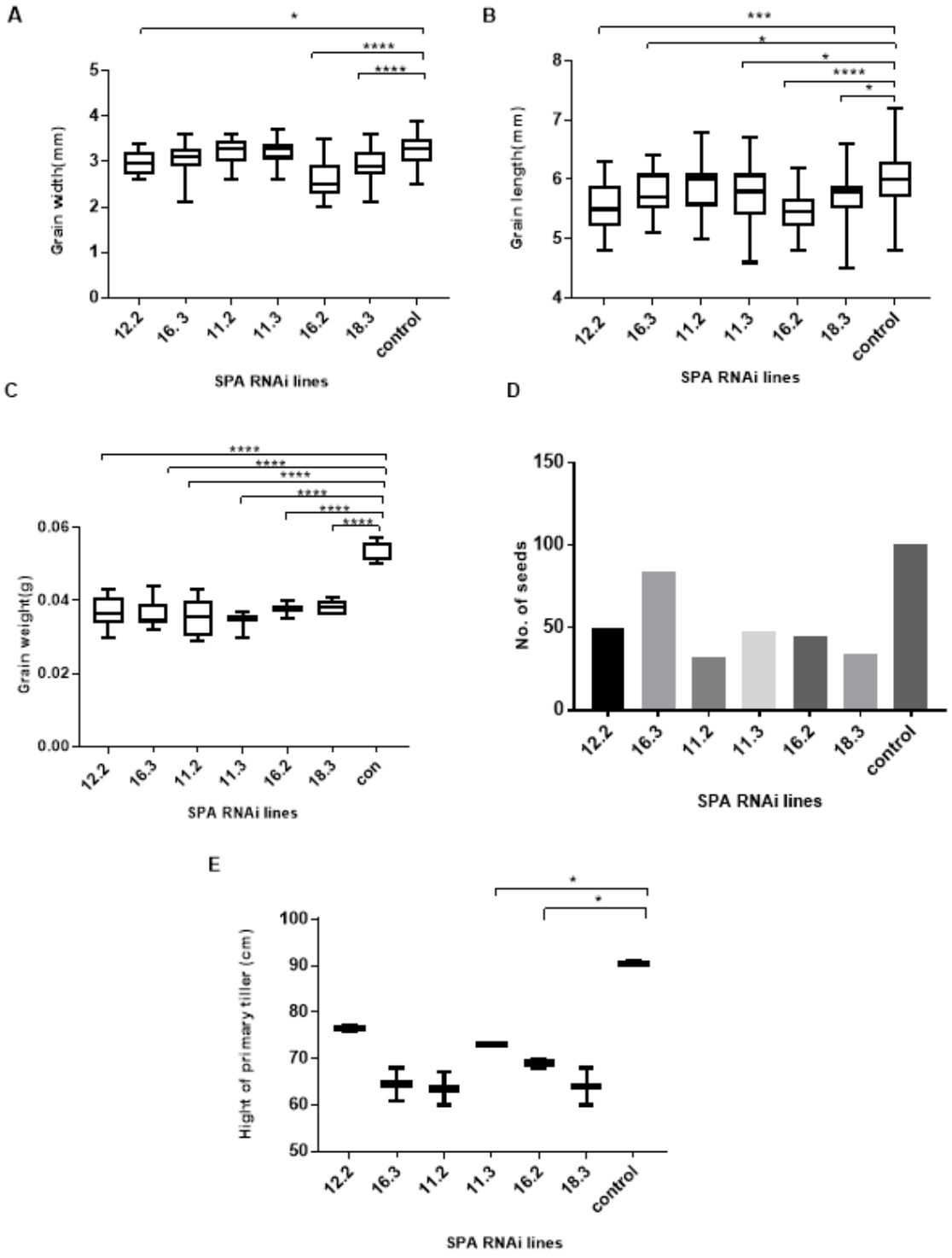


Figure 4.5 The graphs illustrate the the phenotypic features of SPA-RNAi lines.

A) Grain width, **B)** Grain length, **C)** Grain weight, **D)** Number of seeds, **E)** Hight of primary tieller.

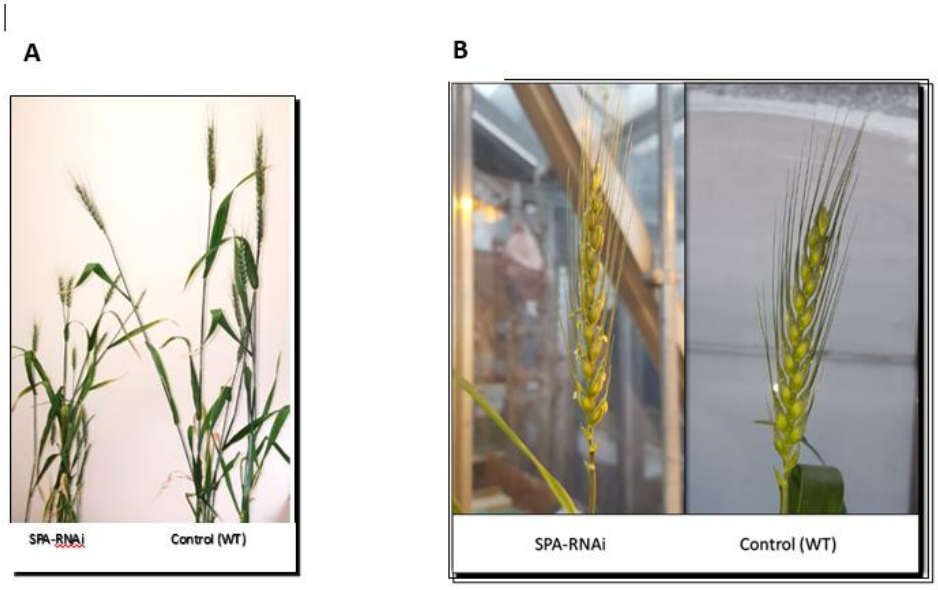


Figure 4.6 (A) Comparing between TaSPA-RNAi (CW53.11.2) (left) and control (right) in the length (80 days after anthesis (DAA).

(B) Comparing between TaSPA-RNAi (CW53.11.2) and control (WT) in the spikelet.

4.4.4 The chemical (quality) properties of wheat grain

Starch was extracted from the grains of the all mutant lines. Several starch properties were examined **(Table 4.3).** Starch samples were analyzed by an ImageJ software to measure the size of starch (A and B granules) **(Fig.4.7).** Also, the size of the protein bodies was measured **(Table 4.3); (Fig.4.8).**

In addition, the total starch analysis was analyzed using matured dry seeds. The total amount of starch was measured by enzymatic methods using the Total Starch Assay Kit (Megazyme International). Generally, the result showed the variation in the total starch comparing with the wild type (control which was around 64.8%, while other samples have around 56.3 %- 51.3 %) which really less than the control and this result was correlated with the grain weight. To clarify the reason for this reduction in the total of starch is likely to be there is the change in the expression level of various starch biosynthetic enzyme genes which are closely associated with the physicochemical properties of starch in wheat endosperm. For that, it needs more analyses to confirm this point.

Table 4.3 Growth metrics and starch characteristics for TaSPA-RNAi lines.

	Con.	18.3	16.2	11.3	11.2	16.3	12.2
Growth metric							
Height of primary tiller(cm)	90	68	70	73	50	68	76
No. of seeds	100	58	44	48	40	84	50
Grain weight(g)	0.054	0.038	0.037	0.034	0.035	0.036	0.036
Grain length(mm)	5.8	5.7	5.5	5.82	5.75	5.58	5.54
Grain width (mm)	3.23	2.95	2.60	3.1	3.2	3.11	2.97
Grain area (mm^2)	15.73	13.81	11.3	15.0	15.4	14.35	13.45
Starch Content							
Starch content %	64.8	53.5	55.8	51.3	54.8	55.9	56.3
A-type starch granules	120± **24.4**	113± **20.4**	99.59± **20.42**	102± **30.08**	103± **29.5**	100± **28.35**	90± **23.0**
B-type starch granules	55± **26.7**	47.3± **24.8**	51.37± **10.74**	38± **14.37**	41± **15.05**	34± **9.70**	38± **29.7**
Protein bodies size							
Area (μm^2)	55.3± **6.2**	43.80± **4.37**			38.10± **9.3**		54.76± **3.56**
long axis length (LAL) (μm)	18.5± **1.32**	14.06± **2.4**			14.2± **1.3**		17.3± **1.33**
short axis length (SAL) (μm)	13.11± **2.08**	8.72± **1.40**			8.9± **1.2**		9.4± **1.76**

Data represent means ± SD of three independent replicates.

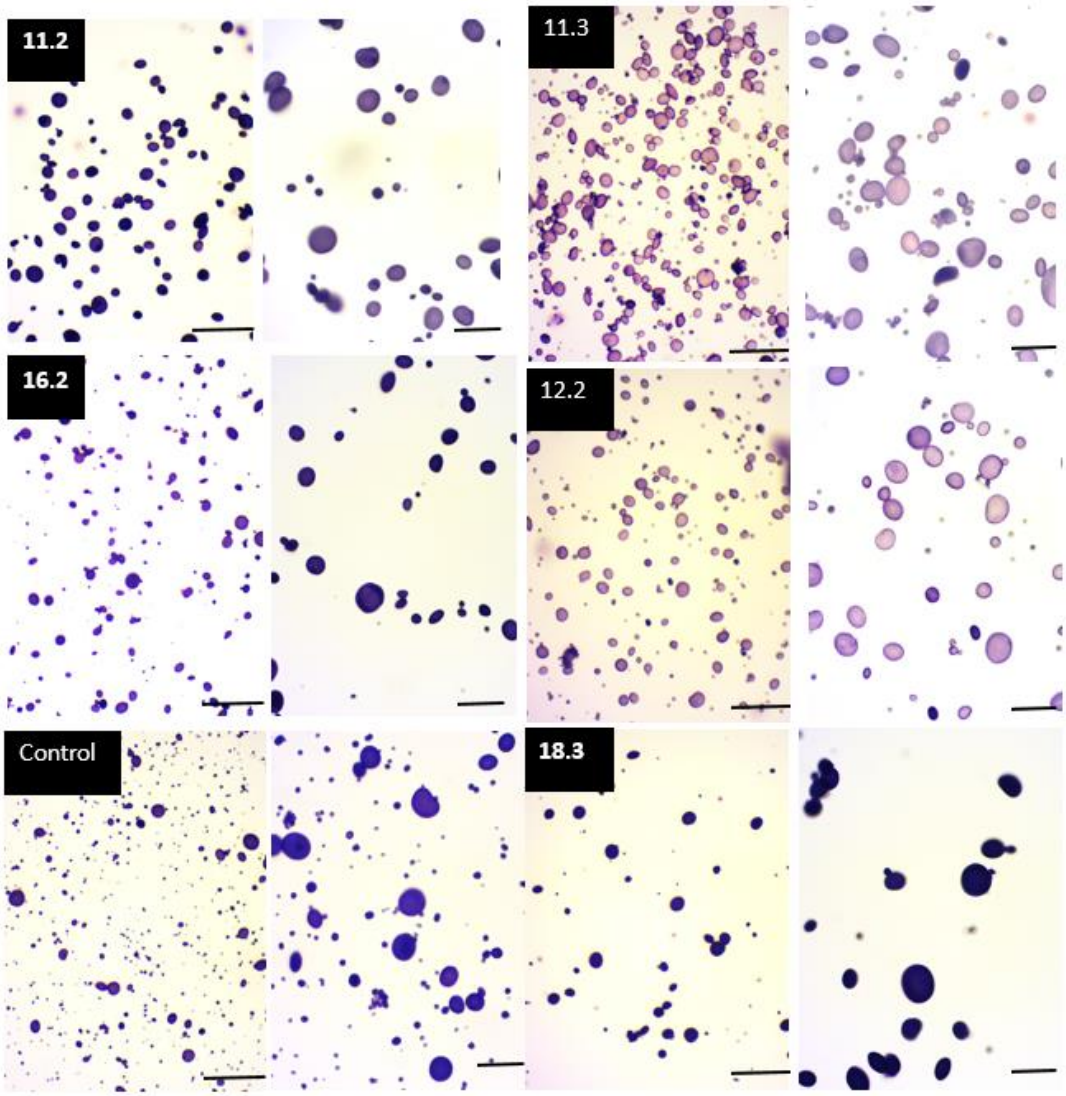


Figure 4.7 Micrographs of starch granules stained with iodine solution by using Light microscopy. (x40) and (x100) to detect the starch granules in the mature grains of TaSPA-RNAi lines.

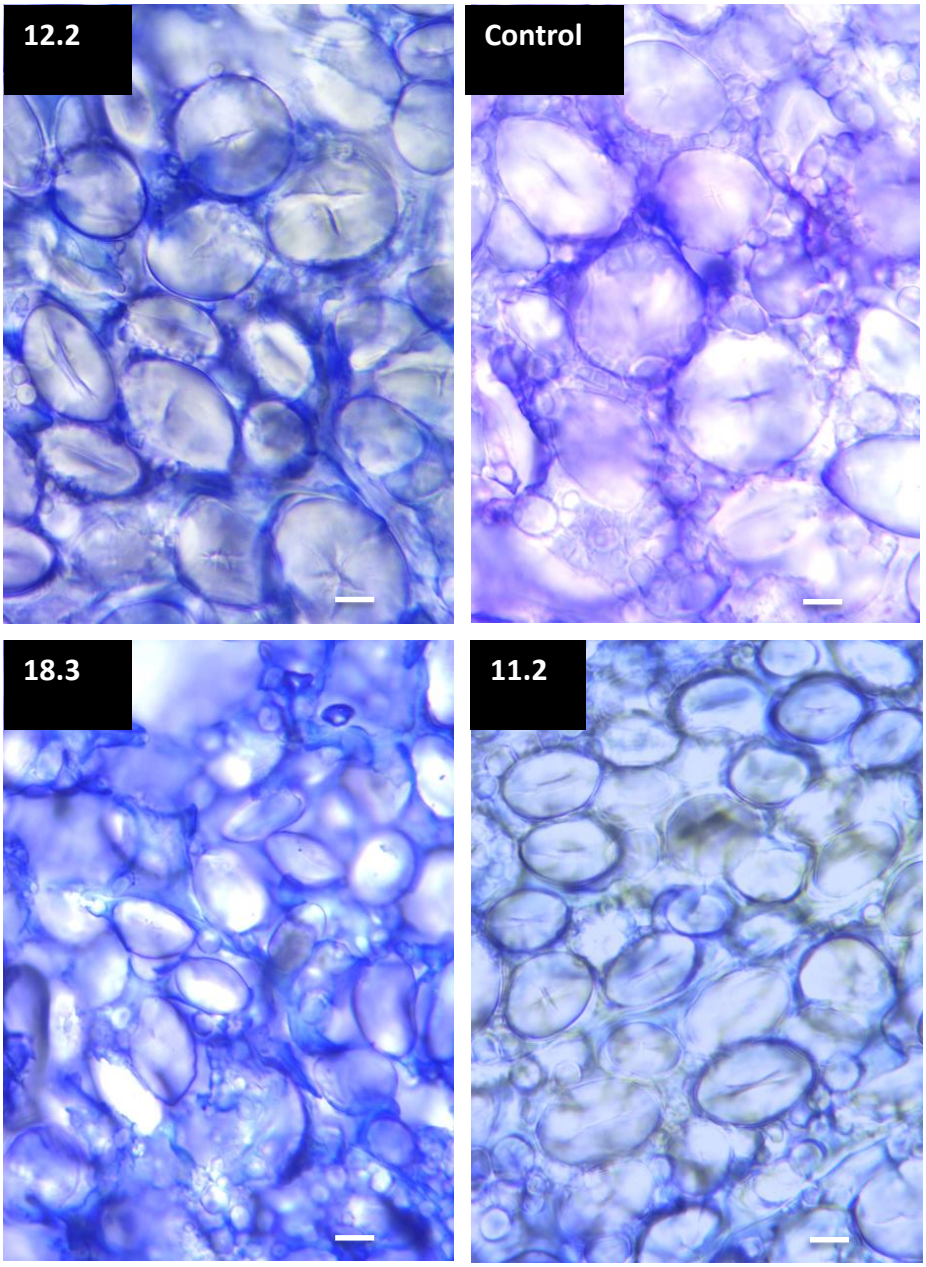


Figure 4.8 Micrographs of sections of mature grains stained with coomassie brilliant blue R250 by using light microscopy, (x100) for protein bodies in SPA-RNAi lines.

4.4.5 Analysis of the putative promoter regions of some genes that express during the filling stage in the endosperm.

As it is known that genes with similar expression patterns, such as co-expressed genes likely share common functionalities; so that, one might infer that genes with similar functions could have similar cis-regulatory elements. From this point, it can be said that the prediction of the common DNA sequence motifs allows the determination of tissue-specific or stress-responsive expression patterns of genes (Saeed *et al.*, 2006), and can reveal whether the genes are closely correlated with cis-regulatory elements in their promoter regions (Walther *et al.,* 2007; Le *et al.*, 2012). To further understand transcriptional regulation and the potential functions of some genes (*TaSPA, TaBLZ1, TaRSR1* and *TaPBF*), and (*HvBLZ2, HvBLZ1, HvRSR1* and *HvPBF*) which have expression patterns in endosperm during the filling stage of the endosperm, the cis-element which were located upstream of promoter regions within 2000 bp were identified (Fang *et al.*, 2008; Wu *et al.*, 2012).

From the analysis of the promoters of these genes, numerous motifs were detected by blasting the promoters of these genes in the Plant Promoter Analysis Navigator PlantCARE, a database of plant promoters and their cis-acting (bioinformatics.psb.ugent.be/webtools/plantcare/html/) (For example, S000454, S000314 and S000028 are present in all genes, whether from wheat or barley. However, some motifs were specific to certain genes such as S000077 and OPAQUE2ZMB32 for *HvBLZ2*, whilst S000002 and 300MOTIFZMZEIN were specific to *HvPBF* only **(Table 4.4).** *TaBLZ1* was found to share these genes with several motifs, which showed that those genes with similar expression patterns share common functionalities. In addition, this analysis confirms the close relationship between the orthologous in wheat and barley.

Table 4.4: Summary of some cis-elements which present in the promoter regions of (SSIII, SSII, SBE, LMW) in barley and wheat. (1) indicates the motif presents, while (0) the motif was absent.

ID motif	Factor or Site Name	Annotation	Motif	HvBLZ2 TaSPA	HvRSR1 TaRSR1	HvBLZ1 TaSPA	HvPBF TaPBF
S000454	ARR1AT	Response regulator	GATT	1-1	1-1	1-1	1-1
S000314	RAV1AAT	RAV1; AP2; VP1; B3 (AP2-family)	CAACA	1-1	1-1	1-1	1-1
S000028	CAATBOX1	found in legA gene of pea, seed	CAAT	1-1	-1-1	-1-1	0-1
S000310	WBBOXPCWRKY1	W box; WRKY Found in amylase gene in sweet potato, alpha-Amy2 genes in wheat, barley, and wild oat, PR1 gene in parsley, and a transcription factor gene in Arabidopsis;	TTTGAC	1 -1	0-0	1-0	0-1
S000408	MYB1AT	MYB recognition site found in MYB; rd22BP1; ABA; leaf; seed; stress	AACCA	1-1	1-1	1-1	1-1
S000077	OPAQUE2ZMB32	opaque-2 binding site" of maize (Z.m.) b-32 (type I ribosome-inactivating protein gene; O2; O2S; O2S and GARE form a gibberellin response complex (GARC)(Rogers and Rogers, 1992)	GATGAYRTGG	-1-1	1-1	1-1	1-1
S000265	DOFCOREZM	Core site required for binding of Dof proteins in maize (Z.m.); Dof proteins are DNA binding proteins, with presumably only one zinc finger, and are unique to plants; Four cDNAs encoding Dof proteins, Dof1, Dof2, Dof3 and PBF, have been isolated from maize; PBF is an endosperm specific Dof protein that binds to prolamin box; Maize Dof1 enhances transcription from the promoters of both cytosolic orthophosphate kinase (CyPPDK) and a non-photosynthetic PEPC gene; Maize Dof2 supressed the C4PEPC promoter;	AAAG	1-1	1-1	1-1	1-1

4.4.6 Yeast two Hybrid analyes

According to Onãte *et al.* (1999), HvBLZ2 could strongly interact with HvBLZ1. In addition, ZmO2 and the ZmPBF could activate the promoter of storage proteins (Zhang *et al.,* (2015), also ZmO2 can interact with ZmMADS47 for the same purpose (Qiao *et al.,* 2016). From this point, the interaction between TaSPA and its paralogous TaBLZ1 wheat and other TFs could play a role during the filling stage in the grain were confirmed in the current study by using (Yeast two-hybrid system) (Clontech). This is based on when bait and prey interact to activate the transcription of reporter genes **(Fig.4.9)** (Fields and Song, 1989).

In this study, TaSPA, TaPBF, and TaRSR1 were chosen as bait proteins, they were cloned into the EcoRI and BamHI sites of the pGBKT7 DNA-BD Cloning Vector (“bait” construction), which were interact with prey proteins, such as different fragments of TaBLZ1, TaPBF, and TaRSR1 which had been inserted into the EcoRI and BamHI sites of the pGADT7 vector **(Fig.4.10).** All these constructs were transformed into yeast strains (*Saccharomyces cereviseae*).

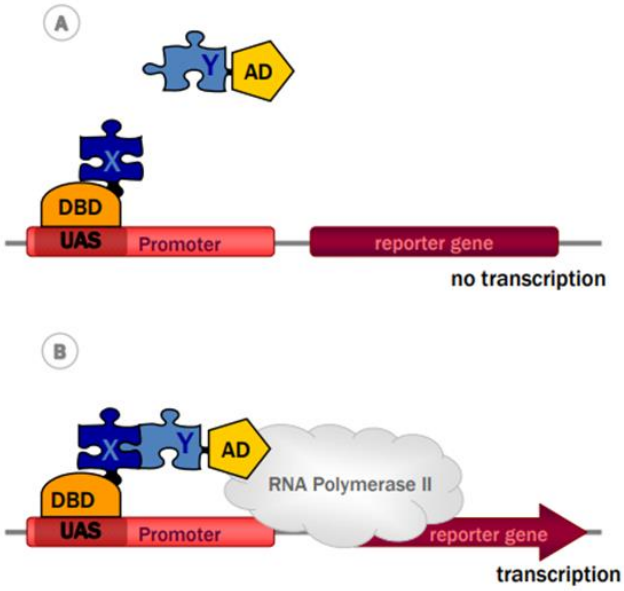


Figure 4.9 Schematic illustrates the mechanism of Y2H to activate the reporter gene (From; Brückner et al., 2009).

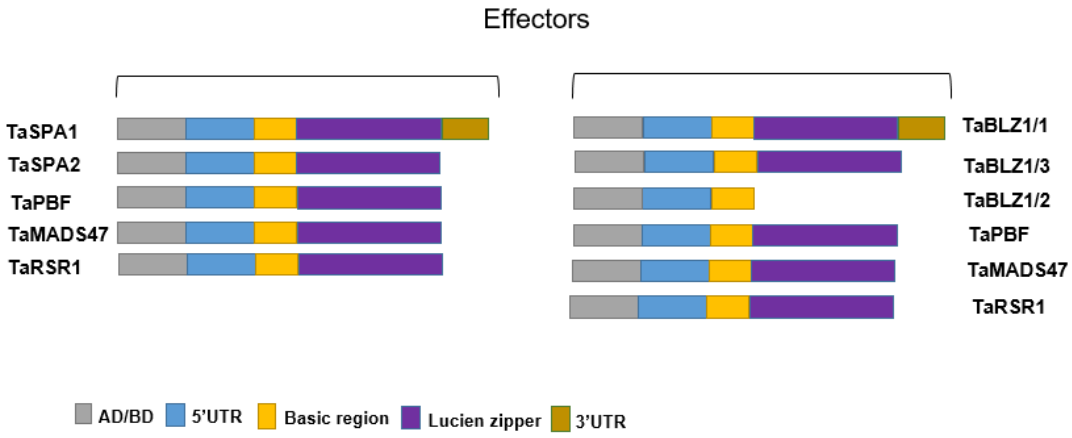


Figure 4.10 Schematic diagram for both the effectors construction which were used for BD (Binding domain) and AD (Activation domain).

4.4.6.1 *Testing Bait for Auto-activation.*

It should be ensured that the bait cannot auto-activate without prey by developing the color in a different medium; a growth test was accordingly conducted to verify that the bait cannot activate the promoter on its own without prey on selective media. The results of this test showed that pGBKT7 containing the target gene did not turn blue in SD/–Trp plates (SDO), and SD/–Trp/ X-a-Gal plates (SDO/X), but rather gave white or blue colonies on selective media, respectively, which means that it is not self-activating. However, there is no growth on SD/–Trp/X-a-Gal/AbA plates, which confirms the positive result. The control constructs were used in parallel with the targets to confirm the result **(Fig. 4.11).**

A

B

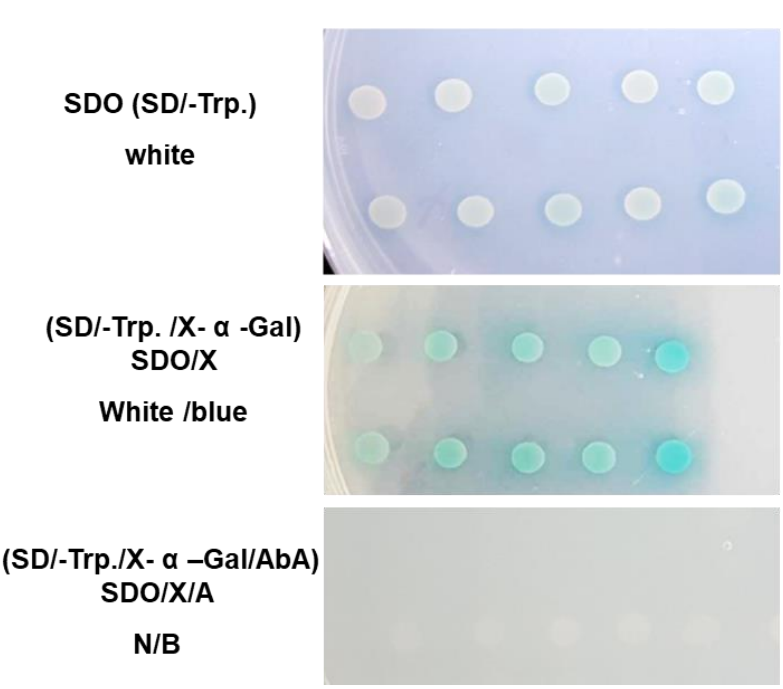


Figure 4.11 (A) The result of mating controls (positive control and Negative control), which was performed before the transformation in different selective media.

B) Colour development to test the bait self-activation for (Ta SPA/A, TaSPA/B, TaPBF, TaRSR1 andTaMADS47).

4.4.6.2 *High Efficiency Yeast Transformation and screening.*

All ORFs individually as a DNA-binding domain fusion (bait/ prey) were cloned in pGBKT7, pGADT7 vectors, respectively, using two yeast strains, which are Y2H for the binding domain (BD, bait) and Y187 for the activation domain (AD, pery). They were mated in different selective media **(Fig.4.12).**

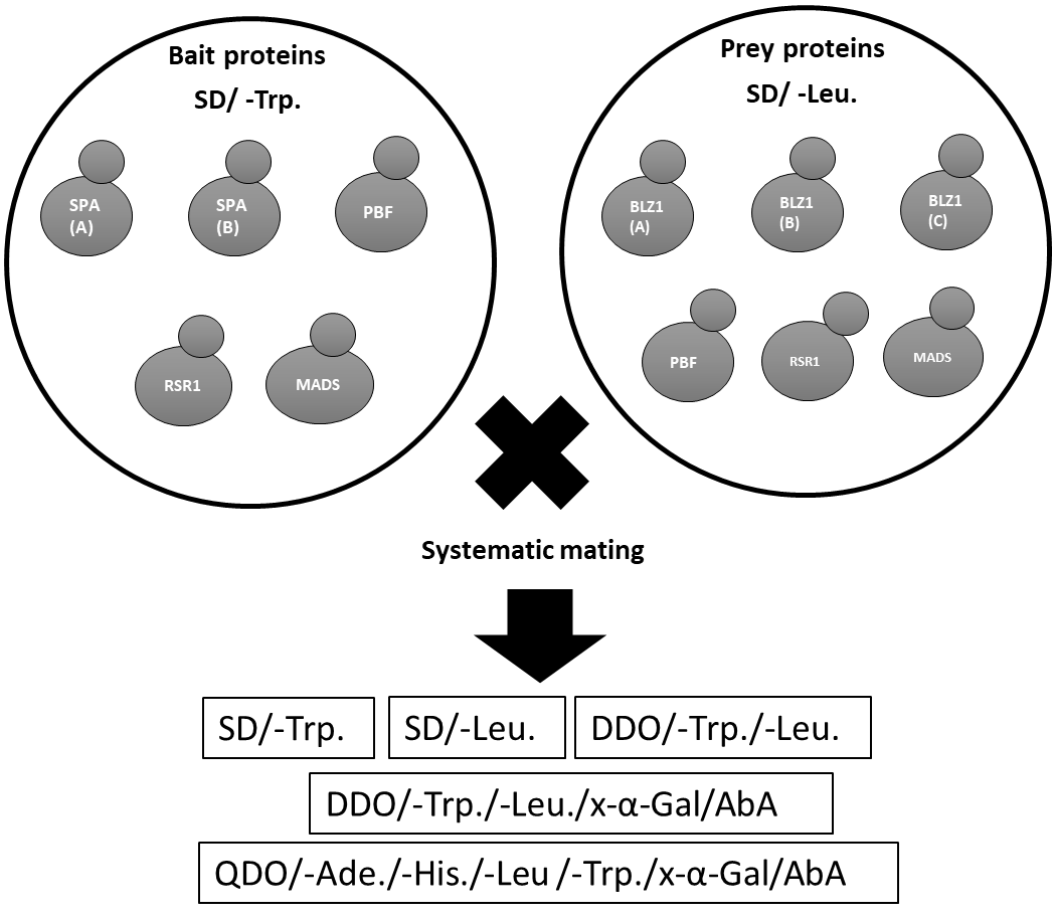


Figure 4.12 Outline of the overall two-hybrid analysis.

These bait and prey clones were mated with each other, and the diploid cells formed were selected for the simultaneous activation the reporter genes.

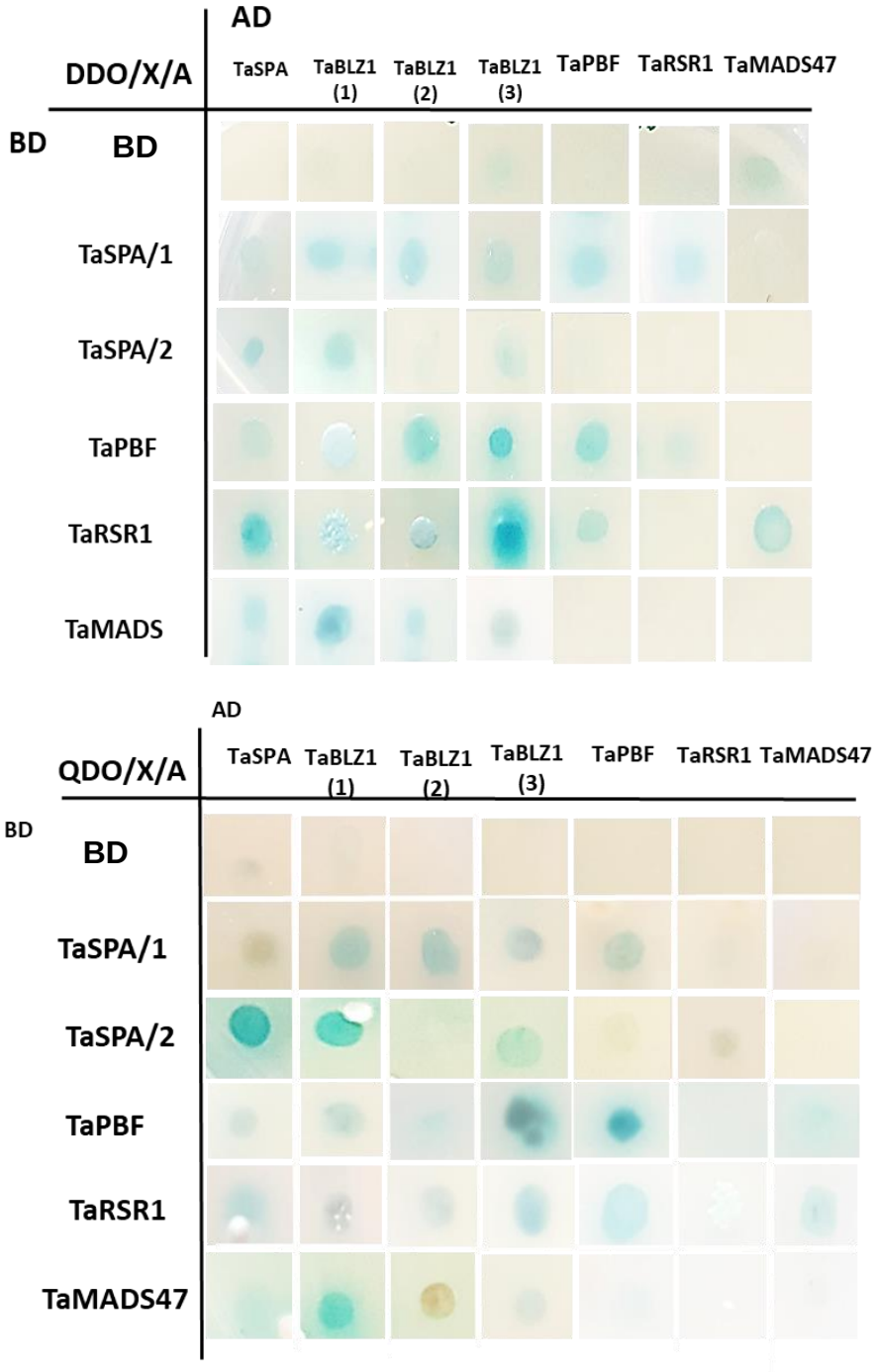


Figure 4.13 Prototrophic growth of the yeast two-hybrid strain Y2H on media.

(DDO/XA and QDO/X/A) indicate hetero-dimerization of the analysed different specific endosperm TFs partners on QDO (SD/-Leu/- Trp/-His/-Ade) plate, monitored by comparing it to the positive and negative controls.

Table 4.5 The summary of the results of the Interaction pattern of TFs (TaSPA, TaBLZ1, TaPBF, TaRSR1 and TaMADS47) which obtained by yeast two-hybrid analysis.

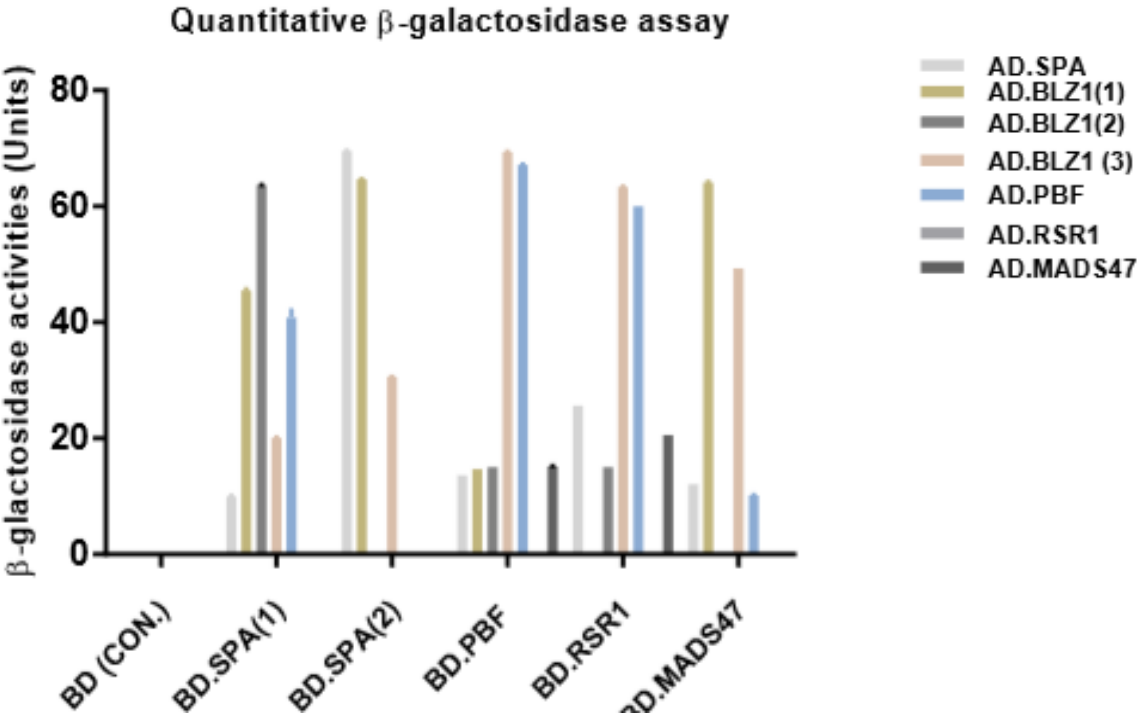
(-) no interaction/activation; + weak interaction; (+) strong activation.

	BD	bZIP (C group)			DOF	AP2/ERBP	MADS-box
AD	**Ta SPA**	**Ta BLZ1/1**	**Ta BLZ1/2**	**Ta BLZ1/3**	**Ta PBF**	**Ta RSR1**	**Ta MADS47**
Ta SPA	-	+	+	+	+	-	-
Ta SPA	(+)	(+)	-	+	-	-	-
Ta PBF	+	+	+	(+)	(+)	-	+
Ta RSR1	+	-	-	(+)	+	-	+
Ta MADS47	+	(+)	-	-	+	-	-

After plating on (DDO/-Trp./-Leu) plates for growth selection, large colonies were selected and streaked again onto plats (DDO/-Trp./-Leu./x-α-Gal/AbA) low stringency, and (QDO/-Ade./-His./-Leu/-Trp./x-α-Gal/AbA) high stringency.
From this plate, it can observe the positive result by develop color in the different medium. A positive interaction between the bait and the prey was indicated with the presence of a blue colony on (DDO/-Trp /-Leu /x-α-Gal/AbA) and a corresponding blue colony on (QDO/-Ade./-His./-Leu /-Trp./x-α-Gal/AbA) **(Fig.4.13).** Blue color developing indicates activation of the reporter genes which encode the enzyme beta-galactosidase which responsible for blue colony with x-α-Gal. From the results, it was investigated the possibility of (TaSPA, Ta BLZ1) bZIP members to interact between them and other TFs. They showed a strong activation capacity which refer to the ability for these genes to form hetero-dimerization. Also, the interaction between (TaSPA, and TaBLZ1) members and Ta PBF was detected, TaPBF (BD) showed an interaction with TaSPA, but it was very weak, and TaBLZ1. However, they did not show any interaction with TaRSR1 or TaMADS47, while TaRSR1 (BD) was activated with TaPBF (AD).

In terms of TaMADS47 (BD) showed an interaction with TaSPA and TaBLZ1, though only weakly. Furthermore, TaRSR1 (BD) showed interactions with all the examined TFs, but again these interactions were weak. **(Table 4.14)** In agreement with this, quantitative β-galactosidase assays revealed protein-

protein interaction was elucidated quantitatively by measuring lacZ activity **(Fig.4.14).** Similarly, TaSPA interacted with other proteins compared with its orthologous HvBLZ2 and ZmO2. Although the protein sequence analysis for the bZIP family in the Poaceae confirms the abilities of the bZIP family members to interact via either heterodimerization or homodimerization, the results of Y2H showed some differences from the expectation of the wide range of the interactions which were essential in terms of the function and control of the biological process. Consequently, from the result of yeast two hybrid, it is difficult to confirm the interaction between two proteins separately out of the biological system, which is controlled by many factors inside cells at specific times (de Folter *et al.,* 2005).



.

Figure 4.14 quantitative β-galactosidase assay for interaction in the yeast two-hybrid assay.

4.4.7 Activation of the *TaLMW, TaSSIII, TaSSII* and *TaSBE* Promoters by TaSPA, TaPBF, TaRSR1 and TaBLZ1.

4.4.7.1 *Identification the motifs in the promoters*

Most the genes especially those having similar expression patterns are co-regulated by the combinational interaction of TFs (Chang *et al.*, 2008), and these genes with similar expression patterns may share biological functions. Hence, from the DNA sequence of the co-regulated genes could predict the network of TFs, either as activators or repressors, especially if it is possible to combine the knowledge of sequences with co-expression analysis (Pilpel *et al.,* 2001). The full understanding of the TFs functionality offers an insight into the biological roles of the query gene.

The methods used to identify cis-regulatory elements in the promoter can be divided into two classes, namely enumerative and alignment methods (Ohler and Niemann, 2001). In this chapter, several motifs were identified in the promoters of different genes to be positively and negatively co-regulated using the Regulatory Sequence Analysis Tool (Plant PAN2.0; (http://PlantPAN2.itps.ncku.edu.tw) (Chow *et al.,* 2015). All these genes have been confirmed their expression patterns as being in the endosperm of the grain during the filling stage. There are numerous binding sites for TFs such as TaPBF, TaSPA, TaBLZ1, and TaRSR1 which belong to the DOF, bZIP and euAP2 transcription factor families, respectively. For example, namely the AAAG (P box) and ACGT (O2 box) core elements were found in all examined promoters **(Fig. 4.15)**. Interestingly, the core elements of the P and O2 boxes were side by side in these promoters.

Furthermore, other binding sites of TFs were identified such as the MADS-box, MIKC, which can recognize (CCATAT), and other members of AP2 of B3 and RAV subfamilies which recognize (CAACA). Also, (AAATA) can be recognized by Myb (Myeloblastosis)-related TFs. Overall, from the promoter analysis, the specificity for TFs is consistent with available data on the binding specificity of the corresponding families, although some remarkable differences were found between these motifs**.**

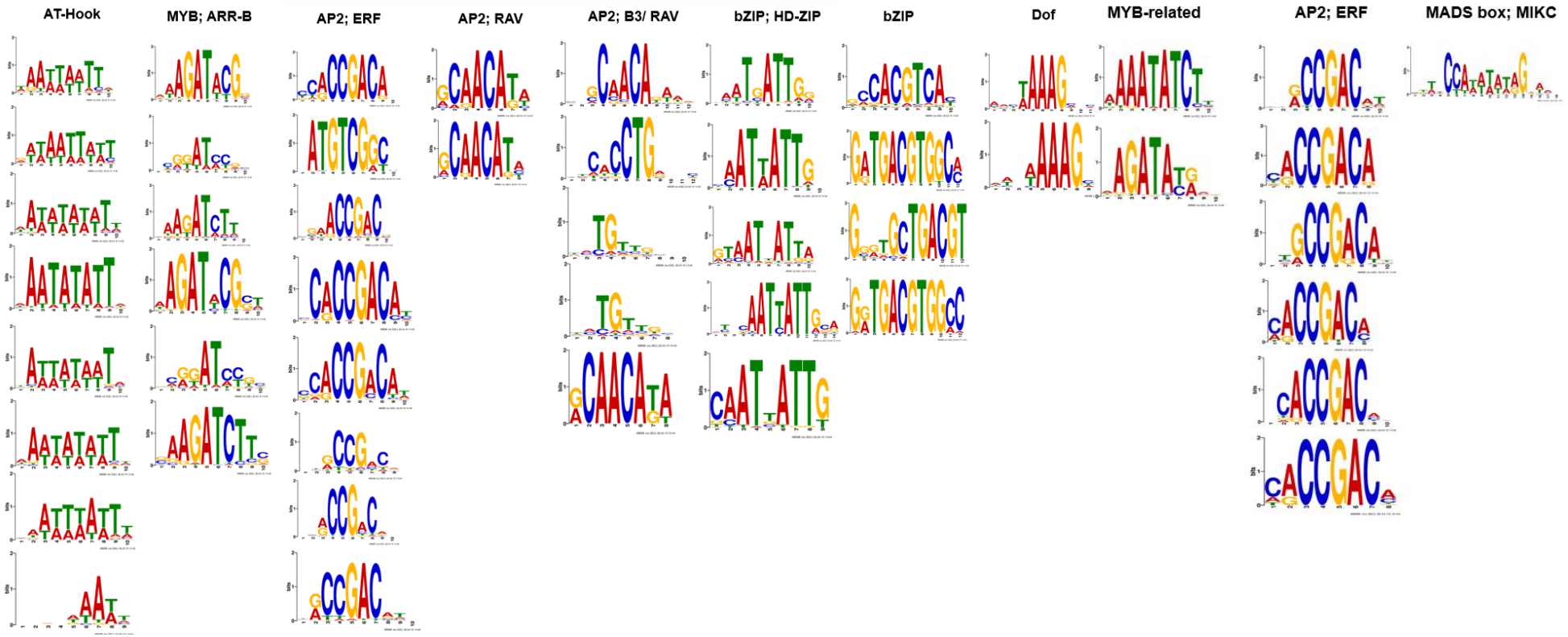


Figure 4.15 The common DNA-binding specificity of some TFs which have been detected in the promoters of our target genes (TaLMW, TaSSIII, TaSSII, and TaSBE. (Source; Plant PAN2.0; (http://PlantPAN2.itps.ncku.edu.tw).

4.4.7.2 *Transient expression assay in the wheat endosperm*

To investigate whether starch-synthetic genes can be directly activated by candidates TFs such as TaSPA (bZIP family), TaPBF (DOF family), or TaRSR1 (euAP2 family), the dual luciferase transcriptional activity assay was used for this purpose. In this system, the reporter construct is generated that consists of two luciferase cassettes, one being the Renilla LUC (REN) internal reporter gene driven by the 35S promoter (35S-REN) that is used as an internal control, and the other being the firefly luciferase (FLUC) driven by the target gene promoter of interest cloned upstream of the FLuc CDS **(Figure 4.16).** In this assay, the starch synthetic gene promoters (TaSSIII, TaSSII, and TaSBE) have been chosen depending on a previous study which confirms the high expression level during the filling stage, also TaLMW as well. To design the constructs of promoters (around ~1 Kb) were cloned **(Table 4.6)** and were fused with FLUC, while the effectors (target of the TFs) were fused with the 35S promoter **(Figure 4.17).** Both these constructs were transiently transformed into tobacco leaf epidermal cells with and without a transcription factor (effector) by agro-infiltration. If the effectors have the ability to activate the promoters of interest by binding specific motifs, the level of FLuc protein will increase. Then, the activity of FLuc can be quantified by using chem-iluminescent substrate D-luciferin and the difference in FLuc activity between reporter alone and reporter and effector is calculated to an indication of the strength of transactivation.

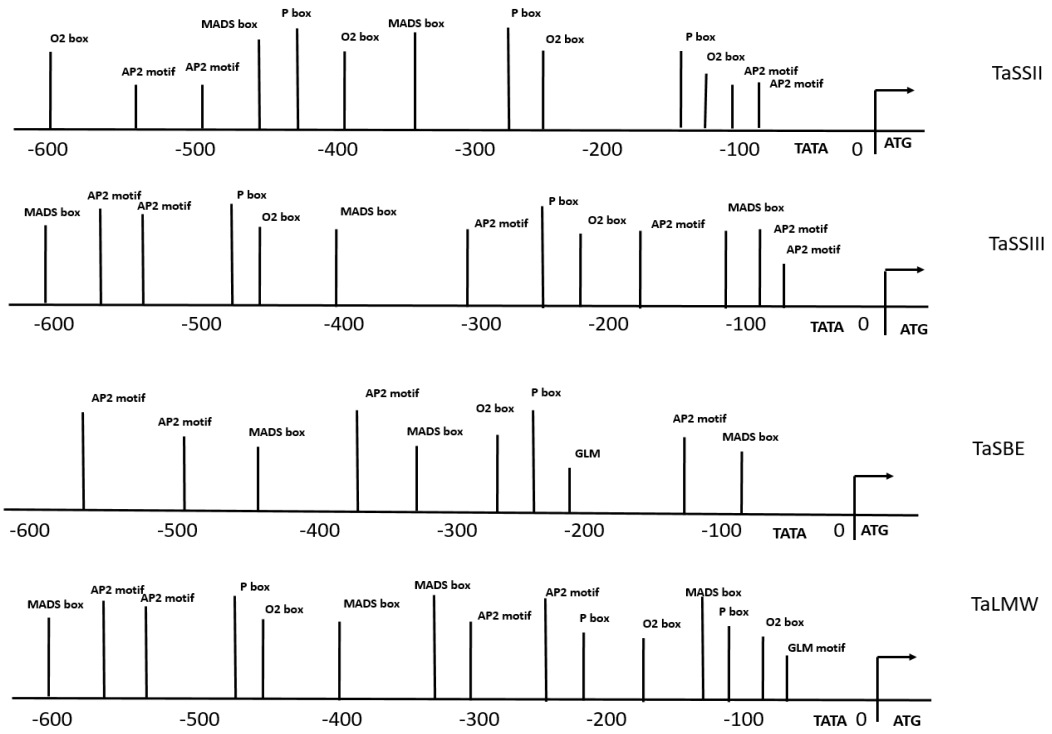


Figure 4.16 Schematic digram of the pormoter regions.

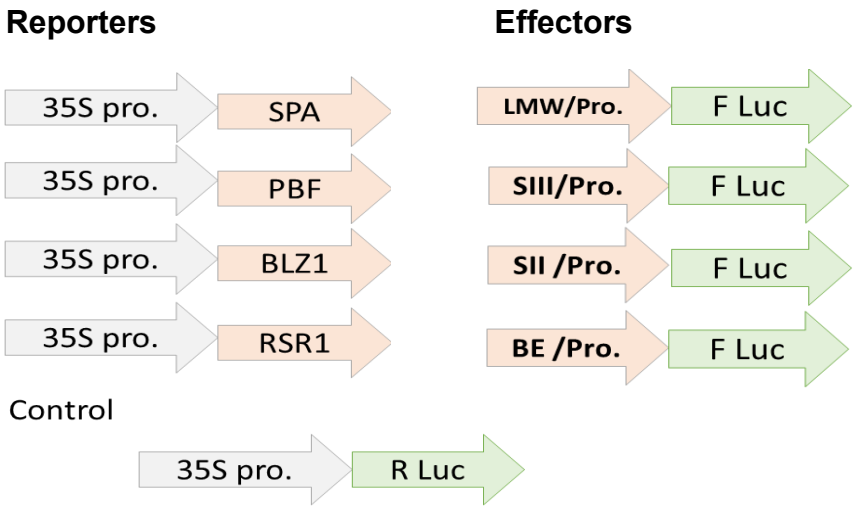


Figure 4.17 Schematic of all constructs which were used for Transient assay.

The objective of our investigation was to test whether these TFs can be bind directly the promoters of these genes, in fact, the differential gene activities were observed between them. Comparing the starch synthesis genes (*TaBE, TaSSII* and *TaSIII*), the differences between them may be noted in terms of their respective activation levels. *TaSBE* and *TaSSII* were negatively regulated by all the TFs tested, such as TaSPA and TaRSR1, while the activity of *TaBE* increased by combining to TaSPA and TaRSR1, which confirmed the ability to control the activity of the *TaBE* promoter. Furthermore, Ta BPF and TaBLZ1 did not show any activation. Although these TFs failed to activate the *TaBE* promoter, there was some variation in the associated levels of activation, which was lowest with TaRSR1 and highest with TaSPA and TaBLZ1, respectively. *TaSIII* showed high levels of activity with TaSPA, while the activity was low with TaRSR1. Generally, TaSPA could regulate the expression of starch synthesis genes a by binding specific motifs, while TaRSR1 negatively regulated both starch synthesis genes. On the other hand, one of the storage protein genes, *TaLMW*, was tested in terms of the ability of different TFs to activate its promoter. TaSPA showed positive activation compared with TaPBF and TaBLZ1. However, the combination of (Ta SPA and TaBLZ1) showed higher levels of activation compared with (TaSPA and TaPBF). These findings further suggest that TaSPA is a direct transcriptional activator for certain starch synthesis and storage protein genes **(Fig.4.18).** In the table (4.6) shows the conclude result of this experiment.

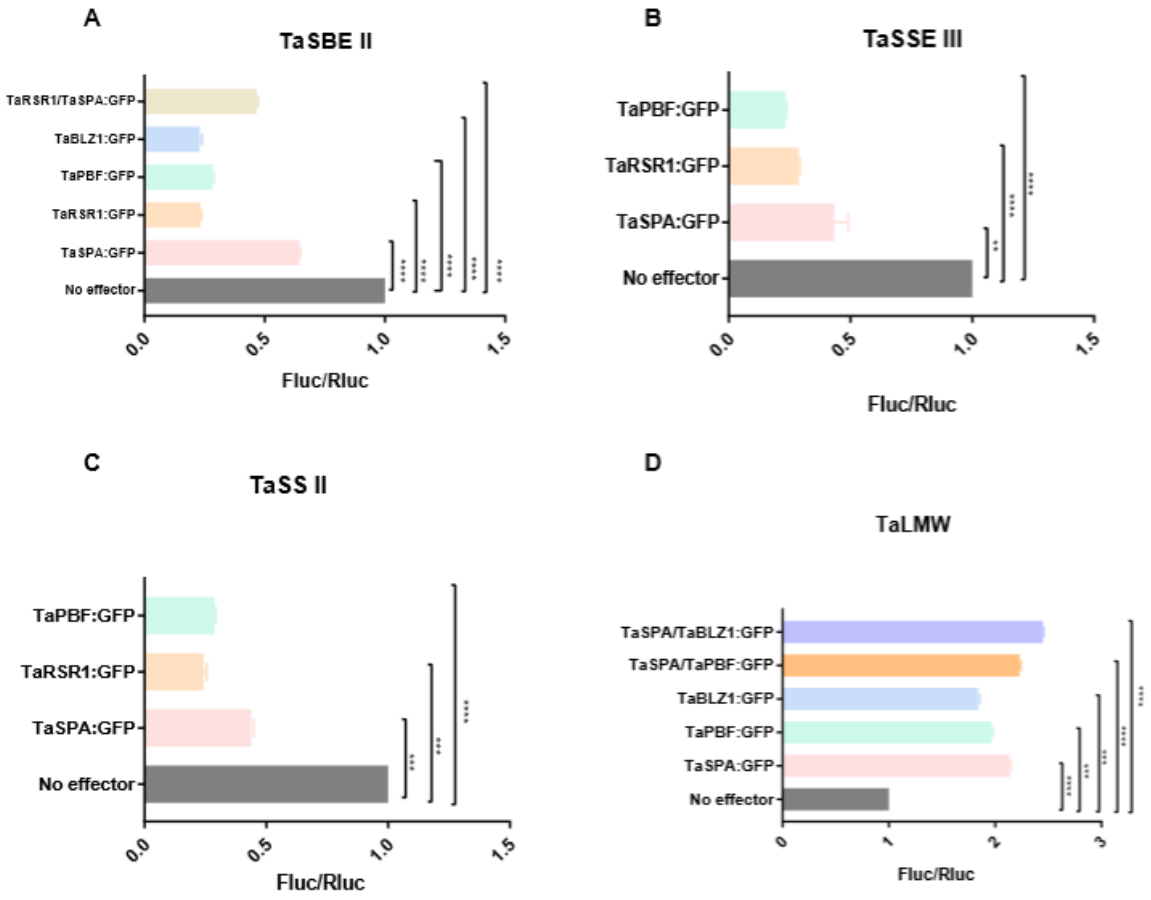


Figure 4.18 Transactivation of thePromoters of A) TaSBEII, B) TaSSIII, C)TaSSII and D) TaLMw promoters by different TFs (SPA, RSR1, PBF, and blz1).

The charts represent the fold-increase of activation compared with the negative control (empty effector). Data represent means ± SD of three independent replicates.

Table 4.6 Candidate target genes tested for promoter transactivation by different TFs in the transient luciferase assay.

Name	ID genes	TaSPA induced	TaBLZ1 induced	TaPBF induced	TaRSR1 induced
Low molecular wheight	*Ta-LMW*	+	+	+	-
Starch synthesis III	*Ta-SIII*	+		+	-
Starch synthesis II	*Ta-SII*	-		-	-
Branching enzyme	*Ta-BE*	-	-	-	-

4.5 Discussion

In this chapter, we carried out a comprehensive functional analysis focusing on TaSPA, and others such as TaRSR1. First, to examine the potential regulatory function of TaSPA, the RNAi method was used to knock down *TaSPA* gene. RNA interference (RNAi) is one of the technologies which has been used to down-regulate multi-genes (Travella *et al.,* 2006). For example, it was reported that the reduction in all x-, a-, and c-gliadins by RNAi in bread wheat lines of cv. 'Bobwhite' which have an effect on the protein content of the endosperm (Gil-Humanes *et al.,* 2010). Another study, Gil-Humanes *et al.,* (2008) reported the reduction of c-gliadins resulted in a decrease in the protein by up to 80%.

In the current study, TaSPA-RNAi lines showed a reduction in the expression of *TaSPA* except in few lines. The level of the reduction *SPA* by RNAi was varied in the lines. In addition, the phenotypic observation of TaSPA-RNAi showed significant differences between the mutant lines and WT (control) in terms of the flowering time, plant length and seed mass. However, it is known that TaSPA is an endosperm-specific TF, which has long been known to regulate SSP genes in the endosperm. So, it has not been previously reported if one of the mutant orthologous of SPA showed the effect on any part of the plant except the storage protein and starch accumulation in the endosperm. Despite all these observations, it is not clear if the knockdown of TaSPA was the main reason, or the knockdown is likely to affect levels of other transcription factors. Also, the grain had a soft and chalky endosperm, comparing to the wild-type, which could be this feature was affected by the accumulation rate of storage protein in the endosperm that impacts on the morphology of protein bodies as well. However, this point needs more investigation to confirm this variation.

Furthermore, the mechanism of regulating storage protein genes (SSPs) in the endosperm is controlled by a network of TFs through interactions between them to enhance the activation of the SSPs promoters. In addition, the prior study that has noted the effect of a triple mutant in maize PbfRNAi; OhpRNAi; o2RNAi was reduced dramatically by both RNA transcript and protein levels of 27-kD g-zein (Zhang *et al.,*

2015). The possibility here is that the mutant line of o2RNAi and PbfRNAi may influence on the levels of some other transcription factor(s), or PbfRNAi and o2RNAi are otherwise able to directly or indirectly cause defects in zein gene (Zhang *et al.,* 2015). This study reported the level of the associated reduction in zein proteins in *o2RNAi*, and which were also reduced markedly in the double mutant PbfRNAi; *o2RNAi*. Furthermore, it was observed that there was a reduction in the kernel weight by 20% and 43% in *o2RNAi* line and *PbfRNAi; o2RNAi* line, respectively. However, the RNA-seq transcriptome analysis referred to that around 1000 genes were down-regulated or up-regulated in *PbfRNAi* and *o2RNAi* mutants.

Summaries, further work is required to confirm the effects of the mutant on different aspects of the grain such as total protein content and the transcript level of storage protein genes in the endosperm. Additionally, it is believed that the specificities of the temporal and spatial expression of storage protein genes and starch synthesis genes in the endosperm are controlled by several endosperm-specific TFs (Vicente-Carbajosa *et al.,* 1997) which interact to gather to activate them. This interaction between the TFs is considered of importance for the understanding the molecular biology (Pawson and Nash, 2003). Several methods have been proposed for studying protein-protein interactions such as yeast two- hybrid screening has been used to successfully identify and/or confirm putative protein interactions in vitro (Fields and Song, 1989). Several studies identified some of the TFs that participate in the regulation of coordinate genes in grain, particularly TFs of bZIP which are transcriptional activators of the seed storage protein by binding DNA motifs in the promoter through dimer formation (Vinson and Boyd, 1993; Foster *et al.,* 1994; Diaz and Carbonero, 1998). In particular, (TaSPA, TaBLZ1); (C group) bZIP family have been confirmed as being able to undergo hetero- and homo-dimerization, although these dimers showed different patterns and amounts between different proteins.

In this project, it is considered TaSPA has the greatest impact on the expression of storage protein genes and starch synthesis genes as the transcriptional activator, and others could have additive and synergistic effects on their expression. So that, we assessed the interaction between TaSPA and TaBLZ1, which showed strong

activation between them especially between the fragments of TaSPA (bZIP domain) and TaBLZ1. This result is similar to the previous study in Arabidopsis (Ehlert *et al.*, 2006). The C group of bZIP showed strong activation between members of bZIP.

In 1999, On˜ate *et al.,* demonstrated that the HvBLZ2 interact strongly with HvBLZ1, suggesting the majority of bZIP proteins could form dimers and the bZIP domain is the most important to support the dimerization sufficient to sustain the dimerization between HvBLZ2 and HvBLZ1. This result seems to be consistent with other studies which found that the interaction between the members of bZIP (C group) showed strong activation between them compared with other groups of bZIP. The reason for this is the structure of the leucine zipper, which has a significant impact on the dimerization, especially the number of heptad (abcdefg) repeats, it was the longest in the bZIP superfamily, also the structural features could play a role in controlling the dimerization (Deppmann *et al.,*2004; Ehlert *et al.*, 2006). In addition, a recent study showed ZmO2 (OPAQUE 2) involves in a complex of regulators to regulate the zein protein genes by interacting with ZmMADS47 (Qiao *et al.,* 2016). Another important finding was that ZmO2 directly regulates zein genes by binding the specific motif of the promoter of these genes; also, ZmMADS47 specifically binds a CATGT motif in the promoter of α-zein genes. This finding suggested that ZmMADS47 binds to zein promoters next to O2 protein. However, ZmMADS47 alone cannot transactivate the zein genes, but both O2 and ZmMADS47 can able to induce the transactivation of these promoters, which in this instance was highly enhanced. The level of enhancement varies, depending on the AD function of ZmMADS47 and the interaction between ZmMADS47 and O2 (Qiao *et al.,* 2016).

According to all these previous studies, I tried to confirm the interaction between (TaSPA, TaBLZ1) bZIP members and TaMADS47. Hence, the results of this current study showed that TaMADS47 could interact with TaSPA and TaBLZ1, although the interactions were varied. These interactions between TaMADS47 and TaSPA or TaBLZ1 were very weak, while the interaction with full-length Ta BLZ1 showed a strong activation with the reporter gene in yeast. The differences between them indicate several possible reasons for that such as yeast two-hybrid analysis has

certain a disadvantage in terms of the interaction between the proteins not working bi-directionally depending on which of the domains (BD or AD) can fuse to the N-terminus or C-terminus of the protein of interest (Ehlert *et al.*, 2006). Another reason is that the length of the fragment, which has been cloned because of it not being fully understood which protein sequence motif is essential to the interaction with other proteins.

Surprisingly, AD-TaMADS47 was found to interact with BD-TaRSR1. This result further supports several studies which indicated the role of TaRSR1 and TaMADS47 activating storage protein genes or starch synthesis genes in the endosperm (Fu and Xue, 2010; Qiao *et al.,* 2016). Overall, the interactions which were observed between the proteins especially those showed weak interactions, can be confirmed by different systems such as P2H (protoplast two-hybrid P2 system), which identified weak hetero-dimerization events which cannot be detected in the yeast system (Ehlert *et al.*, 2006). Together, these results provided important insights into the importance of the dimerization in gene regulation. Also, our results support the existence of a positive relationship between (TaSPA, TaBLZ1) bZIP family and other TFs which play role in the development of grain such as TaPBF, TaRSR1 and TaMADS47, and which are all necessary for the high activation levels of storage protein genes and starch synthesis genes. So that, our results support the recent studies which posited that both the storage protein genes and starch synthesis genes can be regulated with common transcription factors in the endosperm. In addition, the interactions between proteins were detected in the yeast two-hybrid assay can only be biologically related when the two proteins are existing in the same cells at the identical time (de Folter *et al.,* 2005). However, this point needs further studies using different experimental methods such as specific mutations in motifs which play a role in the dimerization, or deletion of a specific motif which will give further insights into how dimerization specificity is achieved.

Furthermore, several attempts have been made to analyze the impact these interactions between the proteins on the activation of their targets. It was investigated by prediction of potential transcription factor binding sites (TFBS) of the promoters of

several genes such as (*TaSSEIII, TaSSEII*, *TaSBE* and *TaLMW*). Here, our result indicated that the target sequence specificity of plant TFs represented in the promoters of these genes. The cis-regulatory elements of the prolamin gene were first identified in barley, and then in different cereals such as wheat, rye and maize (Forde *et al.,* 1985; Ueda *et al.,* 1994). It was known as an endosperm box which has two highly conserved motifs, namely the prolamin box and a GCN4-like sequence which is bound by basic Leucine zipper transcription factors (Schmidt *et al.,* 1992). The second motif is bound by a Dof-type transcription factor termed the prolamin box-binding factor (PBF) (Vicente-Carbajosa *et al.,* 1997). Our analysis of certain promoters was correlated with previous studies which confirmed the specificity of binding sites for these TFs (Franco-Zorrilla *et al.*, 2014; Qiao *et al.,* 2016).

To confirm this analysis, the dual luciferase transient assay was used for detecting the promoter activity of certain starch synthesis genes and low molecular weight (LMW). The reason for choosing these genes was because they are likely to be co-regulated by common TFs where it is known that these genes share a biological function. More recent attention focused on both starch and protein synthesis in the endosperm are temporally integrated via a network of transcriptional regulatory which controls both synthesis processes (Keeling and Myers, 2010). This integration has uncovered how these genes govern the accumulation of both protein and starch during the development of grain. Thus, we have attempted to investigate the ability of (TaSPA and TaBLZ1) bZIP family to activate the promoters of both *TaSSII* and *TaSSIII*, *TaBE* and *TaLMW* to reveal the main function of TaSPA, TaBLZ1, TaPBF and TaRSR1 were chosen to investigate their ability to activate promoters of our target genes. The function of these TFs can be tested via *in vivo* dual-luciferase transient assay which could provide very useful information, as shown in the results above.

However, although the transient assay helps to detect the activation of a target promoter, it was not necessarily translated into a functional network in the natural case or guarantee a stable expression for the protein itself. To date, several studies have investigated this area, for example, Zhang *et al.,* (2015) reported that the roles

of ZmPBF and ZmO2 have the ability to control 89% of zein transcription and they involved in the regulation starch synthesis genes (Wang and Larkins, 2001; Jia *et al.*, 2013), but their roles are not yet completely understood. However, ZmPBF and ZmO2 have the ability to activate the PPDK1, PPDK2, and SSIII promoters in maize by recognizing the respective P and O2 boxes; however, they failed to activate the *TaSSIIa* and *TaSBEII* promoters. Another explanation is ZmPBF and ZmO2 can directly regulate PPDKs and TaSSIII, while they only indirectly regulate *TaSSIIa* and *TaSBEI* (ibid). A model for the role of ZmPBF and ZmO2 on the regulation of kernel nutritional quality and yield has been presented, as shown in (**Figure 4.19)** (Zhang *et al.,* 2015). However, this finding is consistent with our result that TaSPA cannot regulate Ta*SBEII* and *TaSSII*, comparing to Ta*SSIII* which showed activation. Furthermore, TaBLZ1 failed to activate the *TaSBEII*. TaBLZ1 could regulate all these genes but in an indirect manner by forming a protein complex with other proteins. To date, there is no clear evidence of TaBLZ1 being one of the proteins involved in the complex for regulating the starch synthesis genes.

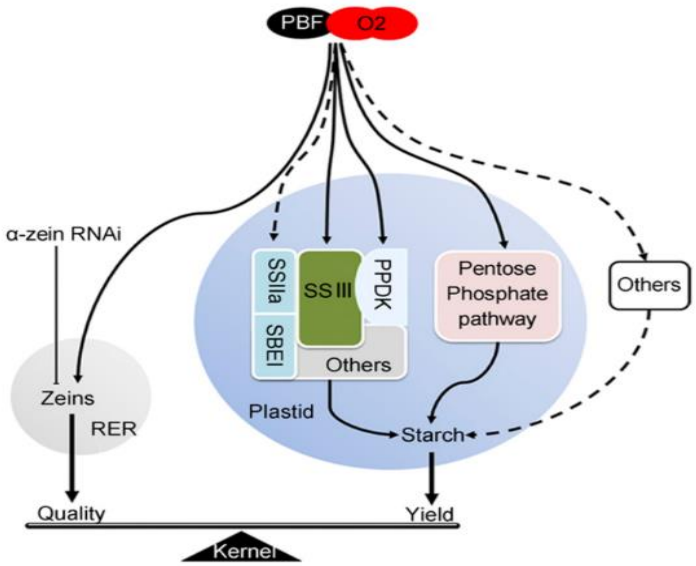


Figure 4.19 A suggested model to explain the role PBF and O2 controlling the quality and yield in maize (From Zhang et al., 2016).

Furthermore, it has been found in other transcription factors (TFs) are involved in starch biosynthesis and its regulation. A small numbers of TFs were identified in

cereal crops such as OsBP-5 (Zhu *et al.,* 2003), FLO2 (She *et al.,* 2010), OsRSR1 (Fu and Xue, 2010), and OsbZIP58 (Wang *et al.,* 2013), which have the ability to regulate the expression of starch biosynthesis genes by binding the promoters of six starch synthesizing genes which are OsAGPL3, Wx, OsSSIIa, SBE1, OsBEIIb, and ISA2. In rice, the null mutant in OsbZIP58 displayed a reduction in the low level of SBE1 (Jeon *et al.,* 2010). RSR1 is known as a regulator of several starch biosynthesis genes (Fu and Xue, 2010). However, Singh *et al.,* (2015) revealed that the quantitative expression of RSR1 is strongly and positively correlated with starch synthesis genes such as SSIII, *ISA1*, *DPE1*, and *SBEIII*, and negatively correlated with *SBEI, GBSSI, SUT*, and *SSII*. However, their results demonstrated that the expression of TaSPA has a strong positive correlation with nine starch synthesis genes, such as SSIII, while it had a strong negative correlation with other genes such as *SBEII* and *BT1*. This result confirmed that both wheat TaRSR1 and TaSPA are regulators of most of the starch genes, which were correlated with other studies which reported the functions of both OsRSR1 and Os bZIP58 in rice (Fu and Xue 2010; Wang *et al.,* 2013).

All these findings broadly support our work in this area, where TaRSR1 was negatively regulated for *TaSIII, TaSII* and *TaBEII*. Taken together, these studies, as shown previously, support the concept that the starch synthesis genes are regulated with a complex of TFs and either some members of bZIP, as mentioned in previous studies, or others such as PBF (DOF family) or RSR1 (euAP2 family) (Kang *et al.,* 2012). In terms of the TaBLZ1 (bZIP family) it is unknown whether it plays a role in the activation the promoter of starch synthesis genes or storage protein genes, but the spatial expression in the endosperm during the development of grain could reflect its functional role during the filling stage.

Taken together, the transient result has led us to the following conclusions. First, TaSPA is a main positive regulator of storage protein genes and some starch synthesis genes. Second, TaRSR1 is generally a negative regulator for most of the starch synthesis genes but has an unknown its role in the regulation of protein genes. However, the interaction between TaSPA and TaRSR1 could reflect the common

function of both TaSPA and TaRSR1. Third, TaSPA and TaBLZ1 can increase the activation of the storage protein gene (TaLMW) to a greater extent tShan TaSPA and TaPBF, it is possible that the TaSPA (storage protein activator) is a master transcriptional activator, but (TaPBF) and (TaBLZ1) could increase transcript of the promoter of wheat genes by binding with SPA. Generally, this indicates that the variation between the grains in terms of the accumulation of protein and starch synthesis genes might be related to the complex of protein regulators, and until this study, it remains unclear as to what and how these TFs are involved in the regulation of these genes.

Chapter 5

The morphological characterization of starch granules and protein bodies in the Endosperm

Abstract

Background: It is known that grain quality and the kernel yield are affected by the structure of the endosperm cells which contain the storage protein and the starch. Both the starch and storage protein are reserved in different regions of the grain. The aim of this chapter is to compare between the hexaploid wheat, tetraploid, barley and rosner them in terms of the morphological features, size of the granules and measurement the total starch in full-length and mature stages.

The result: I used SEM and simple staining to examine the general features of grains of different. The observations showed that the general morphological structure that was similar between examined grains in different species were noted by using SEM. However, preliminary scanning electron microscopy (SEM) observations were extended in this study because protein bodies and starch granules were difficult to distinguish them in SEM, simple staining such as with iodine solution, and Coomassie brilliant blue were combined with SEM to assess the variation in the morphological characterization of starch granules of the endosperm cell and protein. In addition, the total starch and amylose content were analyzed which reflect a significant difference between different grains in terms of the size of the cells (starch, protein), and their distribution.

Conclusion: This study was very useful for investigating of cereal seed development and utilization because of the morphological and chemical differences within the varieties of the grain have a significant impact on end use.

Chapter 5 The morphological characterization of starch granules and protein bodies in the Endosperm.

5.1 Introduction

Cereal grain is important as a renewable resource for human food, and industrial raw material. The major tissue in the grain is endosperm, which holds the plant's nutritional reserves (Liu *et al.,* 2009). However, both starch and protein are accumulated separately in specific granules in the endosperm cells. The variations between granules of the starch and proteins in terms of the size, shape, and distribution within the endosperm cells can have a significant impact on the weight, quality of grain (Andersson *et al.,* 1999; Jing *et al.,* 2014). Furthermore, morphological features from different species can be sourced to understand the relation between the structure of these granules and their function. In addition, several studies have reported the influence of granule sizes on functional baking features (Chiotelli & Le Meste, 2002; Shinde *et al.,* 2003; Park *et al.,* 2004; Liu *et al*., 2007), and pasting behaviour of the starch (Ao & Jane, 2007). Although several studies have observed wheat starch granules in terms of the morphological changes, they have not been comprehensively characterized (Tang *et al.,* 2006; Kim and Huber, 2008; Salman *et al.*, 2009). As a result, we attempted to shed more light on the morphological changes of starch granules during grain filling and seed maturation, also, investigate in the morphology of starch (A,B) granules, and the distribution of storage protein in mature cereal seeds which is of such vital importance to the quality of grain.

5.2 Materials and methods

5.2.1 Scanning electron microscopy (SEM) analysis for mature grain

Dry mature grains were freshly prepared FAA fixative (formaldehyde 3.7%, acetic acid 5%, and ethanol 50%). Then, tissues were exposed to three cycles of moderate vacuum (~500 mbar) to ensure penetration of fixative and fixed overnight at 4°C with agitation. Fixed tissues were transferred to 70% ethanol. Samples were dehydrated through a series of 80%, 90%, and 100% ethanol with 12–24 h in each. Samples were coated in gold using a Polaron SC7640 Sputter Coater for 90 s at ~2.0kV. Samples were analyzed on a Hitachi S3000H scanning electron microscope equipped with digital image capture.

5.2.2 Extraction of starch granules

For studying the starch physicochemical properties, starch was extracted from mature grains. The starch extraction method was modified from South and Morrison, (1990). Ten grains were steeped overnight in 5 ml of water prior to thorough grinding using a pestle and mortar in 20 ml of water. The resulting suspension was filtered through a 100 µm sieve and the filtrate centrifuged for 20 min at 4000 rpm. The pellet was resuspended in 1 ml of water and the suspension layered above 9 ml of 80% (w/v) CsCl in a 15 ml centrifuge tube. This was centrifuged for 15 min at 4000 rpm and the supernatant discarded. The pellet was re-suspended in 1 ml of water and transferred to a 1.5 ml micro-centrifuge tube prior to centrifugation at room temperature for 5 min at 13 000 rpm. The pellet was washed in this way a total of three times with water and then once with ice-cold acetone. Pellets were air-dried and stored at –20 °C.

5.2.3 Measurement of the total starch and Amylose

Flour samples preparation around 36 dry mature grains from different species were selected and grinded, around 100 mg of flour for each samples two duplicates for each test sample with three replicates. For the total starch using a commercial kit which is the Megazyme total starch assay procedure (Megazyme Total Starch Assay Kit). (Amylose/ amylopectin assay procedure kit) was used for the measurement of the amylose and amylopectin contents, according to the manufacturer's recommendation.

5.2.4 Staining and observation of starch and protein granules

The sections of grains were stained with different stains to exhibit the morphology of endosperm cell and starch and the distribution of storage protein. For observation of starch, the sections were stained with iodine solution (25% glycerol, 0.14% KI) or periodic acid-Schiff's reagent. For observation of storage protein, sections were stained with Coomassie brilliant blue R250, or toluidine blue O. The grain stained sections was performed using a GX optical L3200 compound microscope equipped with a GT-vision GXCAM-5 5MP digital USB camera and GXCAPTURE software.

5.2.5 Quantitative analysis of morphology parameters of starch granules and protein bodies

The area and size (long and short axis length) of both starch and protein granules of the whole sections were determined using ImageJ software. Every measurement was repeated 3 times.

5.2.6 Data statistical analysis

Statistical analysis was performed using one-way ANOVA followed by Tukey t-tests with the GraphPad InStat version 7.0 statistical software program (GraphPad Software, San Diego, CA).

5.3 Results and discussion

5.3.1 Scanning electron microscopy (SEM) observation for grain.

Mature grain of different varieties of hexaploid, and tetraploid wheat (*Chinese spring,* and *wheat cadenza*), barley, Rosner and rye were examined by using scanning electron microscopy (SEM) to investigate the general structure of grain and the starch granules in species. The comparison of transverse sections revealed the number of significant similarities and differences in starch granule size, distribution and to assess A- and B- type starch granules characteristics. One of the previous findings suggested that the morphology of starch granules has a significant role in grain digestibility and industrial applications (Sabelli and Larkins, 2009). Our Observations on the grains were made on mature kernels elucidate structural features.

5.3.1.1 *General features*

The outline shapes for all species have the rounded and heavily creased profile in most grains except durum and rye, which have narrow profiles, although their similarity in the basic structure, there is a difference between them in terms of the lobed structure. Some grains have obvious lobes to give the closed crease such as wheat, while the barely has starkly flat **(Fig.5.1; to Fig.5.6).**

5.3.1.1.1 *Peripheral aleurone layer*

In different verities of wheat, the peripheral aleurone layer is recognized by cuboidal cells with thick walls, surrounding the central endosperm, which has larger cells, thin-walled cells. However, all the species have distinct cuboidal cells with thick walls, while in the barley peripheral aleurone layer to be significantly less distinct from the central endosperm as compared to wheat. Furthermore, the variation in this layer between the species could be related to the number of layers, in hexaploid species there is one layer, also in rye and durum. However, the barley occurs similarity to the other species which have rounded cells but there are three layers surrounding the starchy endosperm (Jones, 1969).

These cells have a function during endosperm development as a site of nutrient transfer, and, during grain germination, as a site of nutrient storage and hydrolase production (Richie *et al.,* 2000). Comparative differences between central

endosperm and aleurone cell size were also noticed. Wheat aleurone cells are usually around 40 µm in diameter, approximately a third of the size of central endosperm cells **(Fig. 5.1)**. Comparing between the species, it can be seen the rye the size of aleurone cells are slightly smaller and less distinctive than those in other species **(Fig.5.6)**.

5.3.1.1.2 ***Protein matrix***

It is mainly proteins and residual cytoplasmic constituents between the starch granules. Also, the protein matrix was observed even between the peripheral aleurone layers, which has been confirmed that this layer has more protein matrix than the starch granules in the endosperm. From the transverse sections, there are the boundaries and cell walls, which divide the endosperm tissue into several areas. However, the protein matrix extends to form arrays of prismatic cells which have radially orientation, but some these arrays are big and form continuous sheets as in the wheat, barley and Rosner, while in ray it is small and forms a separate sheet. Consequently, from this structure it can be said that the protein content in ray is less than other species.

5.3.1.1.3 ***Endosperm cells***

It is considered the largest tissue in the entire grain and composes about 87 % of the total grain dry weight (Finnie & Svensson, 2009). It contains starch granules and proteins, which shows distinct differences between the species of Poaceae in terms of the content, the endosperm cell size and wall thickness. For example, in wheat the starch granules are packed tightly and have different size and shape (Tomlinson and Denyer, 2003).

In terms of the shapes of starch granules, there are many different shapes such as spherical, oval, polygonal, and elongated and kidney shapes. However, from our observation can be noticed that wheat, triticale, barley and rye starches have a bimodal shape. The large starch granule which is known (A) has a disk shape with the diameter of >10 µm, with ‘‘equato-rial groove’’ at the surfaces. While the small granule has a spherical shape, and its size less than 10 µm in the diameter which is known (B) (Kim and Huber 2008; Wang *et al.* 2014). Furthermore, there is

variation in the size of the starch granules in the different region such as the starch granules have small size in the area of aluerone layer and the size increase gradually in the center of the endosperm.

However, comparing the A granules in different species, it was found that the largest sizes in rye comparing to wheat and triticale. So that, the morphological features of starch such as the shape and size of the granules have the significant impact on the starch functionality and baking properties (Park *et al.,* 2004). In addition, it has been reported another difference between A and B granules in terms of the biosynthesis time during the filling of grain. The synthesis of A granule started 4 DAA, and it still develops over the next 20 days. While, B granule synthesis occurs 10 DAA, with the significant growth around 20 DAA (days after anthesis) (Bechtel *et al.,* 1990; Shinde *et al.,* 2003). This variation in the synthesis time could affect the size of the granules and the molecular organization of the amylose and amylopectin fractions (Tester, 1997). Moreover, the number of the A-granules is always smaller than that of the B-granules, but the A-granules represent the major mass of the starch (Ao and Jane, 2007).

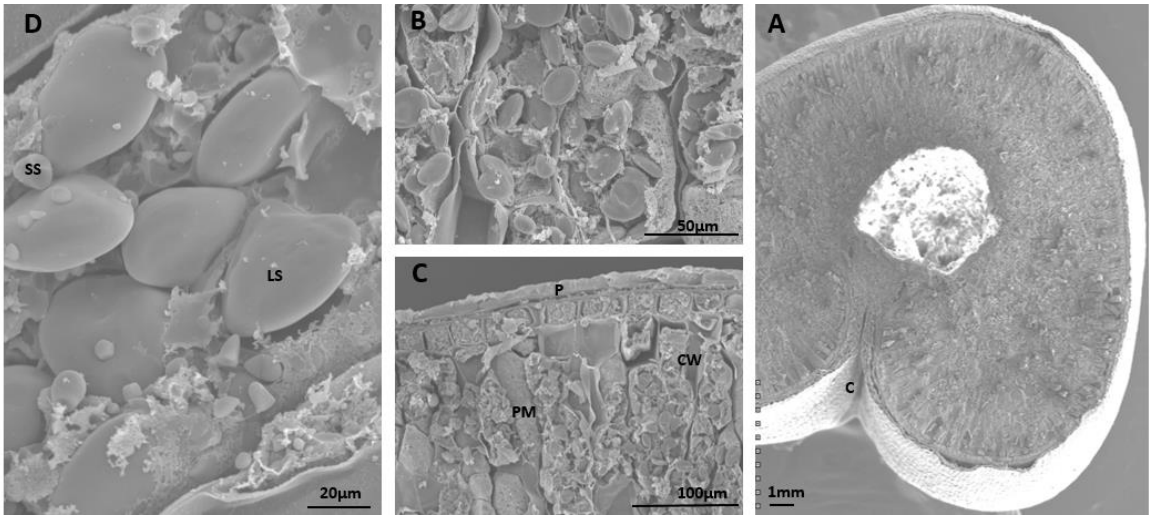


Figure 5.1 SEM for the transverse section of the *Chinese spring* grain.

A. A general view showing the pericarp and aleurone, subaleurone and prismatic endosperm cells. **B.** Cell walls (CW); Protein matrix (PM); **C.** Aleurone cells (Al) and the outermost peripheral endosperm cells (subaleurone layers). **D.** Large and small starch granules as well as protein matrix material in prismatic endosperm.

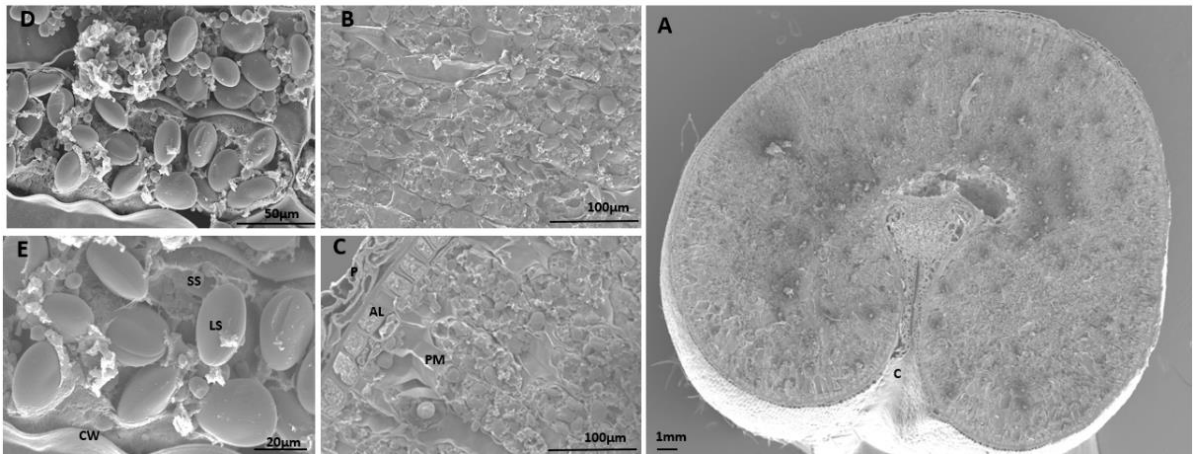


Figure 5.2 SEM for the transverse section of Cadenza grain.

A. A general view showing the pericarp and aleurone, sub-aleurone and prismatic endosperm cells; **B.** Protein matrix and starch granules; **C.** Aleurone cells (Al) and the outermost peripheral endosperm cells (sub-aleurone layer; **D.** Cell walls (CW) of a prismatic endosperm cell; **E.** A higher magnification of large and small starch granules (LS and SS) and protein matrix (PM).

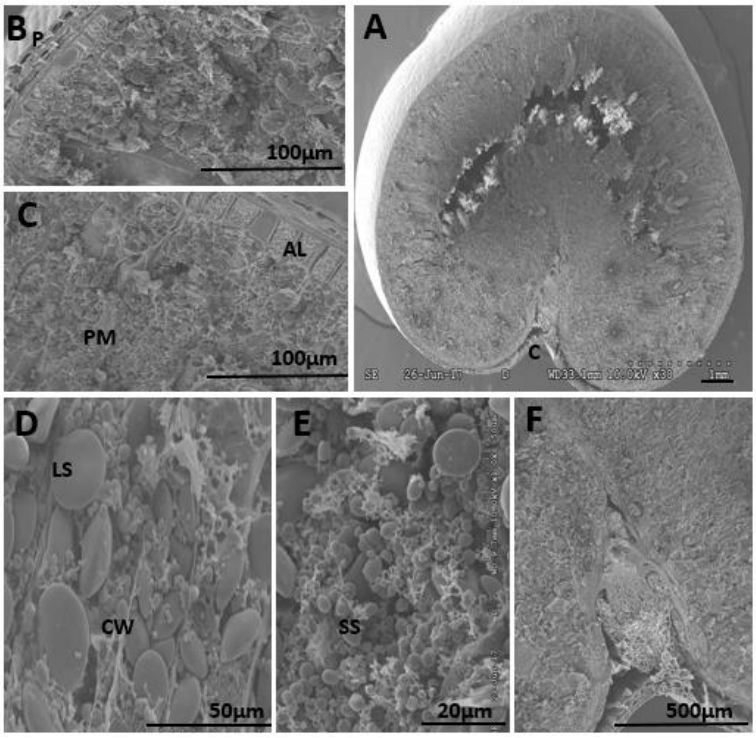


Figure 5.3 SEM for the transverse section of the Durum grain.

A. A general view showing the pericarp and aleurone, subaleurone and prismatic endosperm cells; **B.** Aleurone cells (Al) and the outermost peripheral endosperm cells (subaleurone layer); **C.** Large and small starch granules as well as protein matrix material in prismatic endosperm; **D.** Large and small starch granules, well differentiated cell walls (CW) of a prismatic endosperm cell as well as protein matrix material in prismatic endosperm; **E.** Small starch granules**; F.** A higher magnification of crease region.

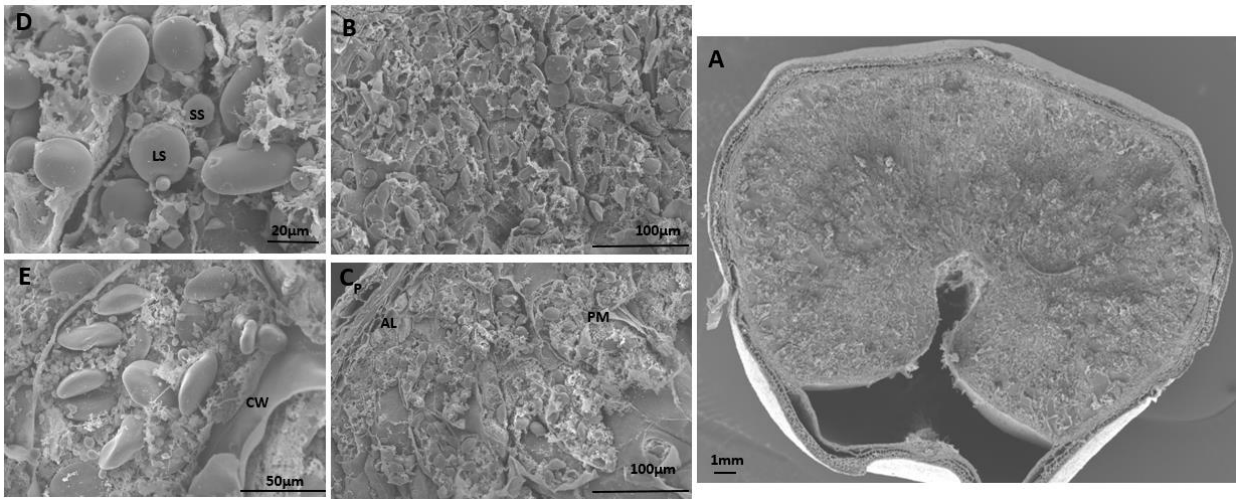


Figure 5.4 SEM for the transverse sction of the barley grain.

A. A general view showing the pericarp and aleurone, subaleurone and prismatic endosperm cells**; B.** Starch granules with protein matriex; **C.** Aleurone cells (Al) and the outermost peripheral endosperm cells (subaleurone layer; **D.** Large and small starch granules as well as protein matrix material in prismatic endosperm; **E.** Well differentiated cell walls (CW) of a prismatic endosperm cell.

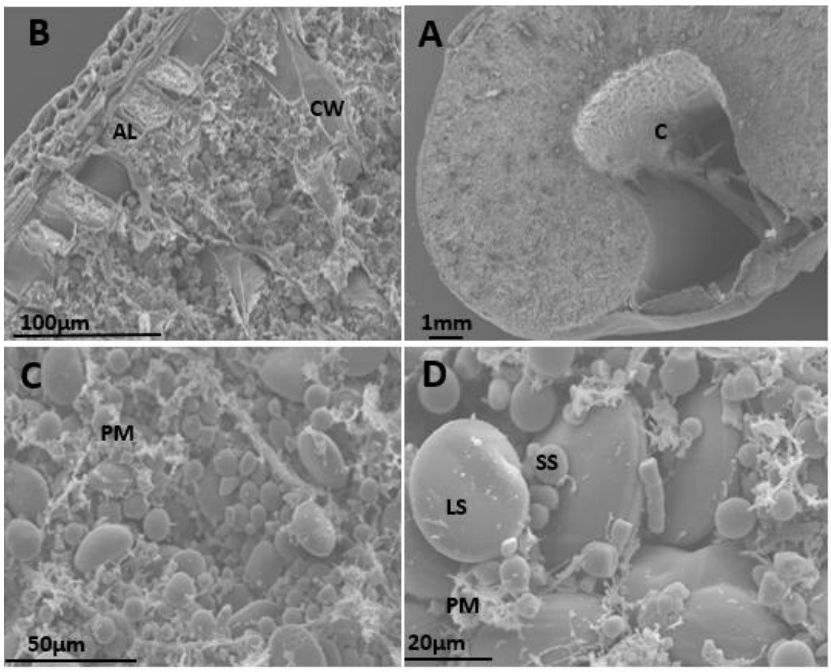


Figure 5.5 SEM for the transverse section of the Rosner grain.

A. A general view showing the pericarp and aleurone, subaleurone and prismatic endosperm cells; **B.** Aleurone cells (Al) and the outermost peripheral endosperm cells (subaleurone layer); **C.** Large and small starch granules as well as protein matrix material in prismatic endosperm; **D.** A higher magnification of large and small starch granules (LS and SS) and protein matrix.

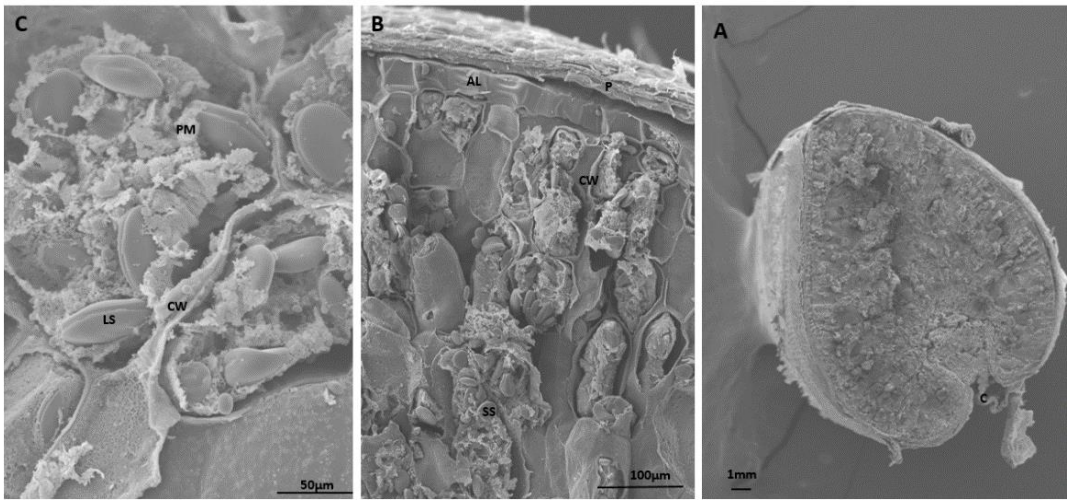


Figure 5.6 SEM for the transverse section of the Ray grain.

A. A general view showing the pericarp and aleurone, subaleurone and prismatic endosperm cells; **B.** Aleurone cells (Al) and the outermost peripheral endosperm cells (subaleurone layer; **C.** Large and small starch granules as well as protein matrix material in prismatic endosperm.

Table 5.1 Summary of observed characters in mature grain.

Species	Grain profile	Crease	The lobe	Cavity	The type of starch	Starch granule shape	Aleurone layer	
							No.	shape
Wheat	Round	Closed	profound	√	A	lenticular	2	Prismatic cells
					B	spherical		
Durum	Narrow	Closed	Slightly profound	-	A	lenticular	1	Prismatic and regular cell
					B	spherical		
Barley	Round	Closed	Slightly profound	√	A	lenticular	3	Compressed and small cells
					B	spherical		
Rosner	Round	Closed	profound	√	A	lenticular	1	Prismatic and regular cells
					B	Spherical		
Ray	Narrow	Closed	Slightly profound	-	A	lenticular	1	Prismatic and regular cells
					B	spherical		

5.3.2 Total starch and amylose measurements

The starch was extracted from dry mature grains of the same species. The procedure for the determination of amylose in cereal grain based on a complex formation between the lectin concanavalin A (Con A) and amylopectin.

Our result shows significant differences between the samples in terms of the total starch and the amylose content. However, the hexaploid cultivated wheat has the highest amount of the total starch comparing to the different varieties of wheat or other species. Furthermore, the total starch hexaploid cadenza is slightly higher than the *Chinese spring* (69.4249%). By comparing the tetraploid wheat (Durum) and other species which are Barley, Rosner and Ray, the total starch was higher in Durum. However, the lowest content of starch and amylose were found in the ray, while durum has the highest content of amylose **(Table5.2).**

Table 5.2 The total starch values were determined in different samples of dry mature grains as percentage on a dry weight of grains, also the ratio of amylose amylopectin

Species	Total starch %	Percentage of amylose (%w/w)	Percentage of amylopectin (% w/w)
***Chinese spring* (*Triticum aestivum*)**	**64.3132**	**11.49**	**88.49**
Cadenza (*Triticum aestivum cadenza*)	**69.4249**	**11.50**	**88.5**
Durum *(Triticum durum)*	**60. 103**	**14.112**	**85.89**
Barley (*Hordeum vulagare*)	**40.5183**	**12.900**	**87.1**
Rosner (*Triticale sp.*)	**41.8999**	**10.555**	**89.45**
Rye (*Secale cereale*)	**19.7603**	**4.311**	**95.7**

It was reported that grains which has high amylose content are more resistant to gelatinisation comparing to the grain with moderate or high amylopectin grains. Also, starch with high in amylose swells more slowly than the starch rich in amylopectin (Colonna and Mercier, 1985).

These results corroborate the findings of a great deal of the previous work in starch in different cereal grains which confirm the barley and maize have high amylose content. These results reflect the importance of the amylose because of that the correlation between the high amylose with the increase of the resistant starch level (Sievert and Pomeranz, 1989; Morell *et al.*, 2003). Also, it is known the amylose is considered as the main factor to reduce the starch digestibility and slowing glucose release (Lehmann and Robin, 2007).

5.3.3 **Observations and analysis of starch and protein granules**

There are several studies which have described the methods used to investigate endosperm structure in different cereal grains such as wheat and barley (Andersson *et al.,* 1999; Holopainen *et al.,* 2005; Dornez *et al.,* 2011; Jääskeläinen *et al.,* 2013). However, in this section, I faced many difficulties in preparing our sections, especially for the entire grain, such that there is no way to obtain the full section of dry mature grain, which is full of starch and storage protein in the

endosperm cells. Another difficulty was how to prepare the entire section of the dry mature grain using microtome (Bright 5030; Bright instrument co., England). For this reason, the images were not very clear because of the thickness of the section and the tight packing of the starch and protein in the endosperm.

Iodine has the ability to penetrate into starch and form a blue complex with lipid-free amylose and amylopectin (Cai *et al.,* 2014). As a result, starch can be specifically stained with periodic acid and Schiff's reagent in cereal endosperm (Faltermaier *et al.,* 2015). To measure the starch granules using different methods, the purified starch from each sample was added on a microscope slide and 20 µl Lugol's solution (Sigma Aldrich), and covered the slide with the coverslip. A light microscope (**Leica DM2500**) with different objectives. Our observations of the quantitative analysis of morphology parameters of starch are reported in **(Table 5.3).** It was measured the size of the area for each starch granule, long axis length and short axis length of the granule of starch, comparing between the starch either small granule and or large granule were the smallest and the largest in the different regions in all samples.

Generally, the morphological changes in granules during grain development, especially at the full stage and maturation, were observed in most of these sections to detect the accumulation positions and distribution in the grains, where it was found that starch accumulates in the pericarp and the endosperm cells during the full-length stage, while the accumulation is restricted in the endosperm cells during maturation because of the thickness of the pericarp is decreased, whilst the protein accumulates during the full-length and maturation was clearly in the aleurone layers and the endosperm cells **(Figs.5.7;5.8).**

This finding was also reported by Zhou *et al*. (2009), who detected that the development of pericarp cells undergoes programmed cell death (PCD) and the storage of starch in the pericarp is only temporary. As a result, the lifecycle of the storage starch in the pericarp can be divided into three phases: (1) biosynthesis in the pericarp cells starts before 6 DPA; (2) both biosynthesis and amylolysis occur at the same time, but biosynthesis is predominant; (3) at the end of the lifetime of

the starch in the pericarp biosynthesis stops. However, it was found that the starch in the thick pericarp varieties is not completely degraded (Earp, McDonough and Rooney, 2004), while in the starch in the thin pericarp varieties was absent, the reasons for which are currently unknown (Li *et al.*, 2012).

This finding broadly supports the work of other studies in this field for different species such as sorghum (Earp, McDonough and Rooney, 2004) and triticale (Li *et al.*, 2012). Also, the number of starch granules increased rapidly during the development of the starch, especially during the maturation, as shown in (**Fig.5.7**), and that not only the number but the diameter of the A-type starch granules gradually increased. Interestingly, it was found that the surface of the A granules are different for the full-length stage and at maturation. During maturation, the surfaces of the A granules were clearly smooth and adopted a ‘‘double disk structure’’ at the end of this stage **(Fig. 5.8).**

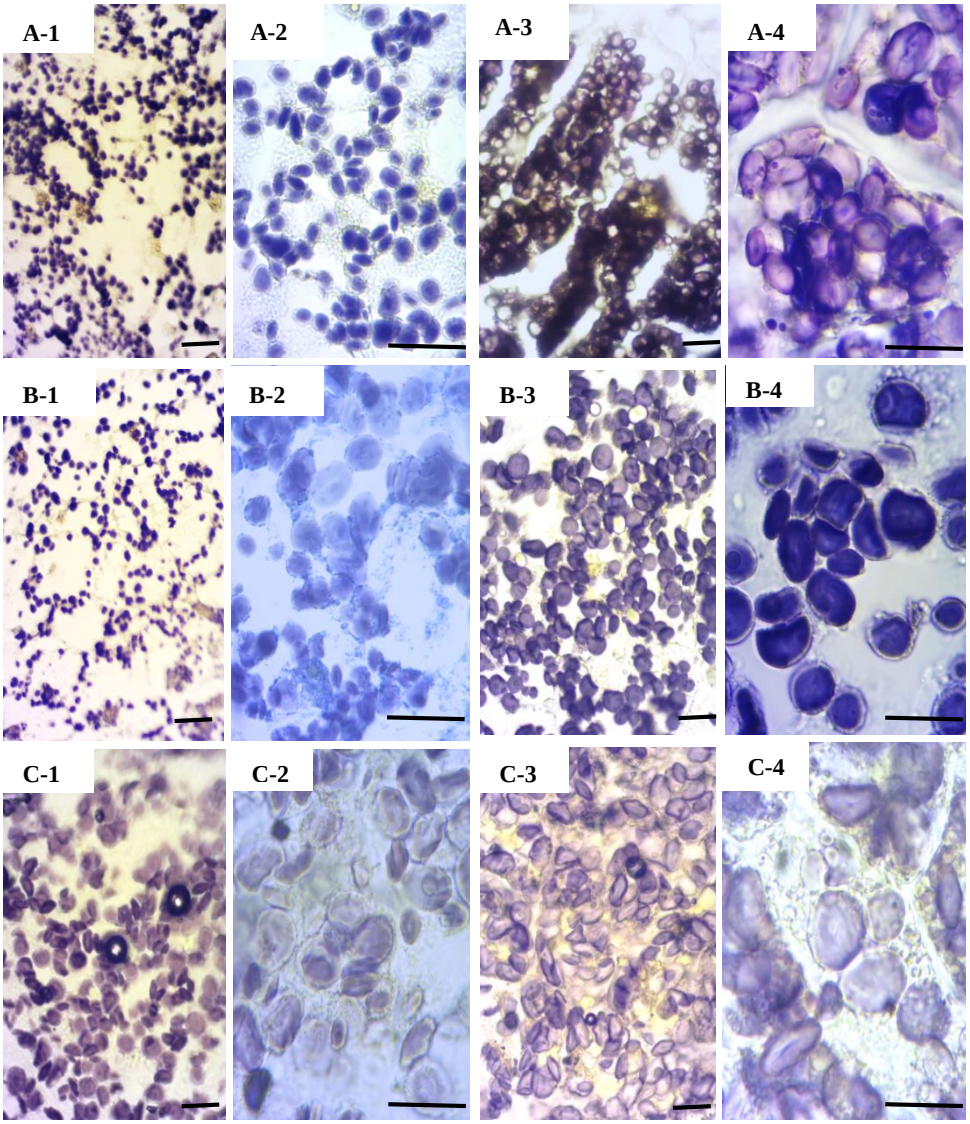


Figure 5.7 Micrographs of sections of grains stained with iodine solution by using Light microscopy (x40) and (x100).

Chinese spring (A1-A2 full length; A3-A4 mature grain), cadenza (B1,B2 full length; B3, B4 mature grain) and Durum (C1,C2 full length; C3-C4 mature grain).

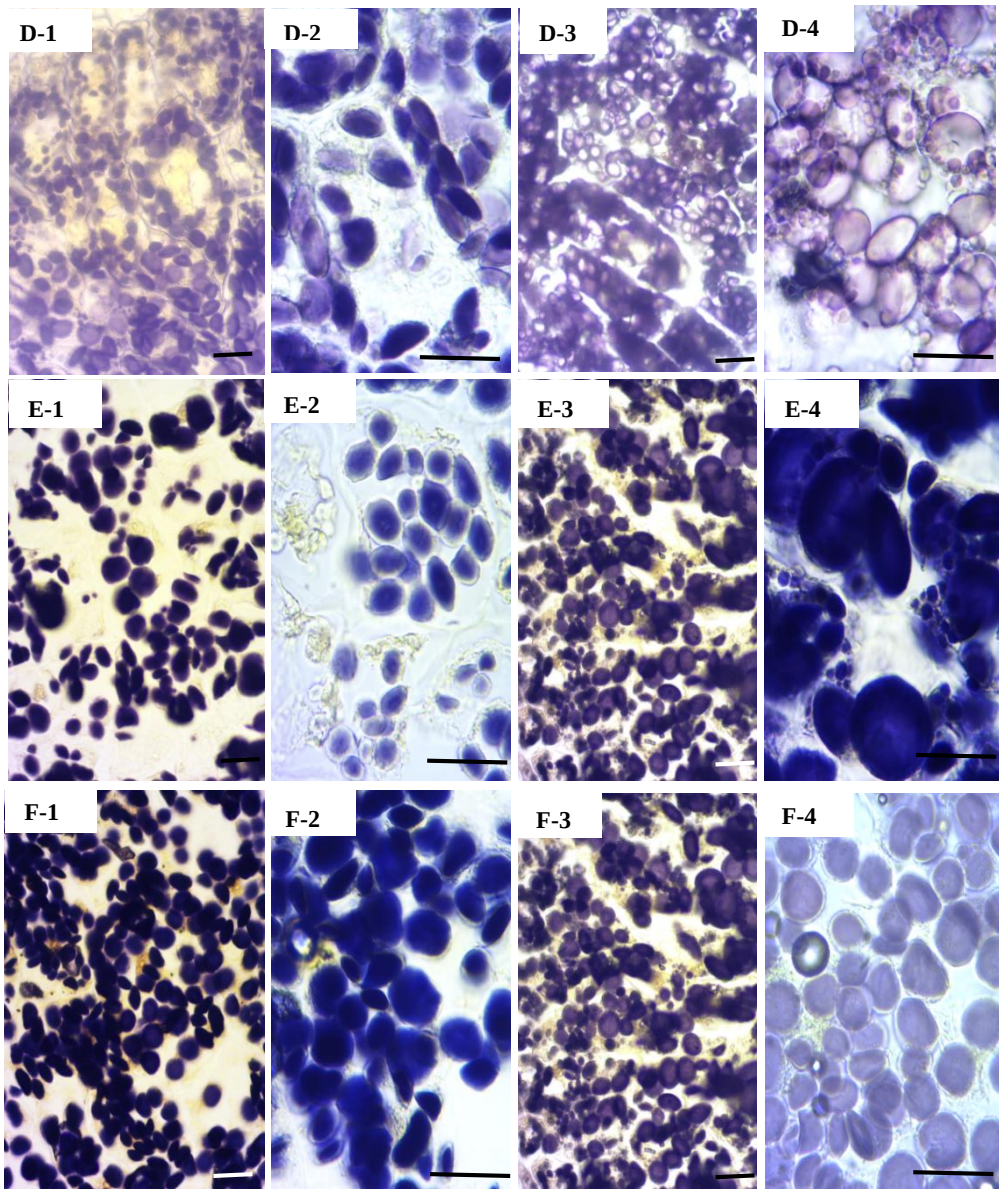


Figure 5.8 Micrographs of sections of mature grains stained with iodine solution by using Light microscopy (x40) and (x100).

Rosner (D1, D2 full length; D3, D4 mature grain), Ray (E1, E2 full length; E3, E4 mature grain); and Barley (F1, F2 full length; F3, F4 mature grain).

The Coomassie brilliant blue has a high affinity for aromatic amino acids and can specifically stain protein blue with high sensitivity (Hu *et al.,* 2015). In addition, the toluidine blue can be used to stain the storage protein (Jing *et al.,* 2014; Leroux *et al.,* 2014). In the present study, the endosperm sections were stained with Coomassie brilliant blue and toluidine blue. However, comparing the two staining, the Coomassie brilliant blue was better than the toluidine blue for storage protein staining. In most the sections, the protein bodies were clearly observed, were shown **in (Fig.5.9).** The storage protein was stained, but the protein bodies could not be clearly observed especially at the maturation **(Fig.5.10)** because of they were packed tightly comparing to middle stages or the full-length stage of the grain. However, they mutually integrate to form amorphous a protein matrix and individual protein bodies are no longer visible at the late stage of endosperm development (Wei *et al.,* 2009a, b). It was suggested that the storage protein in mature wheat grain might exist as a protein matrix and its morphology could not be analyzed. So that, in this stage only the distribution was analyzed (Zhao *et al.,* 2016). The present result agreed with that the storage protein was mainly distributed in the subaleurone endosperm (Oscarsson, 1997). Some pieces of literature have reported the morphology and distribution of storage protein in developing cereal endosperm (Gil-Humanes *et al.,* 2011; Hurkman and Wood, 2011). However, it is no way to compare the storage protein among different samples so that in the present study, the whole section of cereal seed provided the possibility to compare the storage protein in the same region of seed among different samples in the size only. In terms of the distribution, it is likely the different grains have the similar distribution.

Furthermore, the synthesis and accumulation of both storage starch and protein are initiated in the different time and in different regions of the grain. For example, the formation of starch granule in rice starts at the central region of the endosperm and propagates outward to the outer endosperm cells. This process is completed about 1 month after the initiation of starch granule formation in the central region of the endosperm (Hoshikawa, 1993). As a result, during the development of endosperm there are several differences happened which might result in the significant difference in the morphology of endosperm cell and starch.

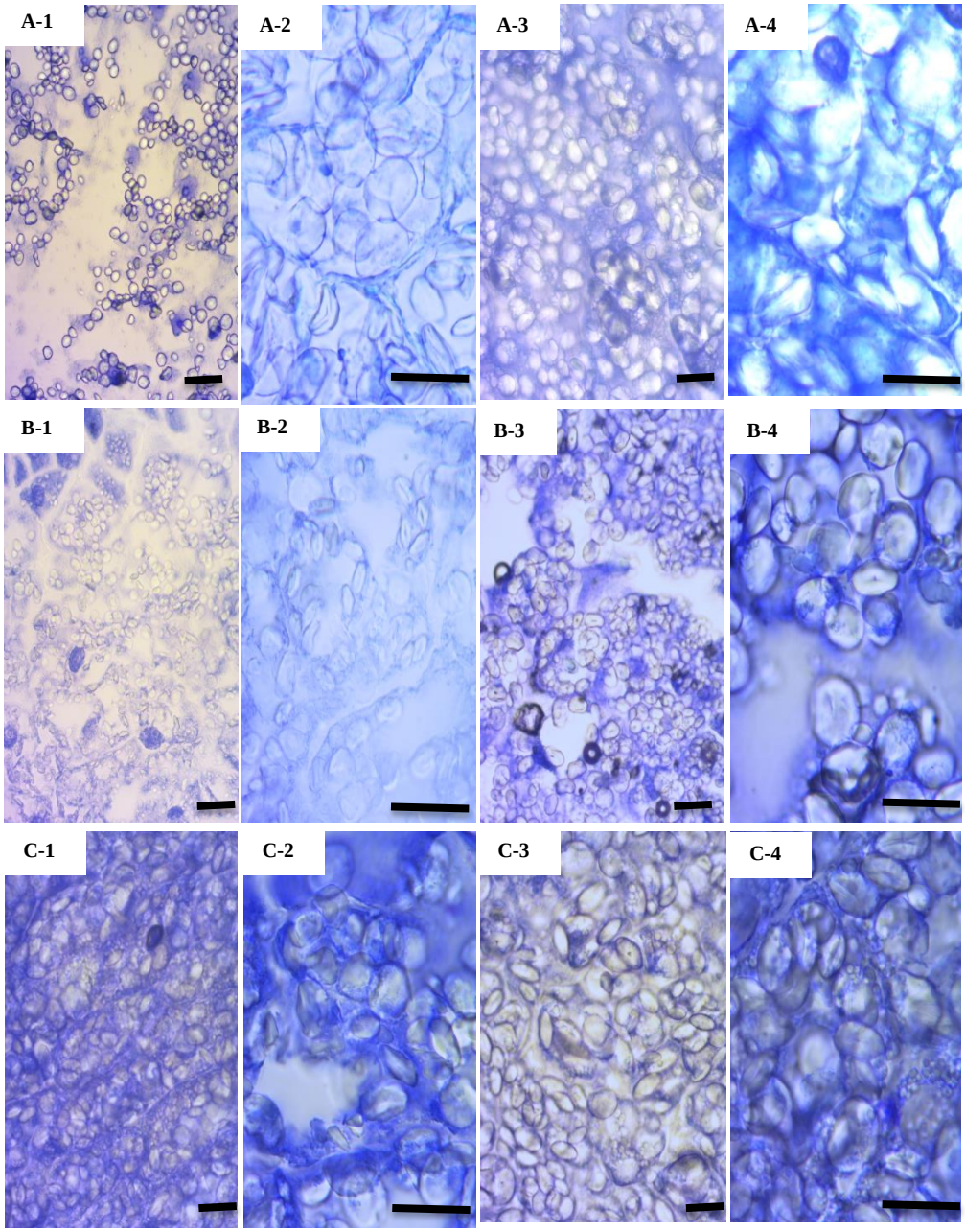


Figure 5.9 Micrographs of sections of mature grains stained with coomassie brilliant blue R250 by using light microscopy (x40) and (x100).

Chinese spring (A1,A2; full length- A3,A4; mature grain), cadenza (B1,B2; full length- B3,B4; mature grain) and Durum (C1,C2; full length; C3,C4 mature grain).

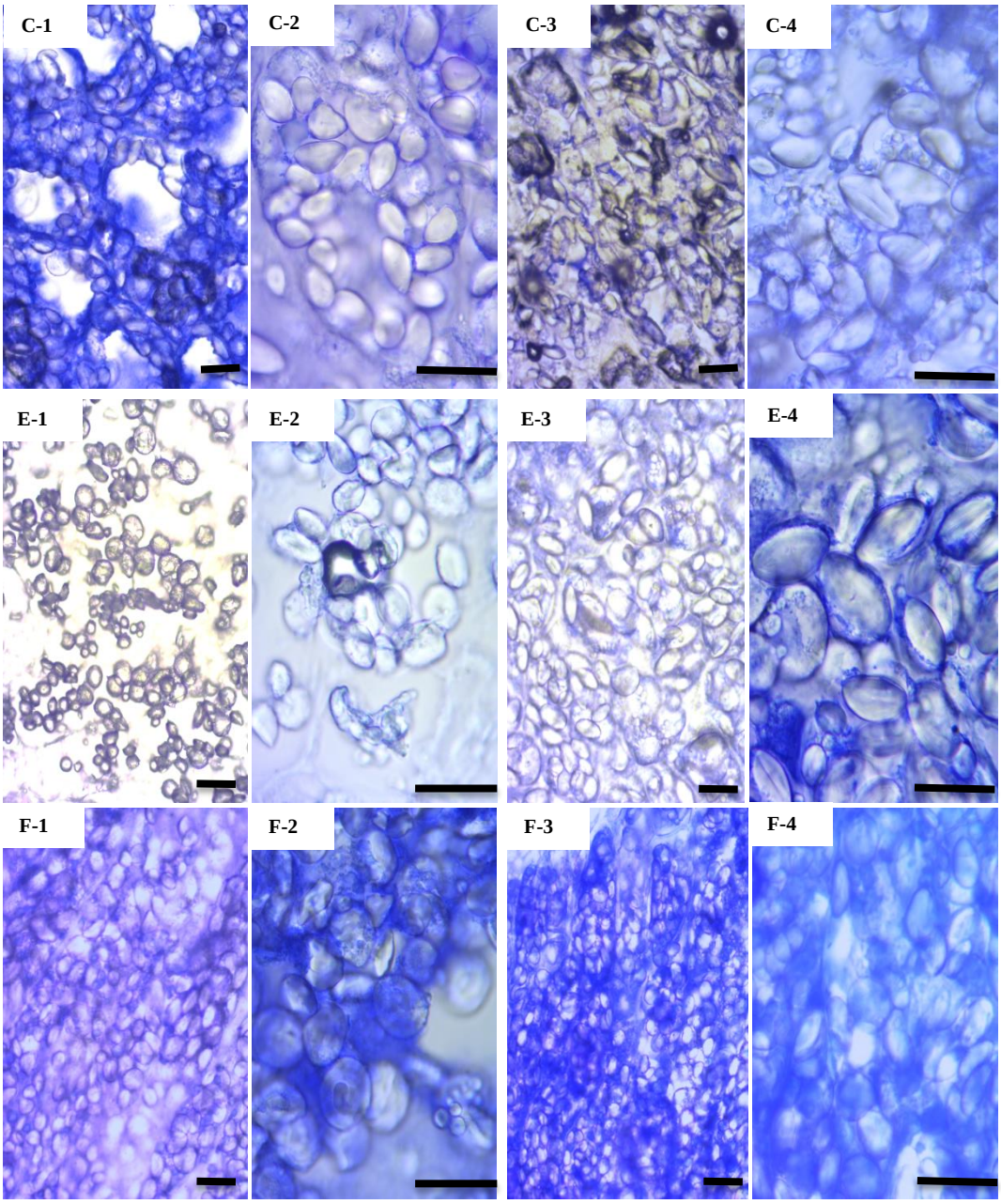


Figure 5.10 Micrographs of sections of grains stained with coomassie brilliant blue R250 by using light microscopy (x40) and (x100).

Rosner (C1, C2 full length- C3,C4 mature grain), Ray (E1,E2 full length- E3,E4 mature grain) and Barley(F1,F2 full length; F3,F4 mature grain).

Table 5.3 Morphology parameters of large and small starch granules, and endosperm cells in different samples of seed.

	Large starch granules				Small starch granules			
	Area (μm^2)	LAL (µm)	SAL (µm)	LAL/SAL	Area (μm^2)	LAL (µm)	SAL (µm)	LAL/SAL
Chinese spring	104± 22.18	16.02± 3.16	13.75± 1.40	1.16	39± 15.9	9.87± 2.04	7.34 ± 2.24	1.3
Cadenza	95± 16.66	15.4± 2.10	11.14± 1.45	1.3	36.0± 12.0	10.77± 1.14	7.55 ± 1.32	1.4
Durum	114± 55.4	17.88± 2.8	11.44± 1.8	1.5	40± 11.48	10.01± 1.60	8.95± 1.1	1.1
Barley	111.7± 30.63	14.04± 1.65	11.97± 2.1	1.17	40± 15.2	11.35± 1.45	8.77± 1.30	1.2
Rosner	108± 28.4	17.75± 1.36	16.75± 2.16	1.05	38.9± 16.14	12.05± 2.29	11.4± 2.29	1.05
Ray	112± 30.3	15.59± 2	13.03± 1.50	1.1	35.9± 17.9	12.40± 2.3	10.08± 1.6	1.2

Table 5.4 Morphology parameters of protein bodies in full length and maturation in different samples of seed.

	Protein bodies in full length				Protein bodies in mature grain			
	Area (μm^2)	LAL (µm)	SAL (µm)	LAL/SAL	Area (μm^2)	LAL (µm)	SAL (µm)	LAL/SAL
Chinese spring	38 ± 1.8	10.29± 1.5	9.82 ± 1.3	1.04	64±8.4	19.5± 2.4	17.0± 1.99	1.1
Cadenza	40.9± 3.6	11.8± 1.33	9.3± 1.15	1.2	53.8± 4.5	14.9± 0.53	14.5± 2.3	1.0
Durum	38± 5.8	14.7± 1.2	13.1± 1.7	1.1	64± 4.3	21.0± 8.2	15.8± 34.4	1.3
Barley	36±2.8	10.7± 0.65	8.9± 1.09	1.2	44± 4.1	13.01± 1.4	11.8± 1.57	1.1
Rosner	57.68± 4.6	18.1± 1.65	12.78± 3.1	1.4	49.7±	14.7± 11.1	27.3± 8.8	0.5
Ray	65.4±4.6	21.20± 3.8	15.9± 5.3	1.3	70±7.4	22.43± 3.0	17.4± 2.25	1.2

Data are means ± standard deviations, LAL: long axis length; SAL: short axis length; LAL/SAL: the ratio of long and short axis length.

5.4 Conclusion

In this chapter, there is an attempt to compare between the examined species in terms of the morphological structure of storage products which reflect the impact the genotype on the morphological characterization of starch, which has a significant effect on the end use and quality of grain. However, in this chapter, there is a limitation which makes this part is not obvious such as most the measurements were restricted in one stage rather than track the change in the structure and the accumulation to detect the relation between the rate of the accumulation and the regulators of storage products.

Generally, SEM results showed that no significant difference between the examined samples in terms of general features, while the most significant point was found is that the difference in the number of aleurone layers and the size of the starch cells. To investigate further, another method was used to look at the entire section of mature cereal seed in this study, after which these sections were stained with iodine solution, and Coomassie brilliant blue R250 to make the morphology of starch and the distribution of storage protein more obvious. The quantity analysis showed that the area and size of starch granules were significantly different in different regions of the endosperms of different grains, so that one region in the center of the grain was chosen to allow comparison between different species. This method could be very useful for the study of cereal seed development and utilization.

Amylopectin was higher than that of amylose, which reflects the importance of amylopectin as the main content in the starch grain (Kang *et al.,* 2012). Consequently, the end use of cereals depends on the properties of starches which in turn are dependent on their amylose/ amylopectin ratios. Thus, the measurement of the amylose content of starch is an important parameter for starch processing. To investigate the relationships between starch structure and the starch content in different species, it was confirmed that the amylose content was higher in A granules than B granules. From this point, it can be said that there is a

correlation between the amylose content in different grains with the size and the abundance of A granules. This finding was consistent with that of Ao and Jane (2007), who confirmed the high amylose content in A granule compared to B granule in wheat, triticale and barley. Another important finding was that the amylose content was higher in durum and barley, which have larger A granules during the full length and mature stages. Although the starch granules in rye were found to be largest amongst the species studied, the abundance of A granules was very low in rye. Rosner showed a certain similarity to the hexaploid wheat in terms of the size of A granules but the number of B granules was relatively higher in the Rosner than for any of the other grains; also, the distribution of starch granules was completely different. It was noted that the starch granules were located around the endosperm cell walls during the full-length stage. However, the granule size is related to the molecular architecture of amylopectin and its molecular arrangement within the granule (Raeker *et al.,* 1997; Geera *et al.,* 2006; Jane, 2006).

To conclude, the proportion of A and B granules can affect the properties of wheat starch. The starch granules and amylose content of different wheat varieties can influence on the structural and functional properties of starch, and a significant correlation was found between A granules and amylose content. In addition, the size of each starch granule has a significant impact on the end use and the physicochemical features. For example, wheat A-granule starch has been used in carbonless copy paper and the wheat B-granule starch is good for plastic film filler. In addition, from our result, it can be confirmed that the starch granules differ in the morphology in different samples which reflect the way of biosynthesis which is controlled the morphology of starch is different between the species that is unknown (Ao and Jane, 2007). However, this aspect of the analysis needs further work, especially on A granules and B granules in isolation, to better understand the relationship between the structure of starch and other features. Structural diversity and complexity of the starch granules have considerable potential to affect starch modification processes.

In terms of the relation between the total starch content and amylose content was found there was no obvious correlation. The highest starch content was found in the hexaploid wheat, and the lowest starch content in the rye, while the highest amylose content was in the tetraploid (durum) and barley. Another point is that it is known that the accumulation of starch increased rapidly during the full length and then decrease during the maturation, the reason could be related to the position of accumulation which is in the pericarp and endosperm during the full length and then is restricted in the endosperm tissues during the maturation. In addition, the rate of the accumulation of starch and protein increase rapidly during the full length which is correlated with the high expression level of some endosperm- TFs such as TaSPA, and it correlated with the increase the size of starch granules and protein bodies as well.

Chapter 6

Conclusions & Perspectives

Chapter 6 Conclusions & Perspectives

Summary:

In this project, I have characterized and established the relationships between TFs (TaSPA/HvBLZ2) (TaBLZ1/HvBLZ1) of O2-like members, and members of euAP2 family (RSR1/Q) which are known to influence in controlling the accumulation of the starch and protein in the endosperm of varieties of wheat (hexaploid species and tetraploid), barley, and other species such as triticale and rye. This was done to address the possibility that the variation between these TFs could the main reason for the variation in the quality of grain derived from different species. The regulation of gene expression is central to any developmental process including grain development, and regulation of gene expression is mediated among other mechanisms by TFs. Most of these TFs have been functionally characterized in maize, barley and rice, however, bread wheat (*Triticum aestivum* L.) is one of the major staple crops for human nutrition (Reynolds *et al.*, 2009). It is known that wheat has a large and complex genome with three A, B, and D genomes (2n = 6x = 42, AABBDD); therefore, each wheat gene potentially occurs as a trio of A, B, and D homologous (Feldman *et al.*, 2012). The complex genome of wheat has led to complicated regulatory mechanisms within the cells of the allopolyploid genome. Given this, the majority of genetic analyses of cereal endosperm development have focussed on diploid species such as barley (*Hordeum vulgare*) and maize (*Zea mays*) (Olsen *et al.,* 1999; Becraft, 2001; Olsen, 2001). To bypass this limitation, for the first family (bZIP) we used *ZmO2* maize as an example of O2-like members, and also to identify all members of O2 in (bZIP family) in the monocots and dicot. For another candidate gene, *RSR1* is a member of the euAP2, I faced difficulties due to the limited information which was published for the function of this gene.

6.1 (RSR1 and Q); euAP2 family.

An attempt was made to identify euAP2 clade in monocots and dicots by constructing the phylogenetic tree, to clarify the relationship between them. Two major duplication events were identified in the euAP2 monocots lineage, and a third one was in the dicots. These two clades in the monocots which are represented by (RSR1 and Q) as

paralogous, and both are specific lineages in the monocots, while the third duplication event in dicots is referred to as the (euAP2/ TOE3) clade in Arabidopsis. To our knowledge, this is the first analysis to explicitly show that Q and RSR1 are paralogous. Furthermore, one of our goals as to identify protein domains, particularly DNA binding sites, linkers, and the N-terminus and C-terminus of the AP2 domain, conserved motifs and the functional domain within the euAP2 gene clade. In addition, a large number of motifs were found across the monocots and dicots, some of which are likely to play a role as repression domains, such as the (LxLxL) motif which presents in all euAP2 proteins. Another motif is (MLDLNV) in the N-terminal of euAP2 genes, which manifests as transcriptional repression activity and has a role in protein-protein interactions. Another conserved motif in all euAP2 sequences is the miR172 binding site. All these features confirmed the close relationship between them, although there is a change in the protein sequences, and which can occur after the duplication event. This analysis may be explained by the fact that the role of duplication events is one of functional diversification (Conant *et al.,* 2014). The change in the gene structure and expression attributed to several reasons such as the evolution and polyploidy events. Also, the duplication events cause new traits between the species or new functions originated. This alteration in gene expression leads to phenotypic variation and diversification between the species.

Moreover, this project has also provided a deeper insight into the temporal and spatial expression of these genes, especially within the context of a comparative approach within the certain samples of Triticeae, which can be considered significant in at least two major respects. First, the expression patterns during the life cycle of the plant for euAP2 confirmed that they have a broad function in the development of the plant at various phases (vegetative and reproductive). Our analyses explored the euAP2 monocots, which have conserved expression patterns such as in the vegetative phase which confirms the essential nature of euAP2 for the vegetative growth of plants. Second, the duplication event leads to divergence in the function of two *euAP2* genes, which likely have divergent roles in specific patterns during the life cycle of the plant, as has been confirmed from mRNA-ISH. Generally, with *euAP2* genes have

overlapping functions in a number of developmental processes such as cell growth and cell division for the ovule and leaf, though there is a specific role for each *euAP2* gene.

Our data shows that from the expression patterns of paralogous (RSR1/Q), euAP2 could have a conserved function in all floral organs in monocots. Furthermore, the result of in situ hybridization suggests that *TaQ* were expressed highly in the early of the establishment of floral meristem identity, and later in all floral organs, which reflect the role of *Q* in the controlling spikelet meristem identity.

One interesting finding was that all *euAP2*, whether monocots or dicots, have one conserved binding site which is miRNA172. However, the *euAP2* can be classified into two groups which are perfectly complementary to miR172d, namely *TaRSR1* and another which has a single nucleotide mismatch at position 20 of the miRNA binding site, such as *Q*. However, the regulatory abilities of miR172 over its targets can be affected by mismatches. In chapter 3**,** I attempted to compare between the expression of miR172 and its targets (*RSR1*-*Q*) in varieties of wheat and barley in different tissues, particularly during the spikelet meristem and seed development, which were missing in the Arabidopsis. As a result, the findings show similar expression patterns of miR172 in examined species. Overall, these results indicate that the miR172 sequence is conserved in most euAP2, which also reflects the conserved function.

6.2 (O2 members); bZIP family.

In chapter 2, bZIP (O2 members) were characterized in terms of the structure, the phylogeny, and the sequence analyses. The phylogenetic analysis of O2 members in monocots and dicots indicated that there are members of O2 are present as a specific lineage in the Poaceae. So that, there is duplication with the Poaceae generating the TaSPA/HvBLZ2 and TaBLZ1/HvBLZ1 as paralogous. This further develops the previous phylogenies where the bZIP divergence in monocots and dicots noted (Vincentz *et al.*, 2003). Therefore, O2 members can be divided into three groups: two groups of monocots (TaBLZ1 and TaSPA), and the third group with dicot species, such as Arabidopsis, which has paralogous such as AtBZO2H3, AtbZIP25, and AtbZIP10, which are the close to O2-like in the Poaceae.

In addition, it was found that wheat has two paralogous (TaSPA and TaBLZ1), which are orthologous (HvBLZ2 and HvBLZ1); both are evolutionarily close and, indeed, wheat probably has a closer phylogenetic relationship with barley and Brachypodium than other species. Also, the presence of paralogous genes in the grass lineage suggests that duplications occurred after the split between dicots and monocots. In addition, the divergence between them because of the duplication, which is likely to provide a new functional specificity (Vincentz *et al.,* 2003), or generate a new protein function or proteins synthesized in different cell types at different levels.

To date, and to the best of our knowledge, there is only one piece of work in the literature on HvBLZ1 (Vicente-Carbajosa *et al.,* 1997), and no data on TaBLZ1 wheat. Hence, the functional relationship and temporal regulation are still poorly understood, and only partial information is available about the HvBLZ1 and HvBLZ2 as regulators for the hordeion gene in barley. Generally, attempts were made to characterize the TaBLZ1 to at least describe the relation between them, and indeed other proteins in O2. It was found that the relation between these TFs was supported by the presence of conserved motifs beside the bZIP domain. The differences and similarities in the bZIP domain and the conserved domain between O2 members lead to identifying key changes in these regions that may be correlated with functional divergence. In addition, it was found that the basic region has a conserved NLS in all the O2-like proteins. In this study confirmed that both paralogous (TaBLZ1/HvBLZ1) have two NLS motifs, one in the N-terminal and another within the basic region which was identified as bipartite. As the phylogenetic analysis shows that the divergence between the TaBLZ1 and TaSPA, also both reflects the divergence in their function which can be detected from the spatial and temporal expression. The result reveals that the mRNAs of TaSPA and TaBLZ1 have overlapping spatial distributions in starchy endosperm cells, being greater in the early stage for TaBLZ1, while the expression for TaSPA/HvBLZ2 starts at the mid-length stage of grain. A possible explanation for this might be that the early expression of BLZ1 (3-5 DAA) could be involved in the initiation of protein synthesis in the endosperm, or with cell development and growth, whereas TaSPA/HvBLZ2 play a role in the accumulation of proteins through regulation of specific endosperm protein

storage genes. In addition, TaBLZ1 is broadly expressed in different tissues beside the starchy endosperm cells, which indicates that this factor could play other functional roles in different tissues such as vegetative tissues (leaves and roots) or in the reproductive tissues. However, the expression patterns of TaBLZ1 support those observed in earlier studies of bZIP, showing that these TFs have roles in diverse aspects of developmental, floral induction and development.

6.3 Comprehensive yeast two-hybrid analysis reveals specific dimerization preferentially between candidate TFs

Another general finding is that most of our target TFs are temporally integrated during the development of endosperm via a network of transcriptional regulatory which controls both starch and protein synthesis processes. In this project, five TFs (TaSPA, TaPBF, TaBLZ1, TaRSR1 and TaMADS47), which have been reported to have a role during the filling stage of endosperm. The results of Y2H shows the interactions between them, also they have the ability to form different complexes with different interaction partners. Consequently, our results strongly imply that these proteins play critical roles during the grain filling, whether as transcriptional activators or repressors of seed storage protein genes or starch synthesis genes. However, the interaction patterns of the TFs studied in yeast show that many, but not all which could be these the interactions could have preferential dimerization between them.

6.4 Do TaSPA and TaRSR1 are able to regulate both starch synthesis genes and seed storage protein genes?

In addition, one of the most important findings in this project was related to the mechanism of integrated regulation of genes involved in both starch synthesis genes and storage protein genes. The important step to identifying a regulatory network is the identification of the various master regulatory elements in diverse target promoters. Accordingly, a major objective in this project has been to explore the relation between bZIP and euAP2 members to regulate starch synthesis genes and storage protein genes which provide an overlapping mechanism to activate these genes. From the analysis of promoters of some starch synthesis genes and storage protein genes which detects several binding sites of different transcription factors (TFBS), that potentially involved in the regulating these genes, such as bZIP, euAP2, DOF. This data suggests

that the most starch synthesis genes and low molecular weight share the cis-elements for several TFs such as O2 members, PBF and RSR1 that are likely to be critical to activate these genes. The result of the transient assay shows that the TaLMW shows activation by TaSPA, but it increases with the combination between other TFs such as TaPBF and TaBLZ1. Nonetheless, current result of this study strongly suggests that TaSPA, TaRSR1, and TaPBF are likely to directly regulate starch synthesis genes and storage protein genes, although TaSPA shows differences in the activation of the examined starch genes as whether activators or repressor. The overall project story is shown in **(Figure 6.1).**

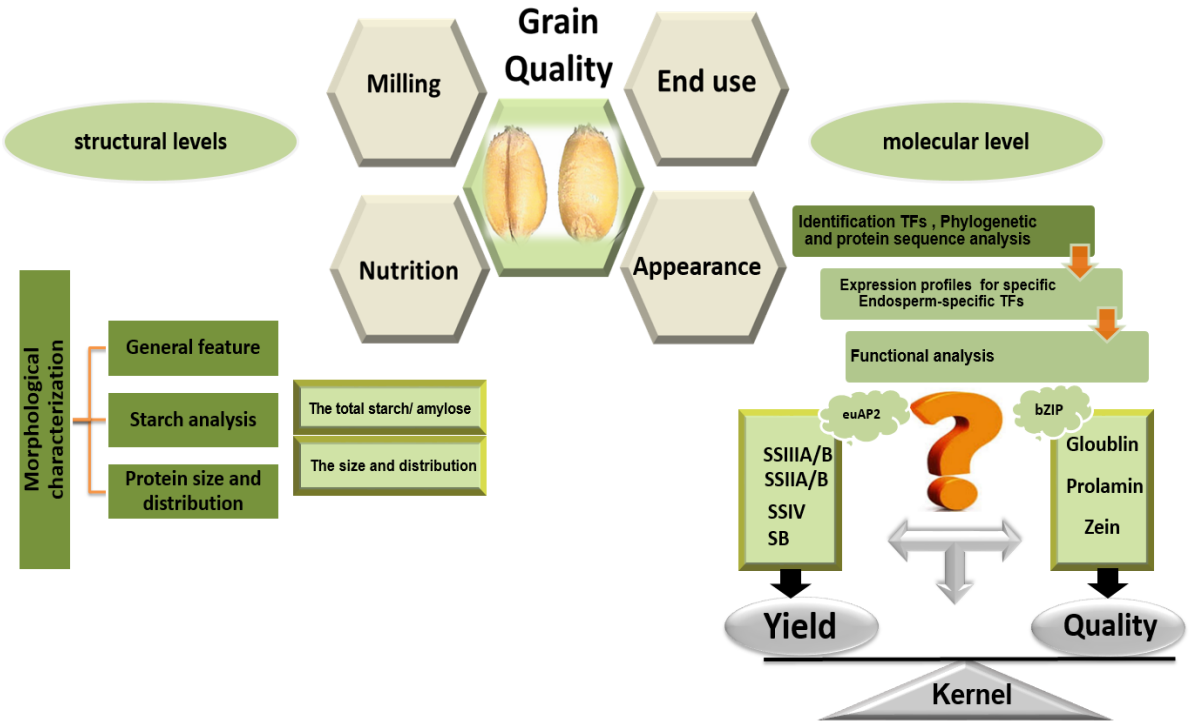


Figure 6.1 Schematic diagram of my research illustrates the overall role of TFs to control the the grain quality.

6.5 Future work

The master regulators of storage proteins and starch in the endosperm of different species are the TaSPA (ZmO2, HvBLZ2), TaPBF and TaRSR1. Although the significant work in maize revealing details of the regulatory network, it is still unclear the detailed regulatory mechanism of protein and starch accumulation in the

endosperm. Therefore, the regulatory network features should the focus of future work. From this perspective, rather than focus on one or two of these regulators, is better to identify the regulators and the relation between them as a network during the development because of it is unclear the reasons for the variation in the accumulation of the protein and starch in the grains of different species is still a matter of debate. Another point is that several motifs have been detected in the structures of TFs which could play a critical role in their function. However, these motifs need a further experimental examination to confirm their function. In addition, the broad expression of euAP2 paralogous (TaRSR1 and TaQ) confirms their pleiotropic function, so that the future efforts must concentrate on identifying the comprehensive array of gene regulation and ideally validate their relationship with these genes through global *in vivo* binding assays such as ChIP-on-Chip. Finally, this study lays the groundwork for future research into the importance of identifying the role of euAP2 in different species' plant lineages, which it must be determined in order to investigate the function of these genes.

Chapter 7

Appendices

Appendices: Media, Buffers and Solutions

Appendix 7: Media, Buffers and Solutions

7.1 Media

1. 1% Luria Bertani (LB) agar

Luria Bertani powder (Oxoid)	12.5g
Agar No 1 (Oxoid)	5g
Distilled water	500ml

2. YPDA Broth

10g Yeast extract
20g Bactopeptone
20g Dextrose (glucose)

3. 2X YPDA Broth

Reconstitute one YPDA Broth pouch in 250 ml ddH2 O and sterilize.

4. 0.5X YPDA Broth

Reconstitute one YPDA Broth pouch in 1 L ddH2 O and sterilize.

7.2 Buffers

1. Tris EDTA (TE) buffer 10X:

Tris HCl pH9	100mM
EDTA	10mM
SDS	1%
Distilled water	Up to 1L

2. AP Buffer:

1M MgCl ÷ 20, **1M Tris/1 M NaCl ÷ 10**, e.g. 20ml MgCl, 40ml Tris/Nacl + 340 ml H_2O

to give 400ml buffe

3. DNA extraction:

Tris-Hcl	10 ml	200mM
Nacl	2.5 ml	25 mM
EDTA	2.5ml	25Mm
SDS	0.5%	2.5 ml of 10%

Up to 50 ml

4. Phosphate buffer saline (PBS)

Phosphate buffer saline tablet (Oxoid)	1 tablet
Distilled water	100ml

Use a magnetic stirrer to dissolve the tablet.

6. Protinase K Buffer

100 mM TRIS, 50 mM EDTA.

7.3 Solution

1. BSA Solution:

1% solution in TBS with Triton (Surfactant). (**0.1g BSA + 30ul Triton to 10 ml TBS**).

2. Developing solution:

Prepare just before use. No surfactant in this mix and so slightly more than the antibody may be needed to completely cover slides. 60ul NBT/BCIP to 40 ml buffer to cover 10 slides

15 ul of NBT50, 75 ul BCIP10 to 10 ml AP buffer.
Add NBT (Yellow) first to avoid precipitation.

4. Caesium Chloride

To prepare 80 % calcium chloride stock solution, add 80 g $CaCl_2$ to 1 L of H_2O.

5. 1M Tris HCl pH 7.5

Tris HCl 78.8g
Distilled waterup to 500ml
Add 400ml of water, then use 10M and 5M NaOH to adjust pH to 7.2. Then top up to 500ml.

6. Proteinase K

Proteinase K (Fisher) 20mg

upH_2O 1ml

7. FAA.

3.7% formaldehyde, 5% acetic acid, 50% ethanol

8. Hybridisation Solution

Hybe solution (HS)	1600 ul (Sufficient for 16 “Sandwich” slides or 8 ½ slide chambers).

10X in situ salts	200
Deionized formamide	800
50% dextran sulfate	400
50Xdenhardts solution	40
tRNA	20
H_20	14

9. 20X SSC

NaCl	175.4
3.0M	
Na-Citrate	88.2
0.3M	

dH_2O to just under 1L, then pH to~7.0, then bring to 1L and autoclave

10. 50X Denhardts

BSA	0.5
PVP	0.5
Ficoll	0.5

dH_2O to 50ml and store at -20°C

11. 0.2x SSC

25ml 20x to 475 ml sterile water to give 0.2X

12. 10X PBS

NaCl	80	1.37M
KCi	2	27mM
Na2HPO4	11.5	
(Sodium Phosphate dibasic)		
KH2PO4	2	
(Potassium Phosphate monobasic)		

13. 1 % solution in TBS

For 1 L: 24 g Tris base; 88 g NaCl; Dissolve in 900 mL distilled water. pH to 7.6 with 12 N HCl, add dH_2O to a final volume of 1 L .

For a 1X solution, mix 1 part of the 10X solution with 9 parts dH_2O and adjust pH to 7.6.

14. 1X TAE (buffer gel)

Make a concentrated (50x) stock solution of TAE

242g of Tris base and dissolving it in approximately 750 ml of dH_2O, add 57.1 ml of glacial acid and 100 ml of 0.5 M EDTA (pH 8.0). After that, adjust the solution to a final volume of 1 L. This stock solution can be stored at RT. The pH of this buffer is not adjusted and should be about 8.5.

Prepare a Working Solution of TAE

The working solution of 1x TAE buffer is made by simply diluting the stock solution by 50x in deionized water. Final solute concentrations are 40 mM. Tris-acetate and 1 mM EDTA. The buffer is now ready for use in running an agarose gel.

15. 10X loading dye

3.9 mL glycerol
500 µL 10% (w/v) SDS
200 µL 0.5 M EDTA
0.025 g bromophenol blue

0.025 g xylene cyanol
Bring to 10 mL total volume with H_2O.

16. Aureobasidin A Stock Solution Recipe:

Dissolve 1 mg Aureobasidin A (Cat. No. 630466) in 2 ml of absolute ethanol for a stock concentration of 500 µg/ml. Store at 4°C.

17. Aureobasidin A Working Concentration

Add just 125 µl of the Aureobasidin A stock solution to 500 ml of dropout agar media for yeast two-hybrid screening, yielding a final concentration of 125 ng/ml.

18. X-a-Gal Stock Solution

Dissolve X-a-Gal (Cat. No. 630463) at 20 mg/ml in dimethylformamide (DMF). Store X-a-Gal solutions at –20°C in the dark.

19. X-a-Gal Working Concentration

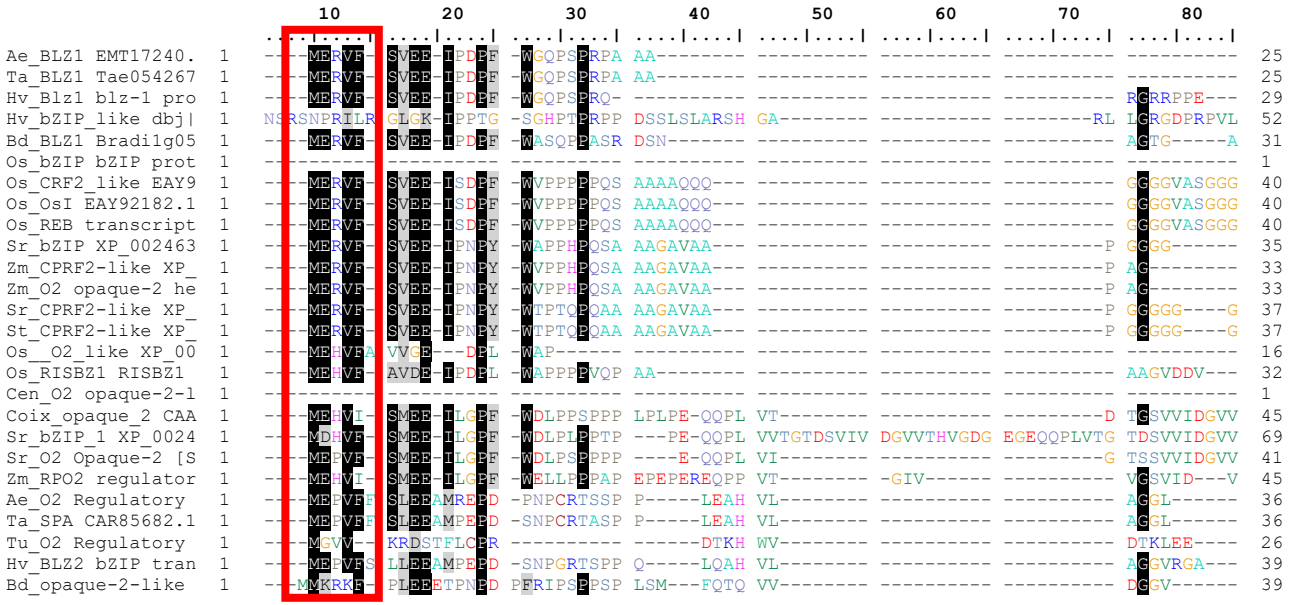
Add 1 ml of the X-a-Gal stock solution to 500 ml dropout media for yeast two-hybrid screening.

Appendices: Supplementary Data Chapter 2

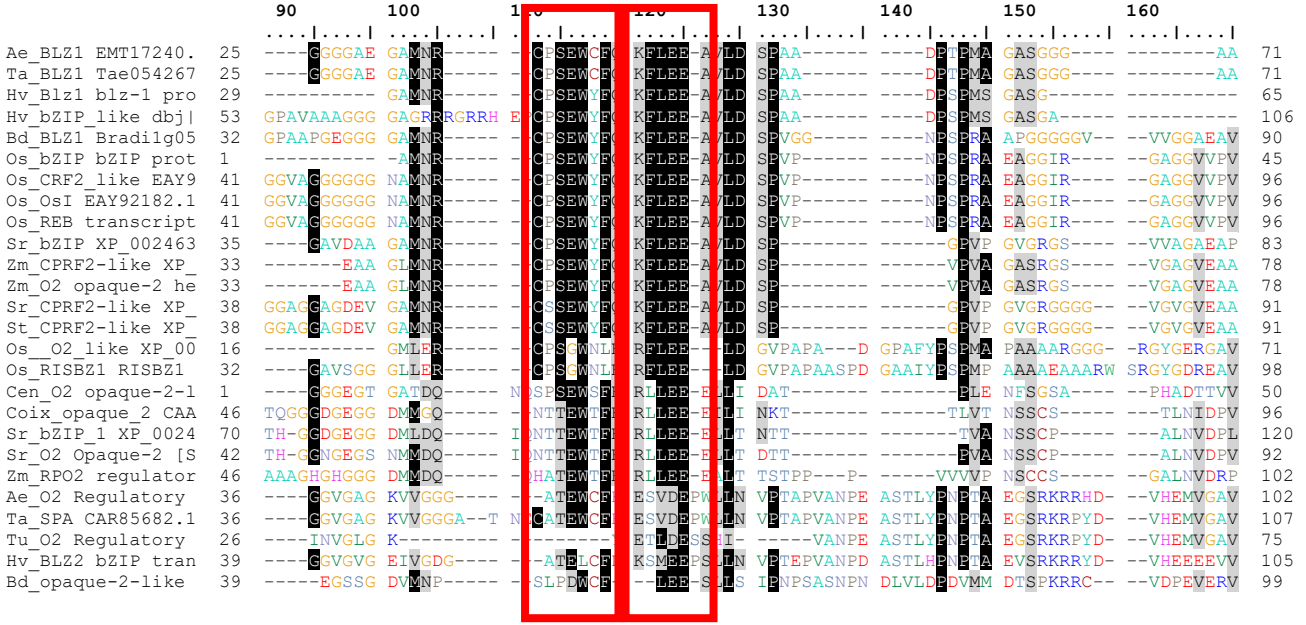
Data chapter 2: Alignment protein sequences.

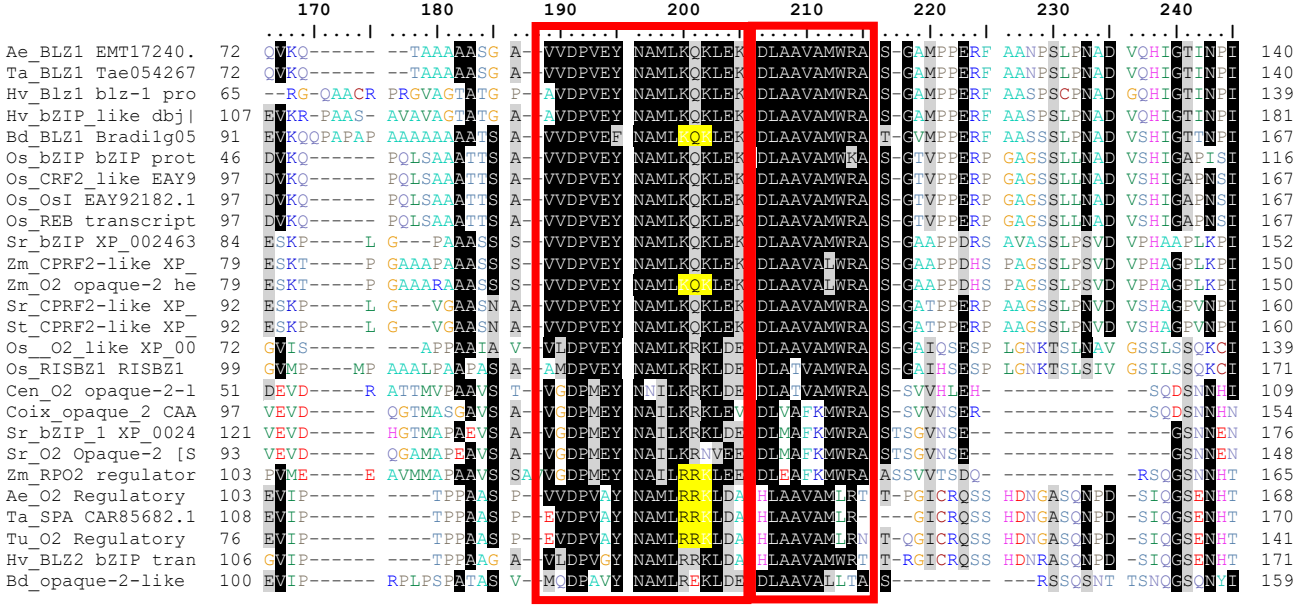
bZIP (O2) family.

A1 group

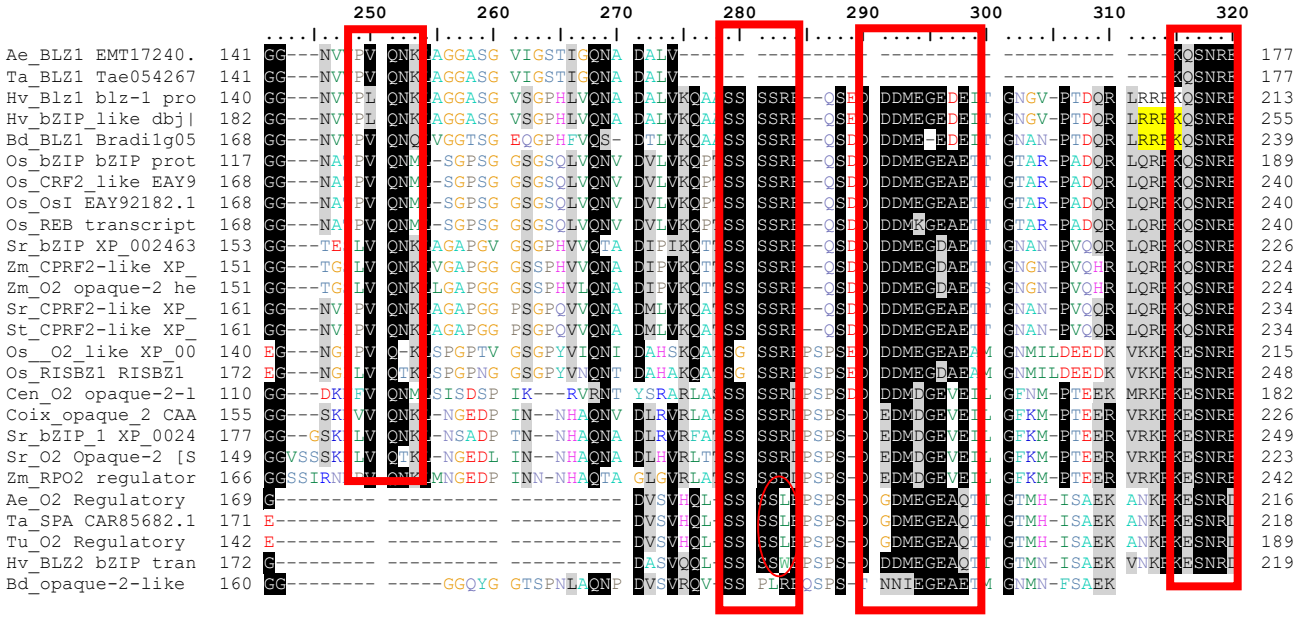


Continued on next page...

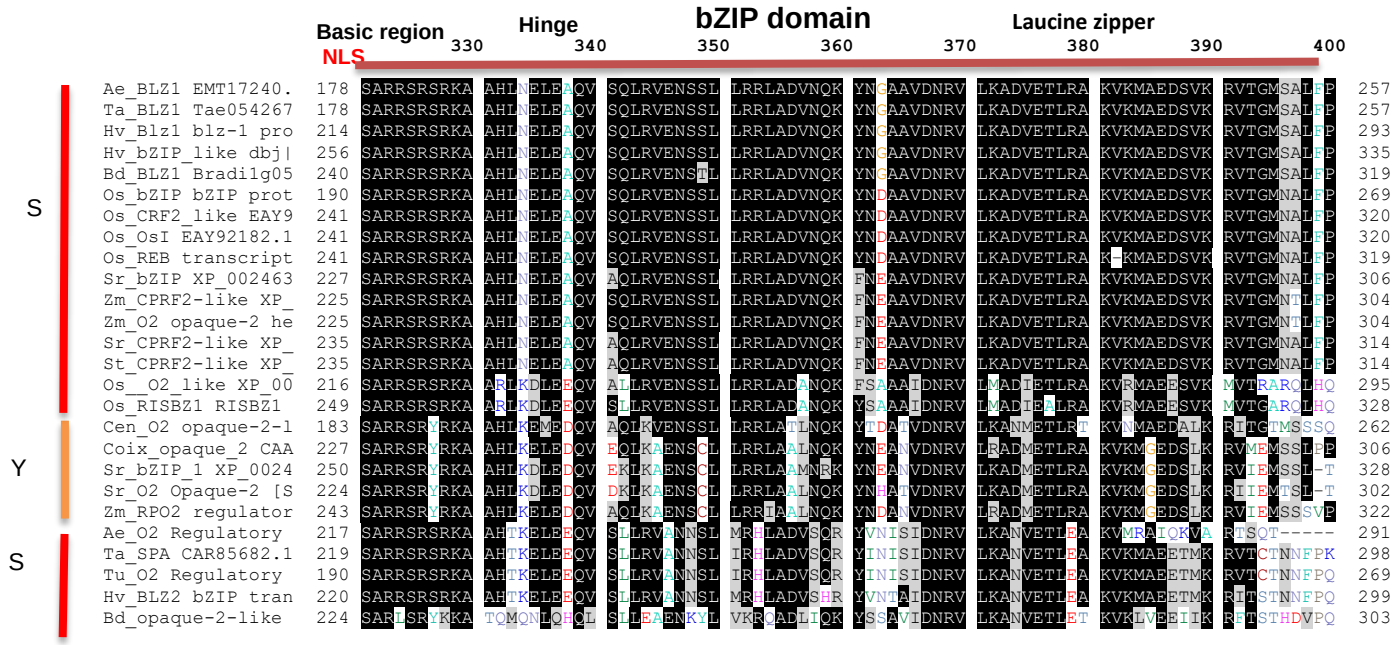




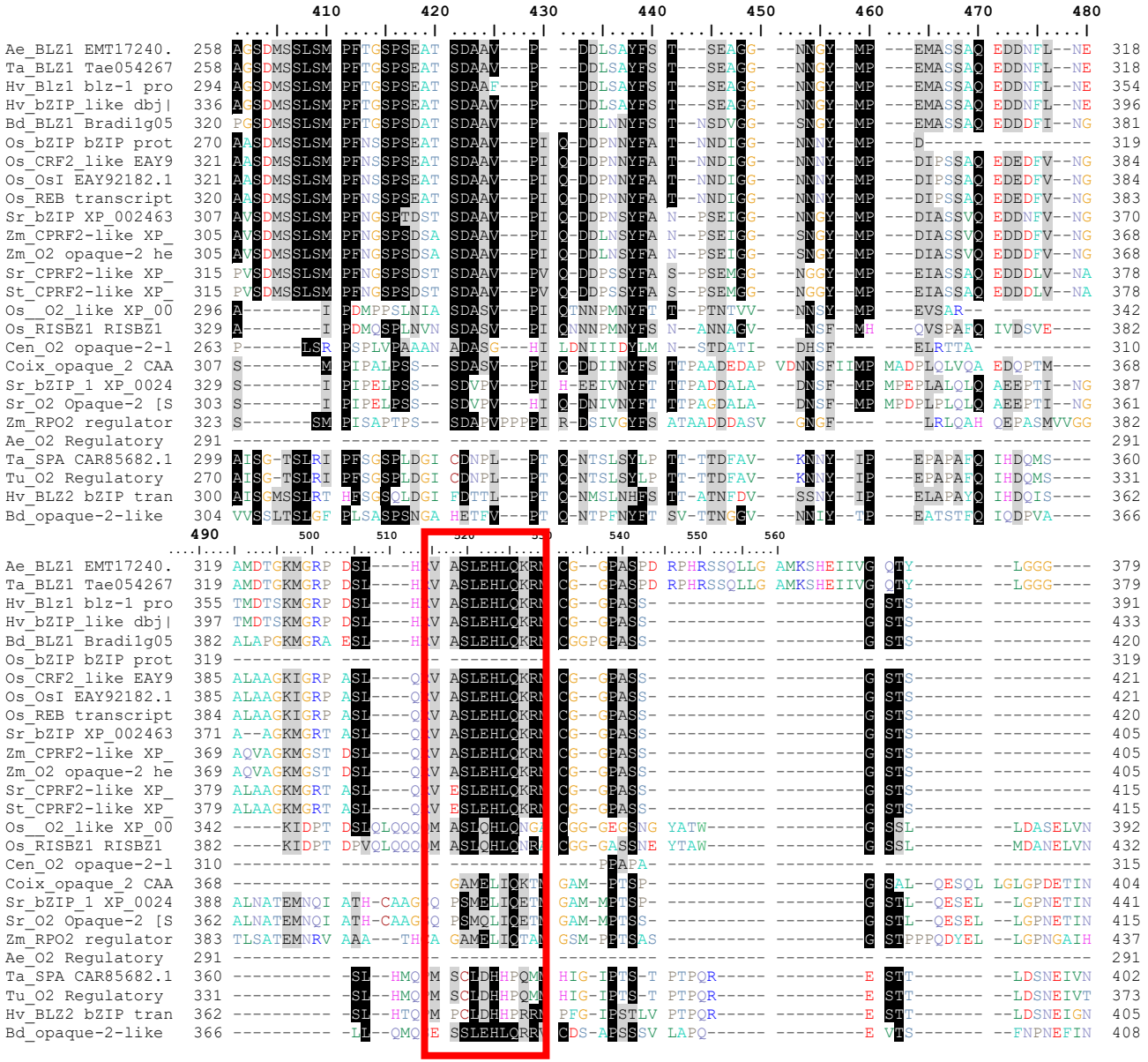
Continued on next page...



Continued on next page...

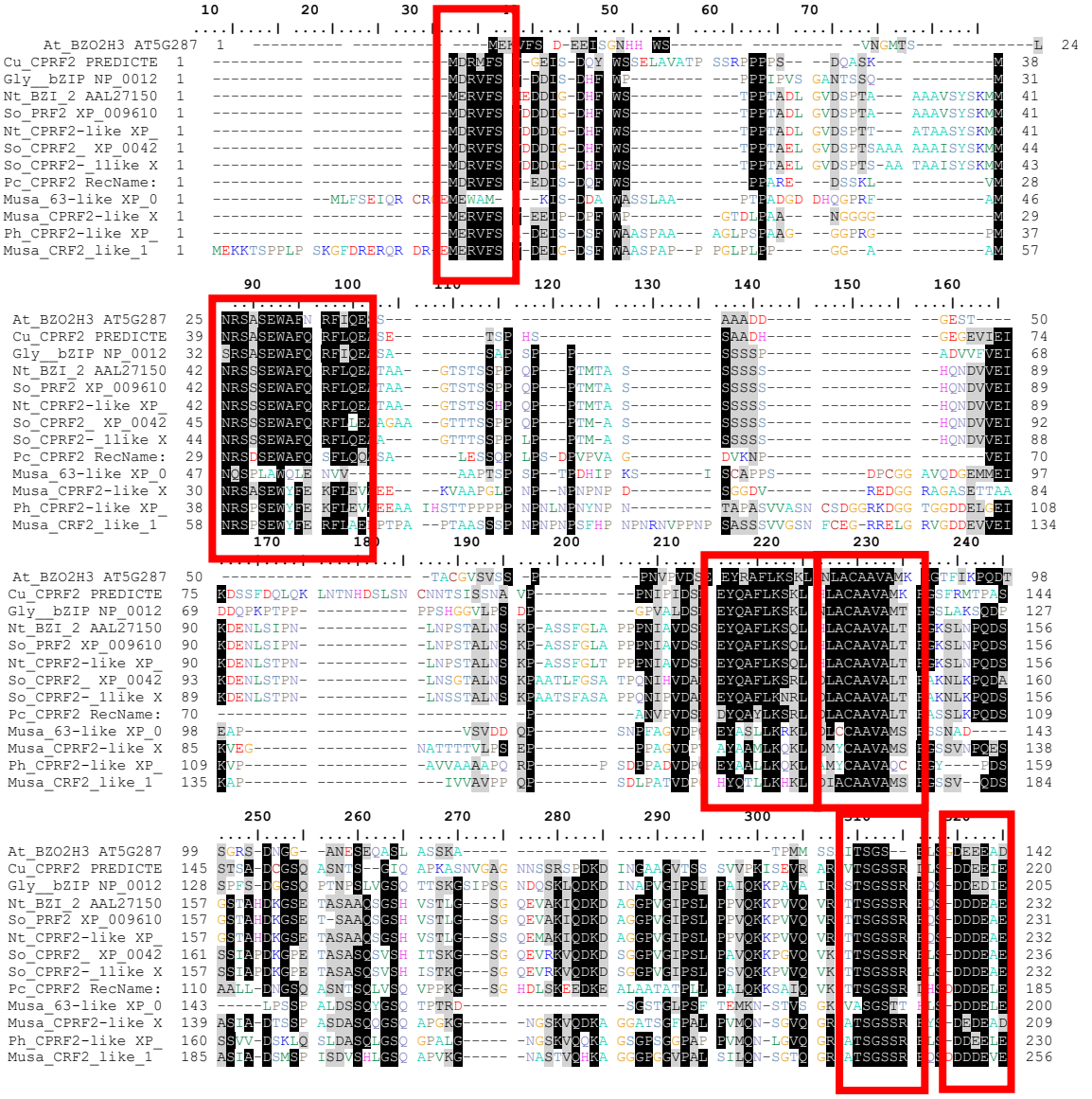


Continued on next page...

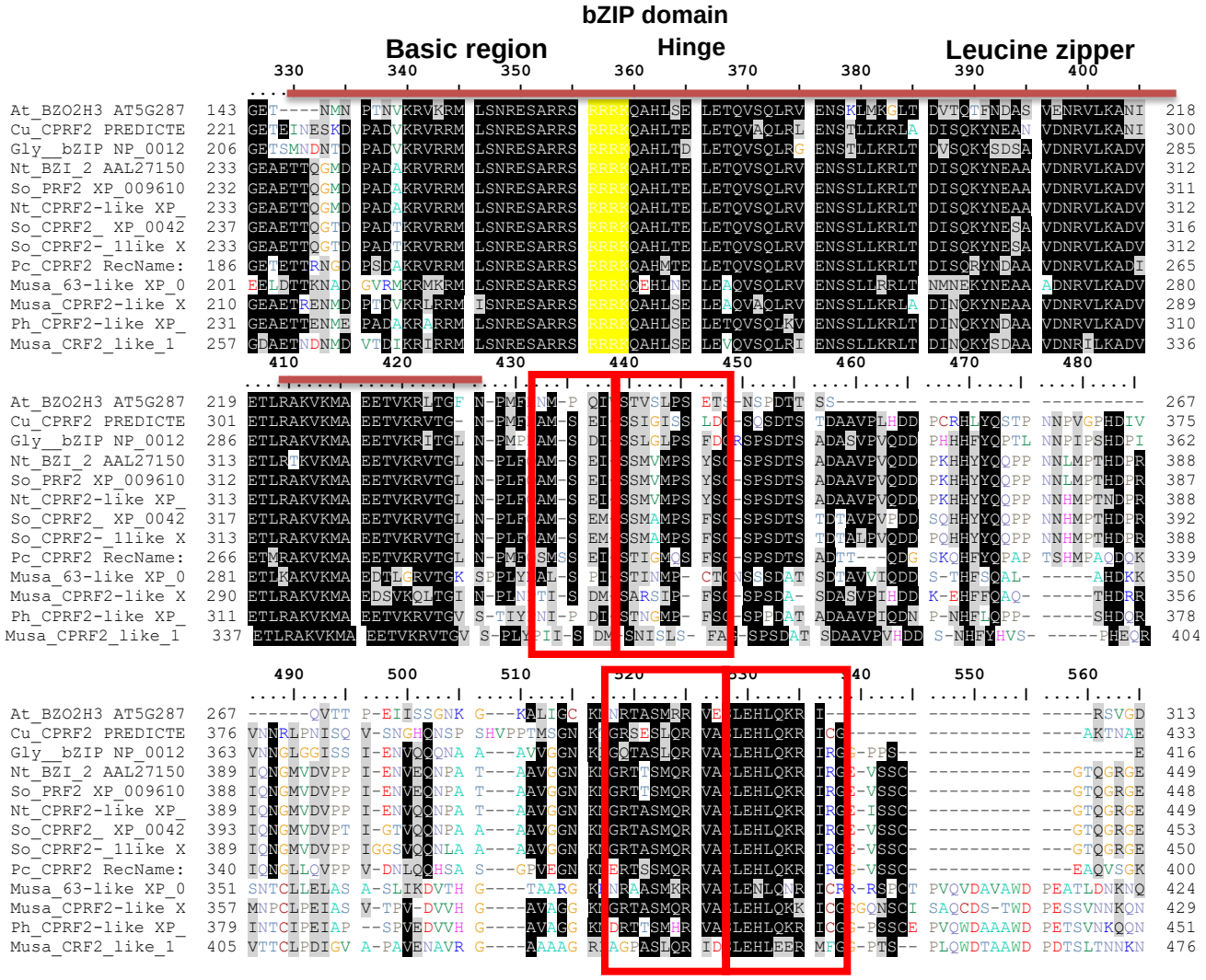


Continued on next page...

A2 group

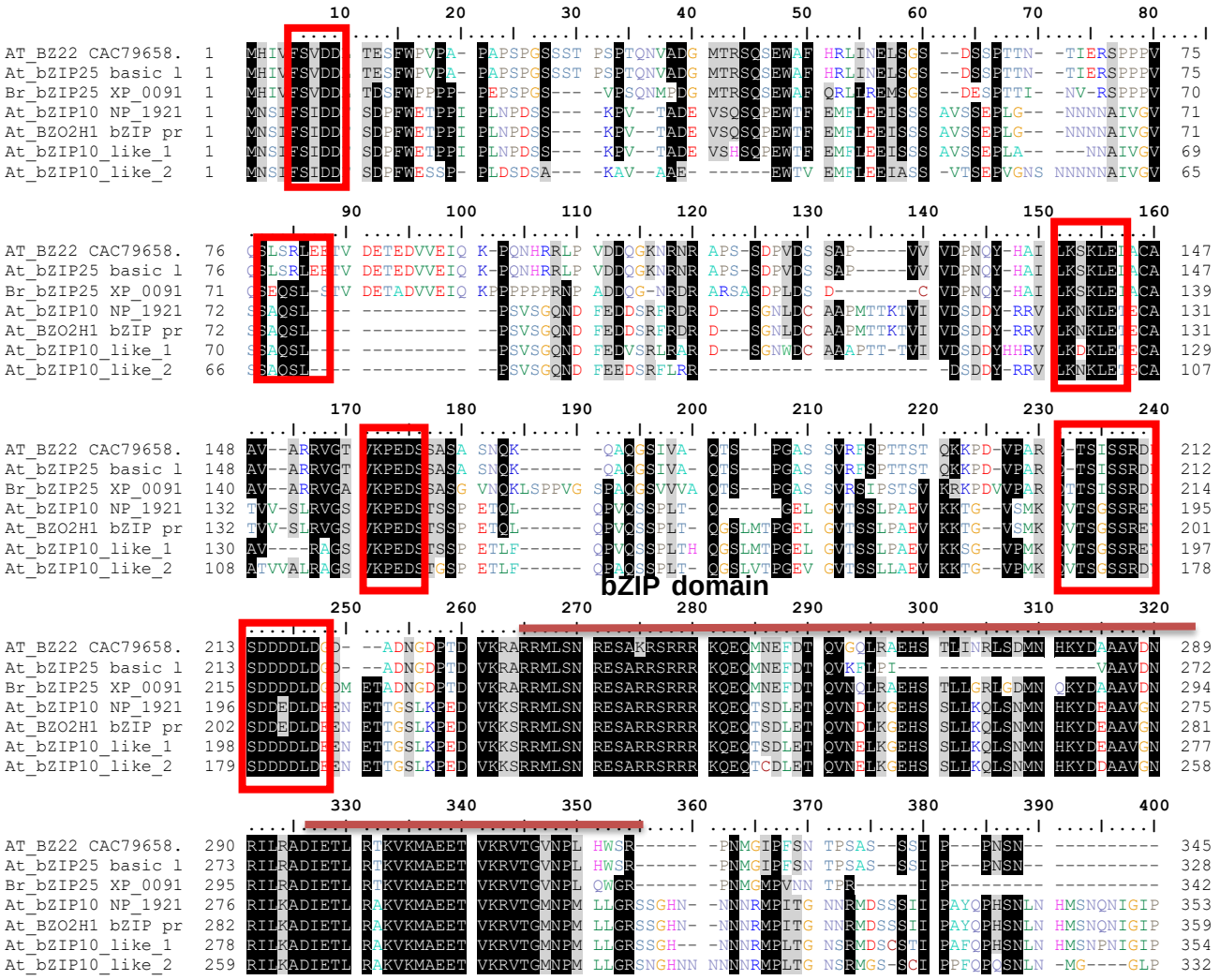


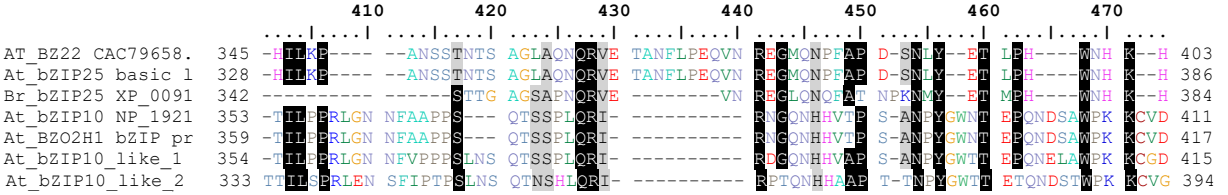
Continued on next page...



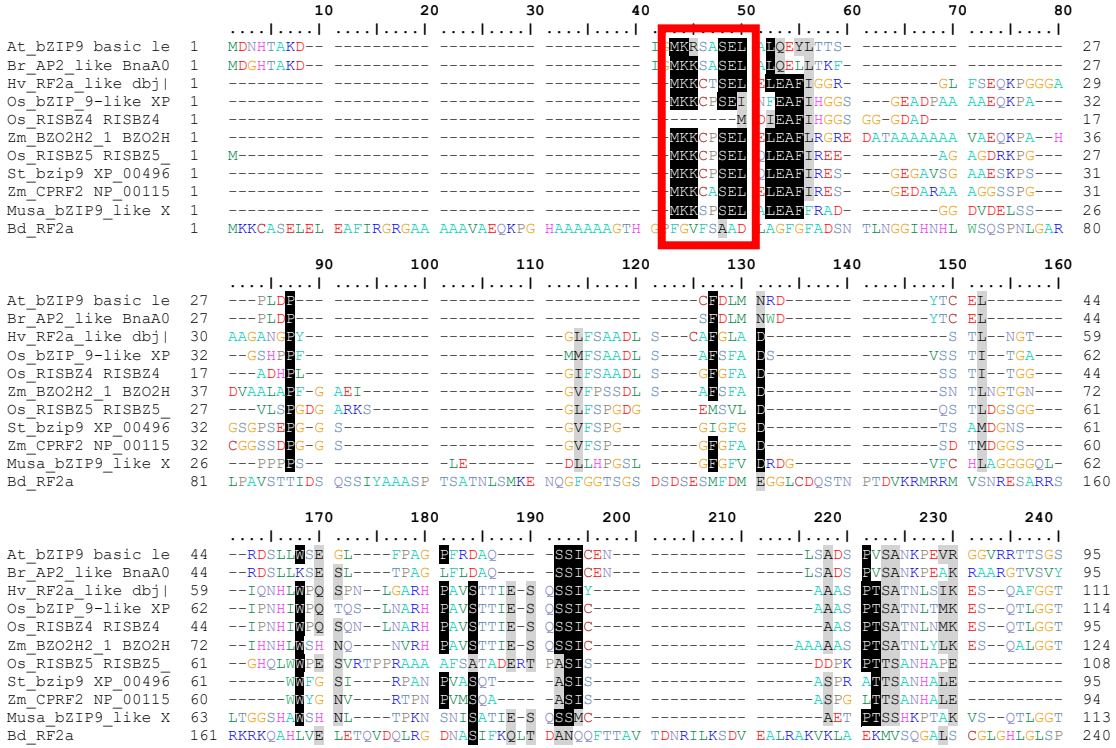
Continued on next page...

A3 group





B group



Continued on next page...

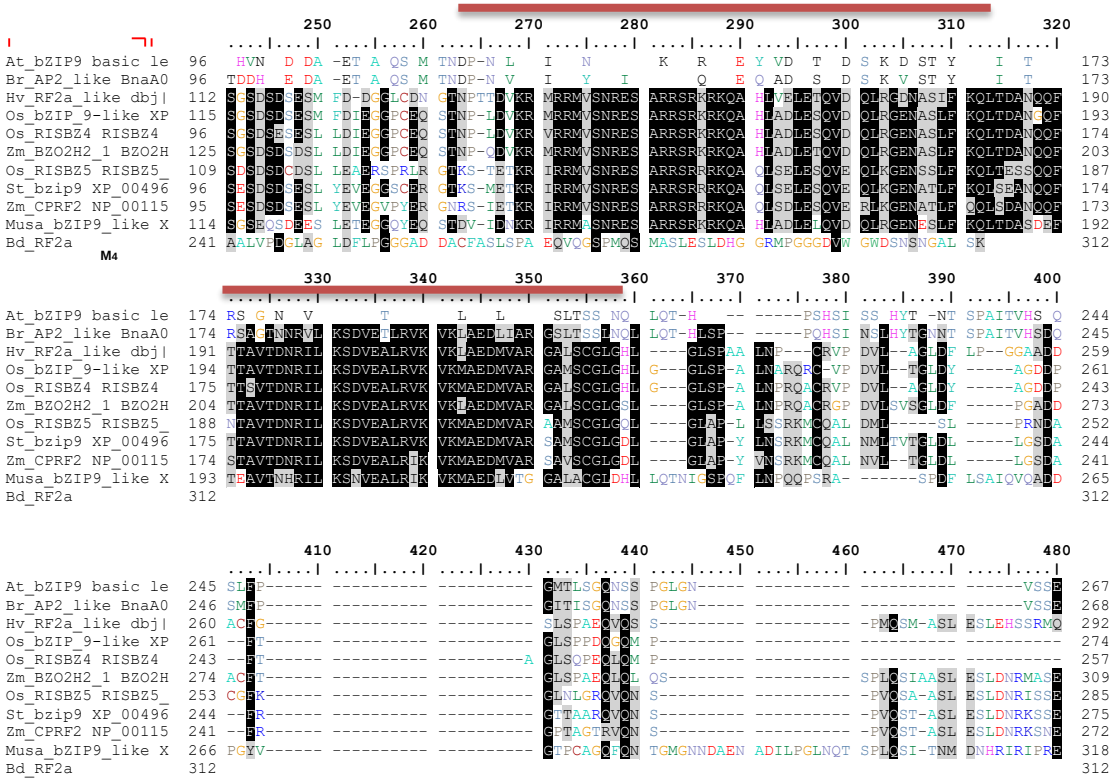
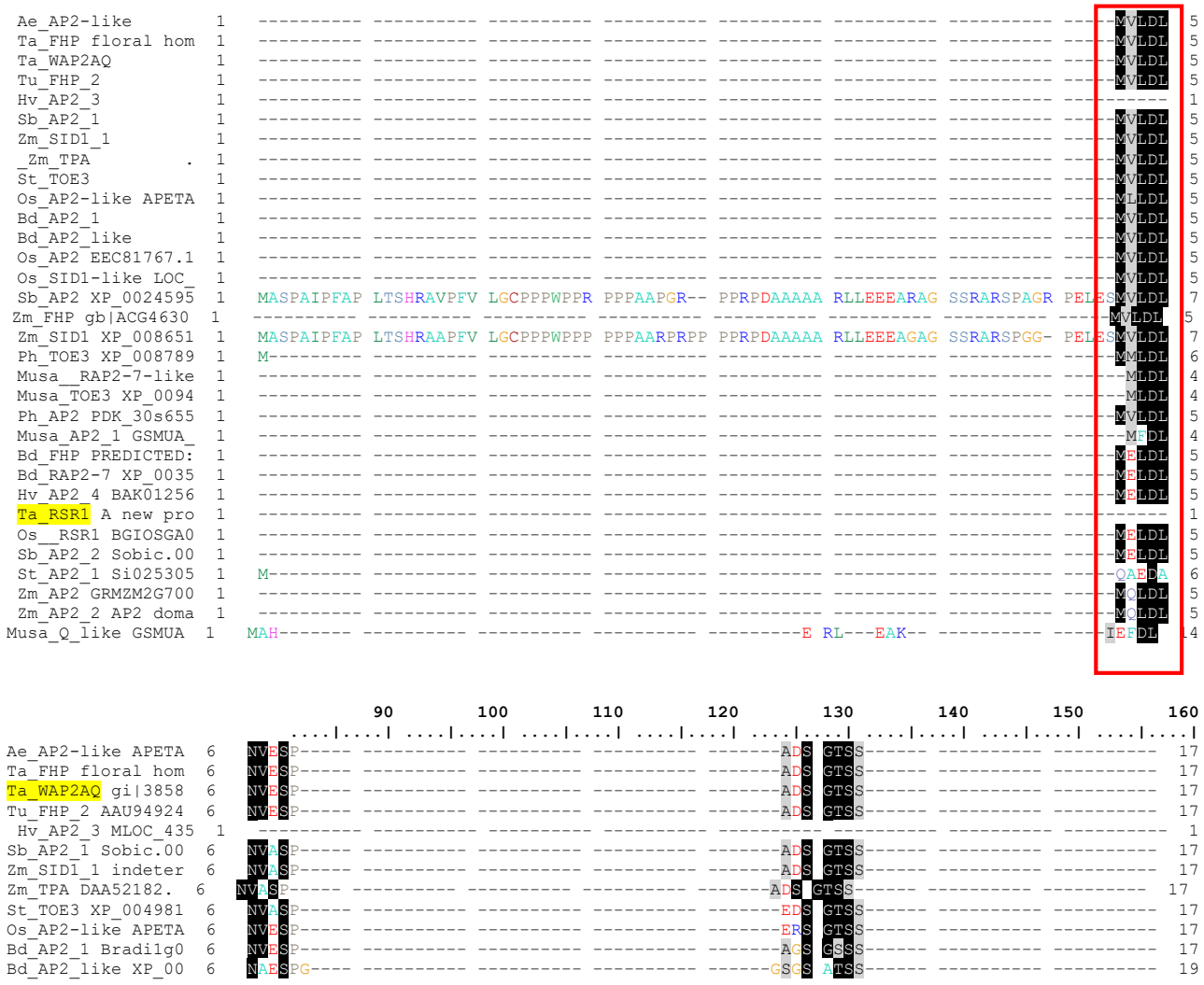
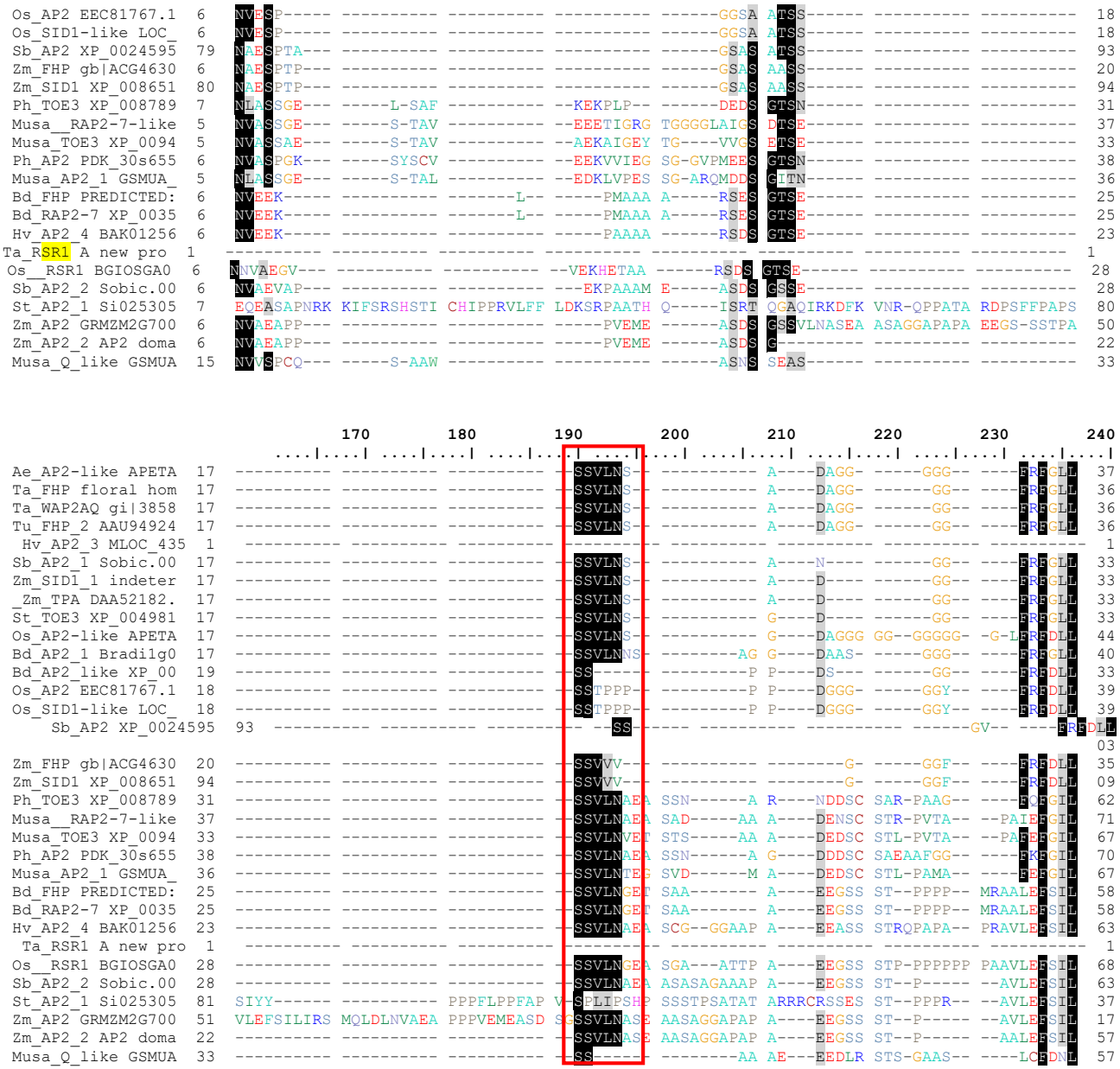


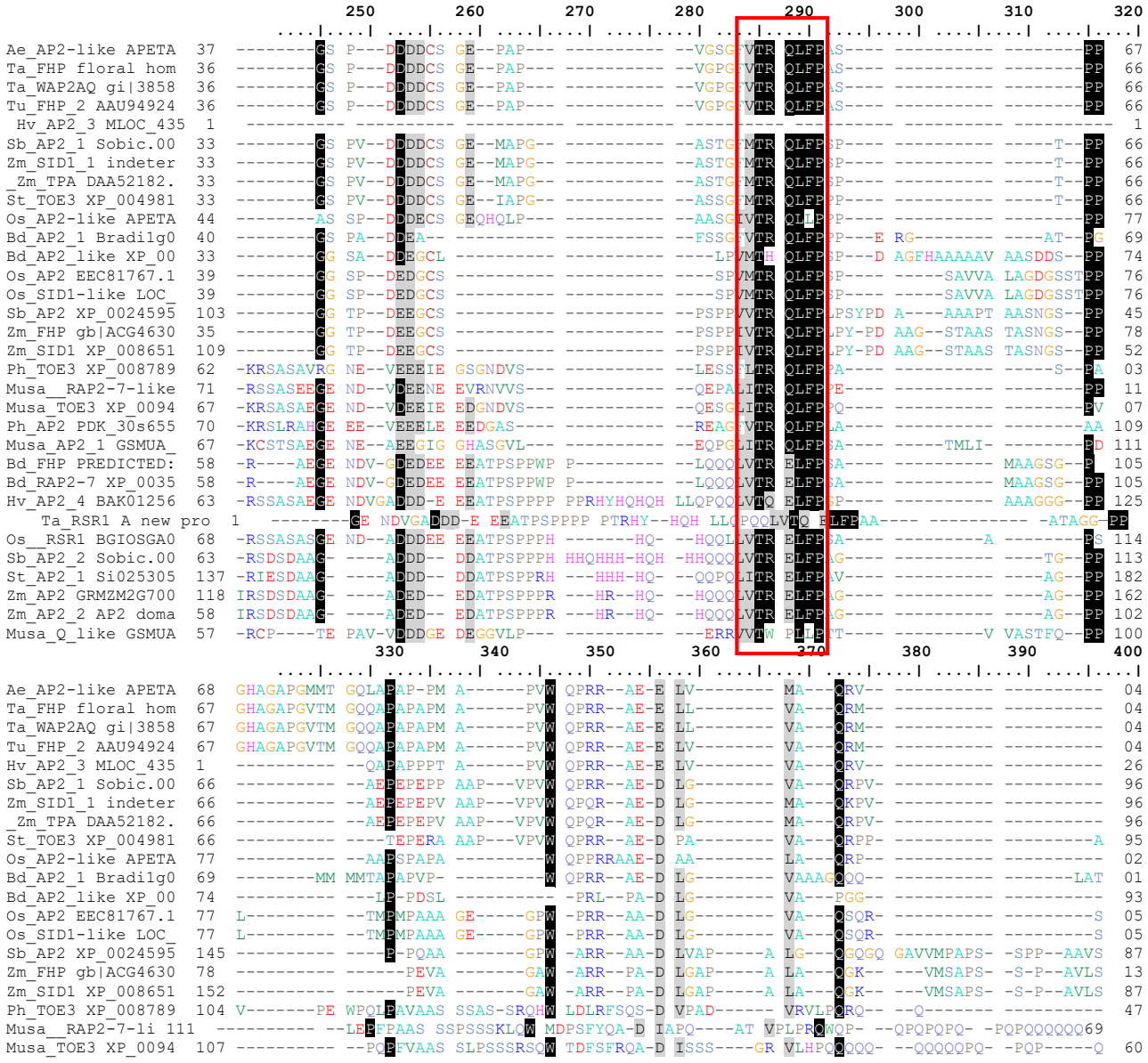
Figure 7.2.1: Alignment the whole sequences of proteins of 4 groups of bZIP family, conserved motifs were boxed, the bZIP domain upper red line, B- the bZIPdomain with the sequence logo, C- Conserved motifs.

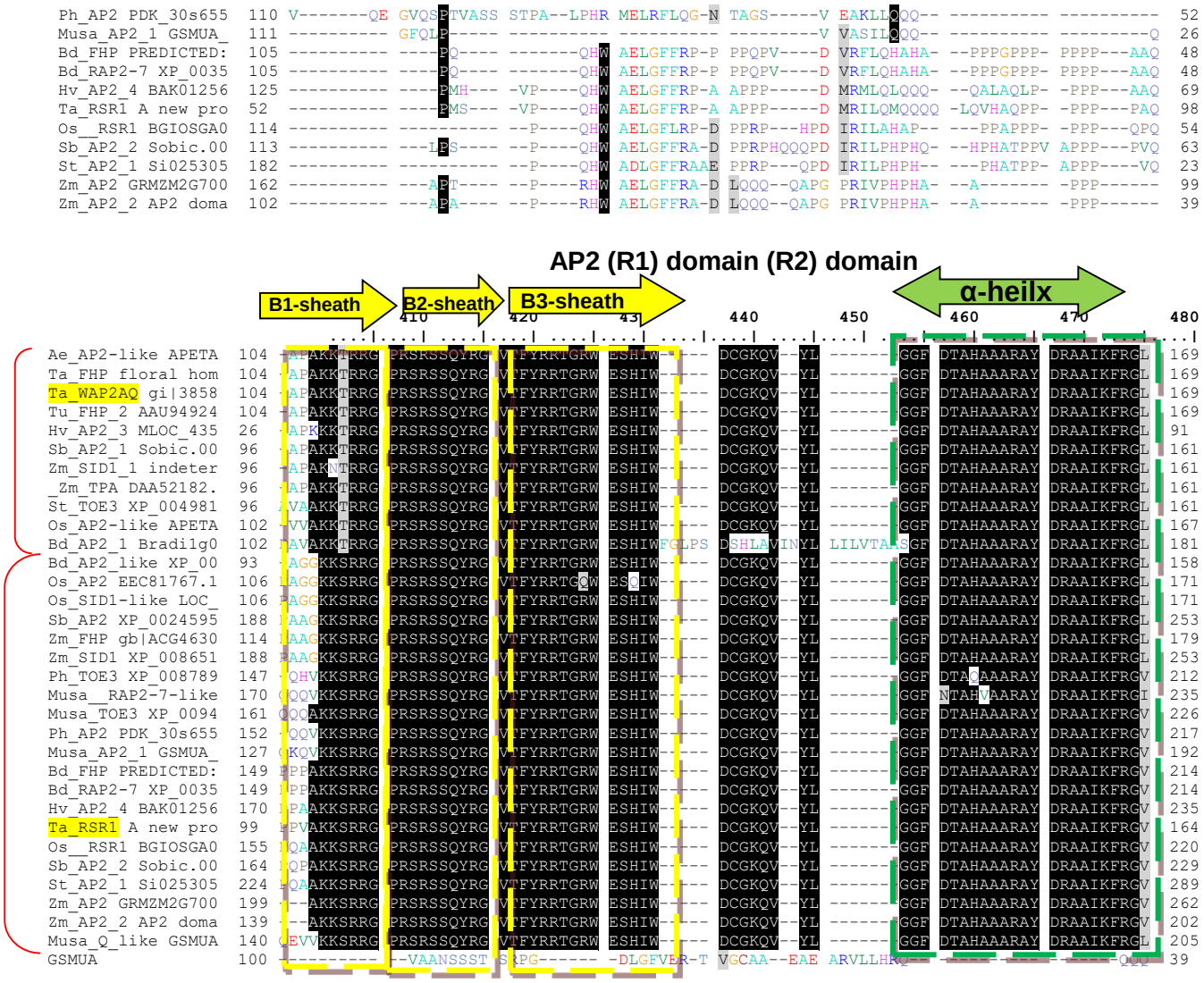
euAP2 family





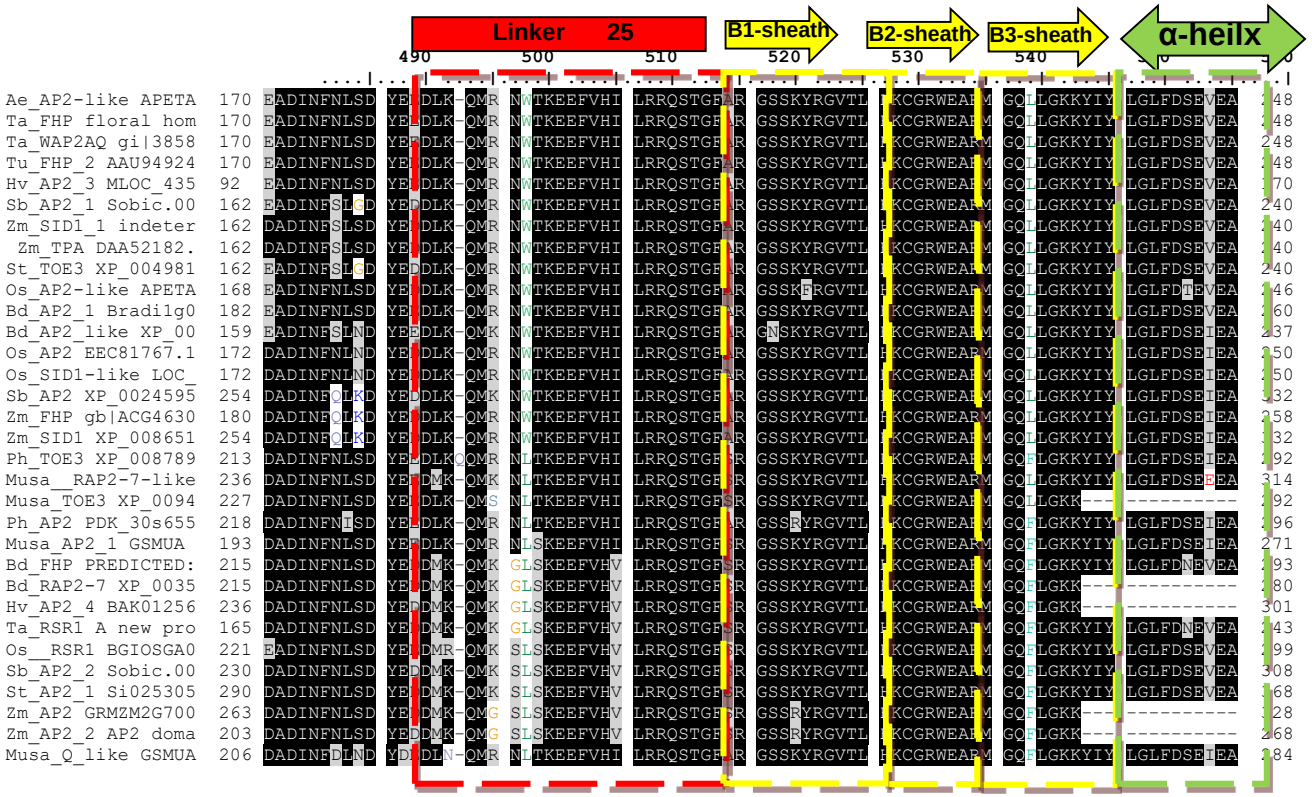
Continued on next page...



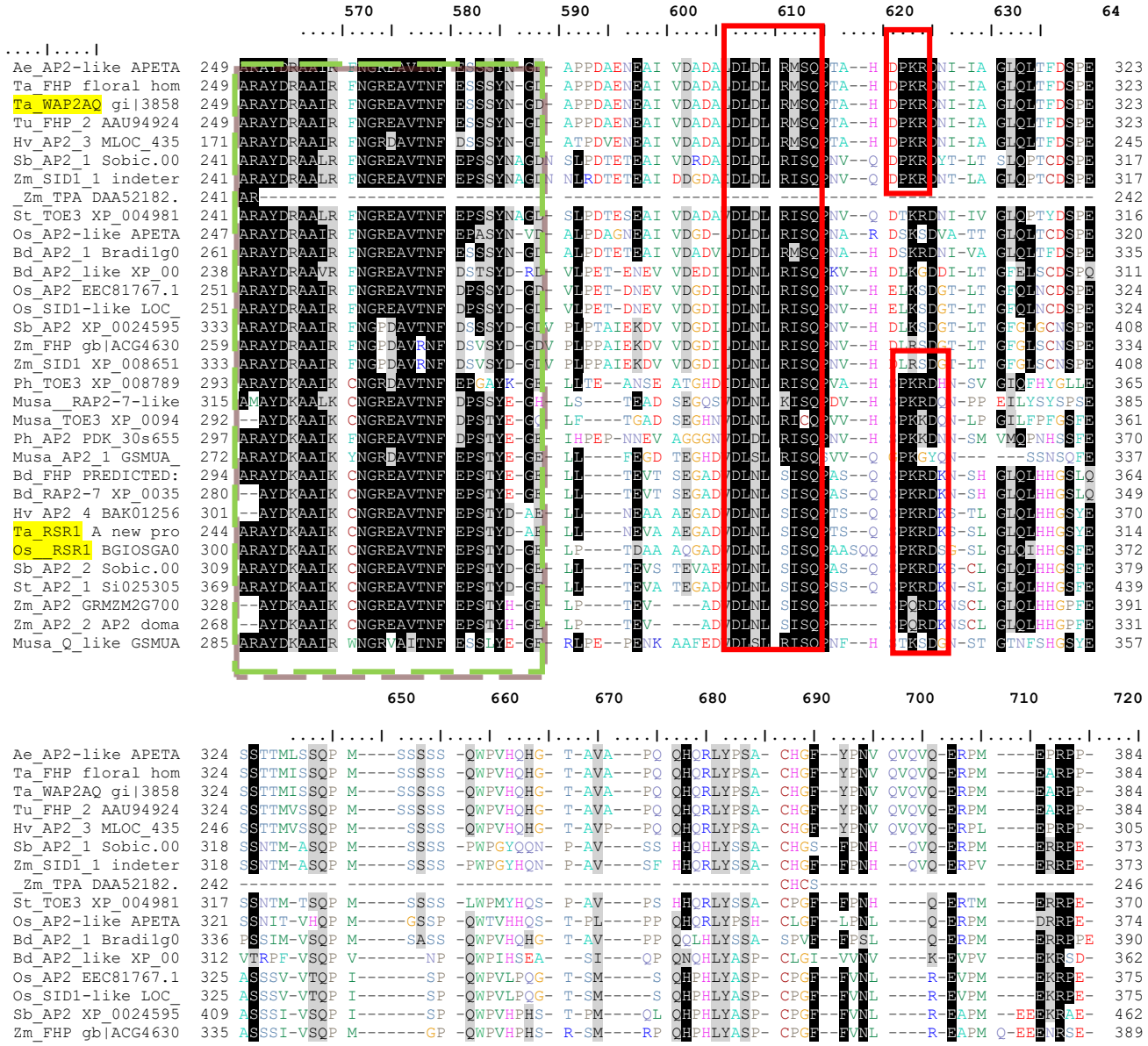


Continued on next page...

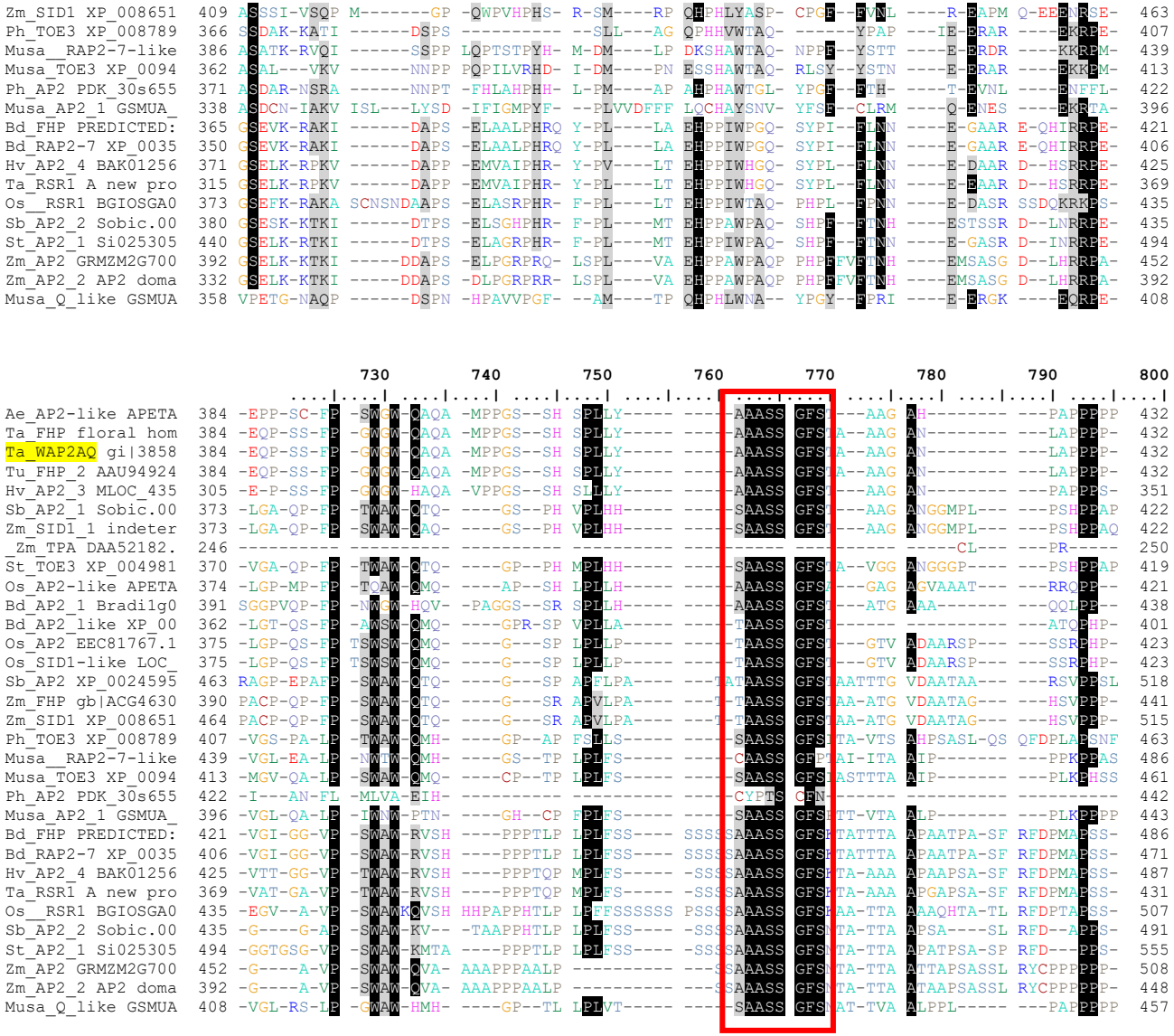
AP2 (R2) domain



Continued on next page...



Continued on next page...



Continued on next page...

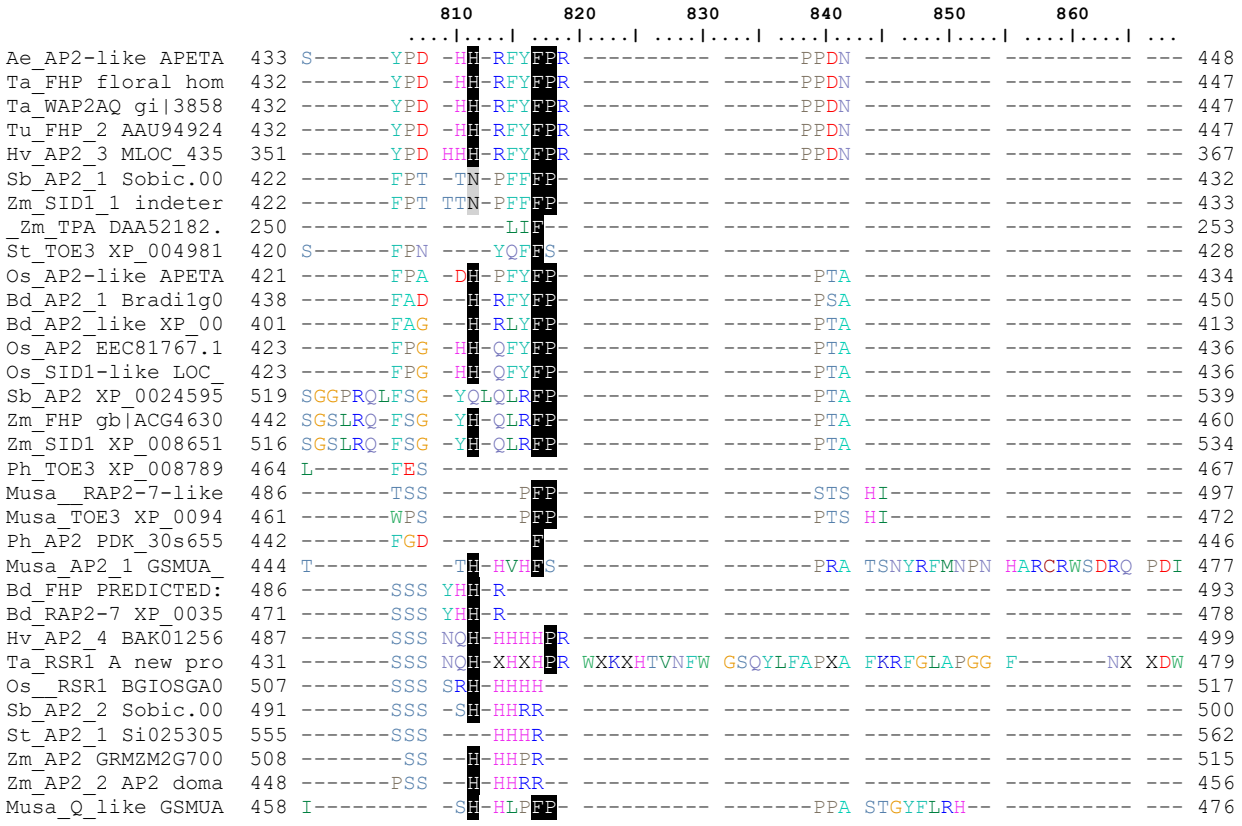


Figure 7.2.2: Alignment the whole sequences of proteins of euAP2 family, conserved motifs were boxed, showing two AP2 domains (AP2-R1)(AP2-R2).

Table 7.2.3: ID genes for bZIP.

Gene name	ID Gene	Species	No. of chromosome	No. of exons
Ae_BLZ1	EMT17240.1	*Aegilops tauschii*	-	6 exons
Ae_bZIP60	EMT18702.1	*Aegilops tauschii*	Chromosome 1	4 exons
At_BZ2	NM_127373.2	*Arabidopsis thaliana*	Chromosome 2	1 exon
At_bZIP9	A0A178UG19	*Arabidopsis thaliana*	Chromosome 5	6 exons
At_bZIP10_like	NM_123775.4	*Arabidopsis thaliana*	Chromosome 5:	1 exon
At_bZIP28	A0A178VDM1	*Arabidopsis thaliana*	Chromosome 3	2 exons
At_bZIP49	NP_191225.1	*Arabidopsis thaliana*	Chromosome 3	2 exons
At_BZO2H1	A0A1P8B7A6	*Arabidopsis thaliana*	Chromosome 4	6 exons
At_BZO2H3	NP_001031962.1	*Arabidopsis thaliana*	Chromosome 5	5 exons
Bd_CPRF2-like	XP_003557420.1	*Brachypodium distachyon*	Chromosome 1	6 exons
Bd-blz1	XM_003557372.3	*Brachypodium distachyon*	Chromosome 1	6 exons
Bd_opaque-2-like	XM_003557372.3	*Brachypodium distachyon*	Chromosome 1	6 exons
Bd_RF2a extraction	XP_003570405.1	*Brachypodium distachyon*	Chromosome 3	6 exons
Beta_bZIP1_like	XP_010679380.1	*Beta vulgaris subsp. vulgaris*		
Beta_CPRF2	XP_010674829.1	*Beta vulgaris subsp. vulgaris*		
Bna_bZIP	CDY12267.1	*Beta vulgaris subsp. vulgaris*		2 exons
Br_bZIP17_like	XP_013637012.1	*Brassica napus*		4 exons
Br_bZIP25	XP_009116138.1	*Brassica napus*		6 exons
Br_bZIP60	XP_013702826.1	*Brassica napus*		
Br_bZIP_like	BnaA02g32330D	*Brassica napus*		3 exons
Cen_O2	XP_004955660.1	*Cenchrus americanus*		
Coix_opaque_2	CAA55092.1	*Coix lacryma-jobi*		
Cu_CPRF2	XP_004139233.1	*Cucumis sativus*		
Gly_bZIP	NP_001237953.1	*Glycine max*		
Hv_Blz1	A0A287SLY8	*Hordeum vulgare subsp. vulgare*	Chromosome chr5H	6 exons
Hv_BLZ2	CAA71795.1	*Hordeum vulgare subsp. vulgare*	Chromosome chr1H	5 exons
Hv_bZIP_like	dbj|BAJ91812.1	*Hordeum vulgare subsp. vulgare*	Chromosome chr1H	5 exons
Hv_CRF2a_like	dbj|BAJ85496.1	*Hordeum vulgare subsp. vulgare*		
Musa_63-like	XP_009407738.1	*Musa acuminata subsp. malaccensis*	Chromosome 10	5 exons
Musa_bZIP1_like	XP_009418466.1	*Musa acuminata subsp. malaccensis*	Chromosome 9:	5 exons
Musa_bZIP9_like	XP_009400629.1	*Musa acuminata subsp. malaccensis*	Chromosome 5	6 exons
Musa_bZIP17_like	XP_009419845.1	*Musa acuminata subsp. malaccensis*	Chromosome 10:	3 exons
Musa_CPRF2-like	XP_009420671.1	*Musa acuminata subsp. malaccensis*	Chromosome 10	2 exons
Musa_CRF2_like	XP_009397697.1	*Musa acuminata subsp. malaccensis*		
Nt_BZI-1	NP_001312947.1	*Nicotiana tabacum*		
Nt_BZI_2	AAL27150.1	*Nicotiana tabacum*		
Nt_bZIP17	XP_009620672.1	*Nicotiana tabacum*		2 exons
Nt_CPRF2-like	XP_009787496.1	*Nicotiana tabacum*		2 exons
Nt_bZIP	NP_001312947.1	*Nicotiana tabacum*		
Os_O2_like	XP_006658337.1	*Oryza sativa Indica*	Chromosome 3	6 exons
Os_bZIP	BAA36492.1	*Oryza sativa Indica*	Chromosome 3	6 exons
Os_bZIP_9	XP_006646954.1	*Oryza sativa Indica*	Chromosome 1	8 exons
Os_CRF2_like	EAY92182.1	*Oryza sativa Indica*	Chromosome 3	6 exons
Os_REB	AAL10017.1	*Oryza sativa Indica*	Chromosome 3	6 exons
Os_RISBZ1	ACT31354.1	*Oryza sativa Indica*	Chromosome 3	6 exons
Os_RISBZ4	XP_004951741.1	*Oryza sativa Japonica*	Chromosome I	6 exons
St_RISBZ5	XP_004965642.1	*Setaria italica*	Chromosome IV	4 exons
Pc_CPRF2	XP_016537747.1	*Capsicum annuum*		
Ph_bZIP17_like	XP_008808577.1	*Phoenix dactylifera*		
Ph_CPRF2-like	XP_008786076.1	*Phoenix dactylifera*		
So_bZIP17_like	XP_006364780.1	*Sorghum bicolor*	Chromosome 10	2 exons

Continued on next page...

So_CPRF2-_1like	XP_006347748.1	*Sorghum bicolor*	Chromosome 1	6 exons
So_CPRF2	XM_002463695.1	*Sorghum bicolor*	Chromosome 1	6 exons
So_CPRF2-like	XP_002463740.1	*Sorghum bicolor*	Chromosome 1	6 exons
So_CPRF2	XP_009610430.1	*Sorghum bicolor*	Chromosome 1	6 exons
So_bZIP	XP_002463740.1	*Sorghum bicolor*	Chromosome 1	6 exons
So_bZIP_1	XP_002459432.1	*Sorghum bicolor*	Chromosome 2	6 exons
So_O2	A0A1B6Q9E3	*Setaria italica*	Chromosome 2	6 exons
St_17-like	XP_004958502.1	*Setaria italica*	Chromosome IV	4 exons
St_bzip9	K3YUA4	*Setaria italica*	Chromosome I	6 exons
St_CPRF2-like	K4A856	*Setaria italica*	Chromosome IX	7 exons
Ta_BLZ1	TraesCS5D01G447500	*Triticum aestivam*	Chromosome 5	6 exons
Ta_SPA	TraesCS1B01G343500.1	*Triticum aestivam*	Chromosome 1	6 exons
Tu_O2	TRIUR3_24233-T1	*Triticum urartu*		6 exons
Zm_bZIP1	NP_001105272.2	*Zea mays*	Chromosome 10	4 exons,
Zm_BZO2H2_1	NP_001131334.1	*Zea mays*	Chromosome 5	6 exons
Zm_BZO2H2_2	XP_008681056.1	*Zea mays*	Chromosome 5	6 exons,
Zm_CPRF2	NP_001150404.1	*Zea mays*	Chromosome 9	6 exons
Zm_CPRF2-like	XP_008665595.1	*Zea mays*	Chromosome 1	2 exons
Zm_O2	A0A1D6GFJ1	*Zea mays*	Chromosome 5	
Zm_OBF1	NP_001105439.2	*Zea mays*	Chromosome 3	1 exon

Table 7.2.4: ID genes for euAP2.

Gene name	ID Gene	Species	No. of chromosome	No. of exons
Ae_RSR1-like		Aegilops tauschii		
Ae-AP2-1like/Q	BAT00742.1	Aegilops tauschii	_____	10 exons
At_AP2-1	AT4G36920.2	*Arabidopsis thaliana*	Chromosome 4	10 exons
At_RAP2.12	AT1G53910.1	*Arabidopsis thaliana*	Chromosome 1	2 exons
At_RAV1	AT1G13260.1	*Arabidopsis thaliana*		
At_RAV2	AT1G68840	*Arabidopsis thaliana*		
At_SMZ	AT3G54990.1	*Arabidopsis thaliana*	Chromosome 3	7 exons
At_SNZ	AT2G39250.1	*Arabidopsis thaliana*	Chromosome 2	7 exons
At_TOE1	AT2G28550.1	*Arabidopsis thaliana*	Chromosome 2	7 exons
At_TOE2	AT5G60120.1	*Arabidopsis thaliana*	Chromosome 5	8 exons
At_TOE3	AT5G67180.1	*Arabidopsis thaliana*	Chromosome 5	9 exons
At -ANT	AT4G37750.1	*Arabidopsis thaliana*	Chromosome 4	8 exons
Bd_APETALA 2-like	BRADI1G53650.4	*Brachypodium distachyon*	Chromosome 1	10 exons
Bradi_AP2-2	BRADI1G02575.1	*Brachypodium distachyon*	Chromosome 1	9 exons
Bradi_AP2-3	BRADI2G37800	*Brachypodium distachyon*	Chromosome 2	10 exons,
Bradi_AP2-4	BRADI1G30337.1	*Brachypodium distachyon*	Chromosome 1	8 exons
Bradi_SHAT1-like	BRADI5G24100.1	*Brachypodium distachyon*	Chromosome 5	10 exons
Bradi_SID1-like	BRADI1G53650.4	*Brachypodium distachyon*	Chromosome 1	10 exons
Bradi RAP2.12-like	BRADI1G46690.3	*Brachypodium distachyon*	Chromosome 1	3 exons
Bradi RSR1-Like	BRADI2G37800.1	*Brassica napus*	Chromosome 2	10 exons
Brasrapa_SNZ-like	Bra007123.1	*Brassica napus*	Chromosome A09	7 exons
Brasrapa_TOE1.1-like	Bra000487.1	*Brassica napus*	Chromosome A03	9 exons
Brasrapa_TOE1-like	Bra011939.1	*Brassica napus*	Chromosome A07	10 exons
Brasrapa_TOE2.1-like	Bra020262.1	*Brassica napus*	Chromosome A02	8 exons
Brasrapa_TOE2-like	Bra002510.1	*Brassica napus*	Chromosome A10	9 exons
Brasrapa_TOE3-like	Bra012139.1	*Brassica napus*	Chromosome A07	9 exons
Brasrapa ANT.1-like	Bra010610.1	*Brassica napus*	Chromosome A08	7 exons
Brasrapa ANT.2-like	Bra011782.1	*Brassica napus*	Chromosome A01	7 exons
Brasrapa ANT-like	Bra017852.1	*Brassica napus*	Chromosome A03	7 exons
Brasrapa AP2-1	Bra011741.1	*Brassica napus*	Chromosome A01	10 exons
Brasrapa AP2-2	Bra017809.1	*Brassica napus*	Chromosome A03	10 exons
Brasrapa SMZ-like	SMZ Bra007123	*Brassica napus*	Chromosome A09	7 exons
Brasrapa TOE1-like	Bra011939.1	*Brassica napus*	Chromosome A07	10 exons
Brasrapa TOE2.1-like	Bra020262.1	*Brassica napus*	Chromosome A02	8 exons
Brasrapa TOE2-like	Bra002510.1	*Brassica napus*	Chromosome A10	9 exons
Brasrapa TOE3-like	Bra012139	*Brassica napus*	Chromosome A07	9 exons
GlyMax_SMZ-like	GLYMA12G07800.4	*SoyBase*	Chromosome 12	10 exons
GlyMax_TOE1-like	GLYMA13G40470.1	*SoyBase*	Chromosome 13	10 exons
Gm SMZ-like	Glyma13g40470.6	*SoyBase*		
Hv_AP2-1	HORVU2Hr1G113880.12	Hordeum vulgare	Chromosome chr2H	6 exons
Hv_AP2-3	HORVU7Hr1G116220.4	Hordeum vulgare	Chromosome chr7H	9 exons,
Hv_APETALA2-like	HORVU5Hr1G112440.10	Hordeum vulgare	Chromosome chr5H	7 exons,
Hv_Q	HORVU5Hr1G112440.10	Hordeum vulgare	Chromosome chr5H	
Hv_RSR1-like	HORVU1Hr1G011800.3	Hordeum vulgare	Chromosome chr1H	7 exons
Os-Q-like	P0407H12.124 OS07T0235800-01	Oryza sativa Japonica	Chromosome 7	10 exons,
Os_GLOSSY15-like	AP2D22 OS06T0639200-01	Oryza sativa Japonica	Chromosome 6	10 exons
Os_SHAT1	AP2D23 OS05T0121600-01	Oryza sativa Japonica	Chromosome 5	8 exons
Os_SID1-like	P0407H12.124 OS07T0235800-01	Oryza sativa Japonica	Chromosome 7	10 exons,
Os_supernumerary bract	OS07G0235800	Oryza sativa Japonica	Chromosome 7	10 exons
Os_TOE3	P0407H12.124 OS07T0235800-01	*Oryza sativa Japonica*	Chromosome 7	10 exons
Os AP2-2	P0407H12.124 OS07T0235800-01	*Oryza sativa Japonica*	Chromosome 7	10 exons,
Os_RSR1	AP2D23 OS05T0121600-01	*Oryza sativa Japonica*	Chromosome 5	8 exons

Continued on next page...

Os SID1-like	OS07G0235800	*Oryza sativa Japonica*	Chromosome 7	10 exons,
PD_SID-like TOE3	XP_017697250.1	*Phoenix dactylifera*		
Sb_RSR1-like	KXG21168	*Sorghum bicolor*	Chromosome 9	10 exons,
Sb_SID1-like	EER93202		Chromosome 2	10 exons
Sb AP2-2	EER93202	*Sorghum bicolor*	Chromosome 1	10 exons,
Sb GLOSSY15-like	KXG20453	*Sorghum bicolor*	Chromosome 10	10 exons
Sb RSR1-like	SORBI_009G024600	*Sorghum bicolor*	Chromosome 9	9 exons
Sb SID1-like	EER96101	*Sorghum bicolor*	Chromosome 2:	10 exons,
Si_AP2-2	Si036068m		Chromosome 9	9 exons
Si_AP2-4	Si025305m		Chromosome 3	11 exons
Si_GLOSSY15-like	Si012635m	*Setaria italica*	Chromosome 7	10 exons
Si_SHAT1-like	Si012635m	*Setaria italica*	Chromosome 7	10 exons
Si_SID1-like	Si029636m		Chromosome 2	10 exons
SL_AP2a	Solyc03g044300.2.1	*Setaria italica*	Chromosome 3	9 exons
SL_AP2c.1	Solyc02g064960.2.1	*Solanum_lycopersicum*	Chromosome 2	10 exons
SL_SNZ-like	Solyc09g007260.2.1	*Solanum_lycopersicum*	Chromosome 9	9 exons,
SL ANT	Solyc04g077490.2.1	*Solanum_lycopersicum*	Chromosome 4	9 exons,
St_AP2c.1-like	PGSC0003DMP400007138	*Setaria italica*		
St_AP2c.2-like PHAP2A	PGSC0003DMT400065503	*Solanum tuberosum*	Chromosome 2	10 exons
St_AP2c.3-like	PGSC0003DMT400065501	*Solanum tuberosum*	Chromosome 2	10 exons
St_AP2c-like	PGSC0003DMT400010229	*Solanum tuberosum*	Chromosome 2	10 exons
St ANT-like	PGSC0003DMT400012588	*Solanum tuberosum*	Chromosome 4	9 exons
St AP2	PGSC0003DMT400016584	*Solanum tuberosum*	Chromosome 3	10 exons
St AP2-4 Relative to APETALA2 1	PGSC0003DMG400027904	*Solanum tuberosum*	Chromosome 9	7 exons
St SNZ-like RAP2-7-like	XP_006366410.1 PGSC0003DMT400065311	*Solanum tuberosum*	Chromosome 11	9 exons
Ta_AP2-1	Tae002676	*Triticum aestivum*		
Ta_AP2-2	Tae023554	*Triticum aestivum*	Chromosome 5A	10 exons
Ta_Q	TRIAE_CS42_5BL_TGACv1_408657_AA1364080.1	*Triticum aestivum*	5BL	6 exons
Ta AP2-1 TaAP2-B	TRIAE_CS42_2BL_TGACv1_129463_AA0384870.01	*Triticum aestivum*	2BL	10 exons,
Ta AP2-2 (RAP2-7)	TRIAE_CS42_5BL_TGACv1_408657_AA1364080	*Triticum aestivum*	5BL	8 exons
Ta RSR1	TRIAE_CS42_1BS_TGACv1_049682_AA0159520	*Triticum aestivum*	1BS	8 exons
Zm_AP2-3	GRMZM5G862109_P01	*Zay mays*		
Zm_AP2-4	GRMZM5G862109_P02	*Zay mays*		
Zm_AP2-5	Zm00001d034629_T002	*Zay mays*		
Zm_GLOSSY15	Zm00001d046621_T001	*Zay mays*	Chrom 9	9 exons,
Zm_indeterminate spikelet 1	Zm00001d034629_T003	*Zay mays*		9 exons
Zm_SID1 tasselseed6	Zm00001d034629_T004	*Zay mays*	Chromosome 1	9 exons
Zm AP2-3	Zm00001d034629_T002	*Zay mays*	Chromosome 1	
Zm AP2-4	Zm00001d034629_T004	*Zay mays*	Chromosome 1	9 exons
Zm AP2-5 (rap2.7)	Zm00001d010987_T001	*Zay mays*	Chromosome 8:	
Zm SHAT1-like	Zm00001d010987_T001	*Zay mays*		

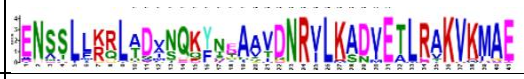
Appendix. 7.2.5: Comparison all members of euAP2 in the miRNA 172 binding sites.

The type of miRNA	Gene ID	miRNA sequence	miRNA binding site	miRNA_aligned_fragment	Target_aligned_fragment
miR172d	JX473823.2	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCCAAAA	Cleavage	Starch negative regulator RSR1 wheat
miR172d	AK370055.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCCAAAA	Cleavage	Starch negative regulator RSR1 Horduem vulgare
miR172d	KT439184.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCCAAAA	Cleavage	Triticum aestivum cultivar Cadenza target of EAT1-B1 (TOE1-B1) gene
miR172d	XM_014899606.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCCAAAA	Cleavage	Brachypodium distachyon floral homeotic protein APETALA 2
miR172d	XM_015783844.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACG	Cleavage	Oryza sativa Japonica Group AP2-like ethylene-responsive transcription factor TOE3
miR172d	XM_008658315.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCG	Cleavage	Zea mays floral homeotic protein APETALA 2
miR172d	AB697000.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTTTCTAC	Cleavage	Triticum aestivum Q mRNA for transcription factor WAP2AQ
miR172d	AB697001.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTAC	Cleavage	Triticum spelta var. duhamelianum q mRNA for transcription factor WAP2Aq
miR172d	AY702957.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCG	Cleavage	Triticum turgidum subsp. dicoccoides floral homeotic protein (Q) gene
miR172d	AK355002.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTAC	Cleavage	Hordeum vulgare subsp. vulgare mRNA for Floral homeotic protein APETALA 2
miR172d	XM_003559195.3	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCT	Cleavage	Brachypodium distachyon floral homeotic protein APETALA 2-like
miR172d	KT439185.1	AGAAUCCUGAUGAUGCUGCAG	GCTGCAGCATCATCAGGATTCTCCA	Cleavage	Triticum aestivum cultivar Charger target of EAT1-B1 (TOE1-B1) gene
miR172d	NM_001111434.2	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACC	Cleavage	Zea mays indeterminate spikelet 1 (ids1)
miR172d	KJ727838.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACC	Cleavage	Zea mays clone pUT5791 AP2-EREBP transcription factor (EREB11)
miR172d	XM_015772883.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTGC	Cleavage	Oryza sativa Japonica Group AP2-like ethylene-responsive transcription factor TOE3
miR172a	AY714877.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCCCACTGG	Cleavage	Zea mays GLOSSY15 (gl15) gene
miR172d	EU551085.1	AGAAUCCUGAUGAUGCUGCAG	GCAGCATCATCAGGATTCTCTCCTCAAC	Cleavage	Arabidopsis thaliana isolate CS6626 TOE3 (TOE3) gene
miR172d	EU551038.1	AGAAUCCUGAUGAUGCUGCAG	GCAGCATCATCAGGATTCTCTCCTCAAC	Cleavage	Arabidopsis thaliana isolate CS6626 TOE1 (TOE1) gene
miR172d	XM_009396259.2	AGAAUCCUGAUGAUGCUGCAG	TCTGCAGCATCATCAGGATTCGCAAGAC	Cleavage	Musa acuminata subsp. malaccensis AP2-like ethylene-responsive transcription factor TOE3

Continued on next page...

miR172d	XM_008652800.2	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCGCC	Cleavage	Zea mays sister of indeterminate spikelet 1 (sid1)
miR172d	EU974186.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCGCCGCTGCC	Cleavage	Zea mays floral homeotic protein mRNA
miR172d	XM_004955796.2	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCGCCG	Cleavage	Setaria italica floral homeotic protein APETALA 2
miR172d	AY069953.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCGCCG	Cleavage	Hordeum vulgare APETALA2-like protein (AP2L1) mRNA
miR172d	EU974186.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACCGCCGC	Cleavage	Zea mays floral homeotic
miR172d	XM_009403966.2	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACTGCCACC	Cleavage	Musa acuminata subsp. malaccensis floral homeotic protein APETALA 2-like
miR172d	XM_002466159.2	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCGACCGC	Cleavage	Sorghum bicolor floral homeotic protein APETALA 2
miR172d	XR_001547818.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCTACGGG	Cleavage	Oryza sativa Japonica TOE3
miR172d	XM_015837726.1	AGAAUCCUGAUGAUGCUGCAG	CTGCAGCATCATCAGGATTCTCCA	Cleavage	Oryza brachyantha floral homeotic protein APETALA 2

Table 7.2.6: Conserved motifs in O2 family by using MEME.

	Sequence logo	Value	size	width	Function
1		4.6e-1274	59	3	bZIP domain (basic region)
2		4.3e-1589	54	4	bZIPdomain(leucine zipper)
3		3.0e-672	41	2	
4		1.2e-424	51	1	
5		1.9e-423	25	2	
6		1.6e-234	37	1	Phosphorlation sites
7		2.3e-322	34	21	
8		1.1e-182	32	11	
9		1.5e-168	38	15	
10		3.9e-148	25	15	
11		2.5e-250	15	41	
12		1.7e-174	18	29	
13		1.4e-104	8	41	
14		2.0e-083	25	11	
15		1.2e-066	8	29	
16		6.9e-063	14	16	
17		2.6e-054	7	28	
18		1.8e-051	12	21	

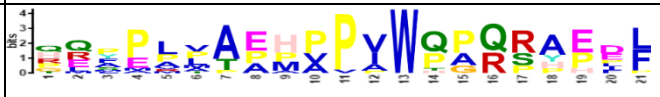
Continued on next page…

19		4.4e-050	10	21	
20		5.1-067	5	39	Phosphorlation sites
21		2.0e-049	12	15	
22		4.0e-048	4	41	
23		7.9e-037	4	29	
24		8.9e-034	3	40	
25		2.2e-028	6	15	

Table 7.2.7: 25 conserved Motifs of euAP2 proteins were identified by using MEME.

	Sequence logo	E-value	Size	Width	Annotation
1		2.9 3.274	61	50	AP2 domain
2		1.0 - 3090	62	50	AP2 domain
3		8.9-1520	53	41	AP2 domain
4		5.3e - 1072	62	21	
5		2.5e-652	61	21	
6		4.3e - 348	59	15	LxLxLxLPP, consereved motif in the C-terminal regions.
7		3.9e-335	59	21	MiRNA binding site
8		5.0e - 279	59	15	
9		1.1e-186	36	15	function as an activation domain (Jofuku et al. 1994). It is found in Q and some dicots
10		9.1e-187	23	21	
11		5.2e-280	35	29	

Continued on next page...

12		3.3e-163		15	
13		1.1e-151	55	8	
14		1.6e-158	43	15	
15		1.3e-092	8	21	
16		2.3e-064	21	21	
17		1,5e-082	17	21	
18		5	30		
19		8.1e-049	13	15	
20		3.1e-047	12	11	
21		1.5e-082	17	21	
22		2.1e-036	13	14	
23		2.8e-035	4	27	
24		6.2e-030	3	41	
25		4.0e-028	5	21	

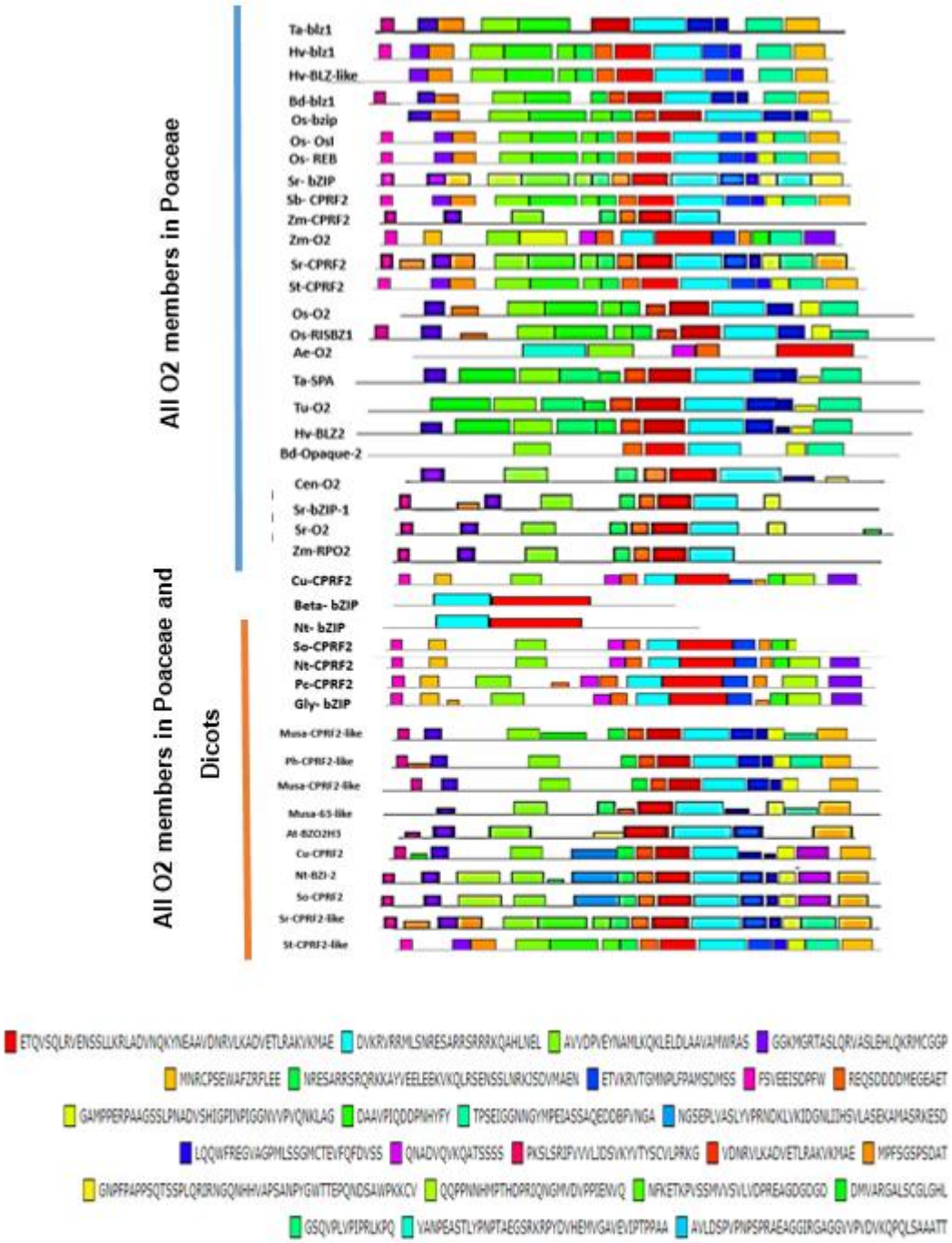
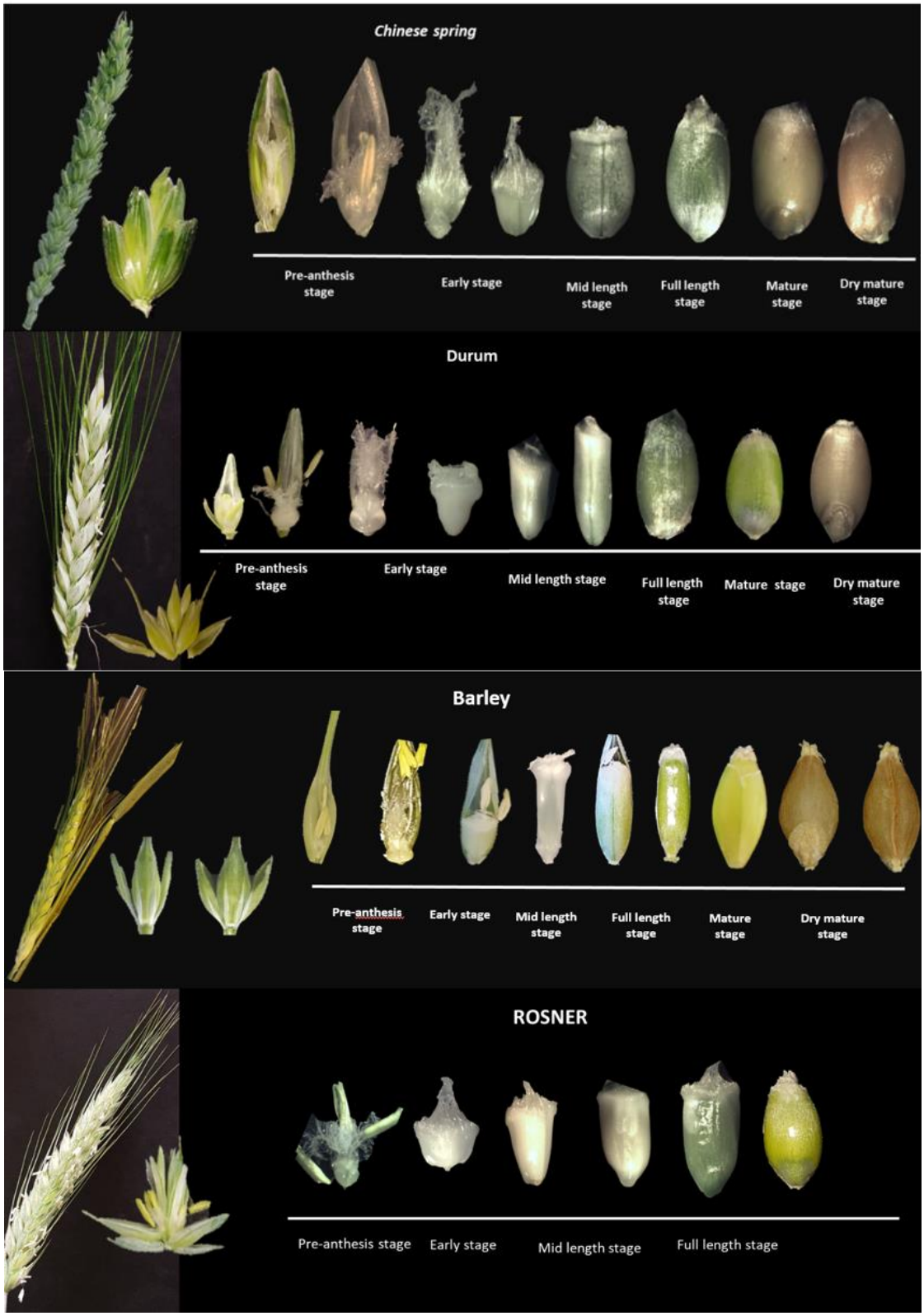


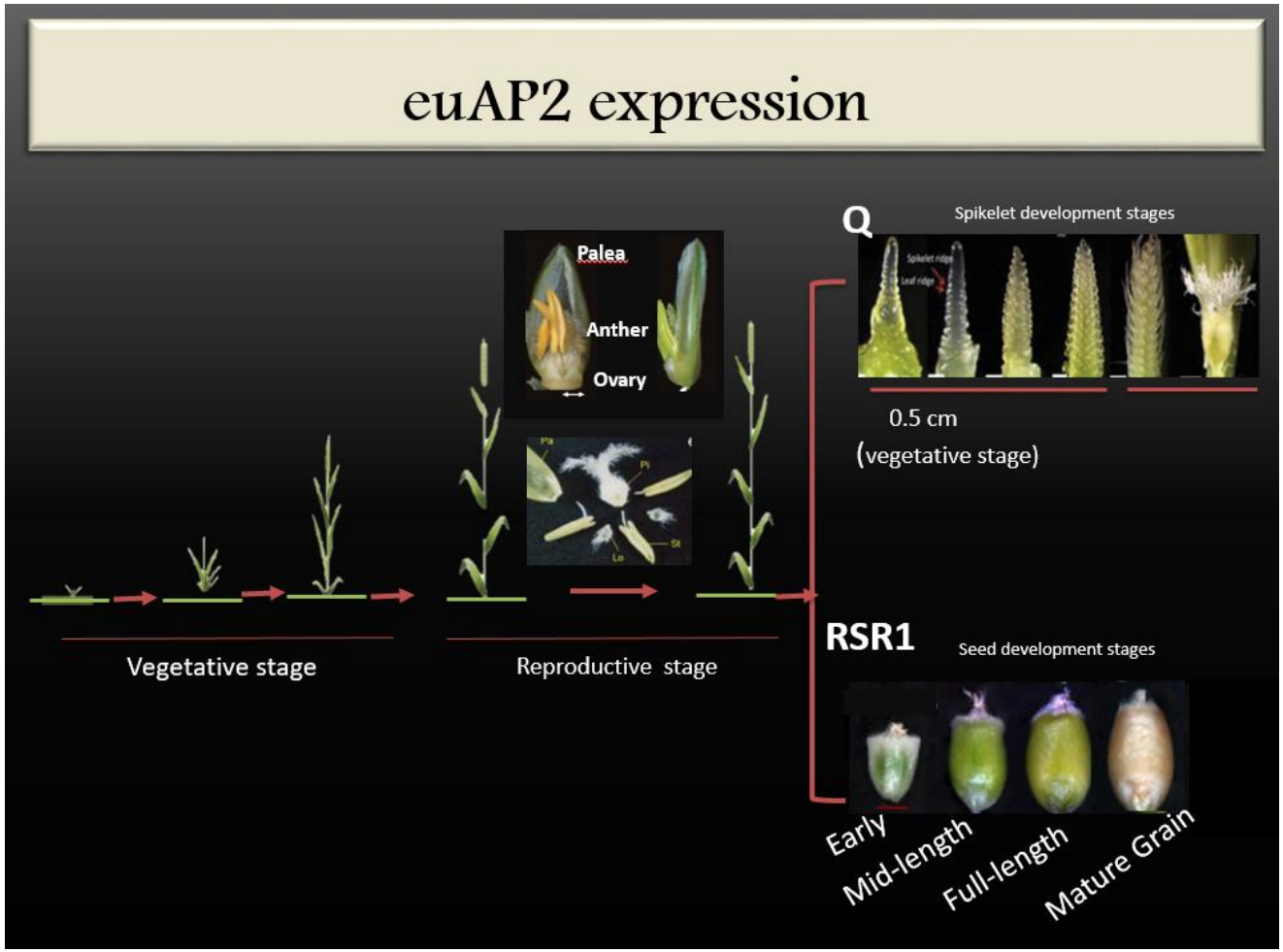
Figure 7.2.8: Distribution and position of consereved motifs along the O2 by using MEME.

Appendices: Supplementary Data Chapter 3

7.3.1 Materials which were used for RNA extraction.



Appendix 7.3.1: Whole tissues which used for RNA extraction in this project from Chinese spring, Durum, barley and rosner.

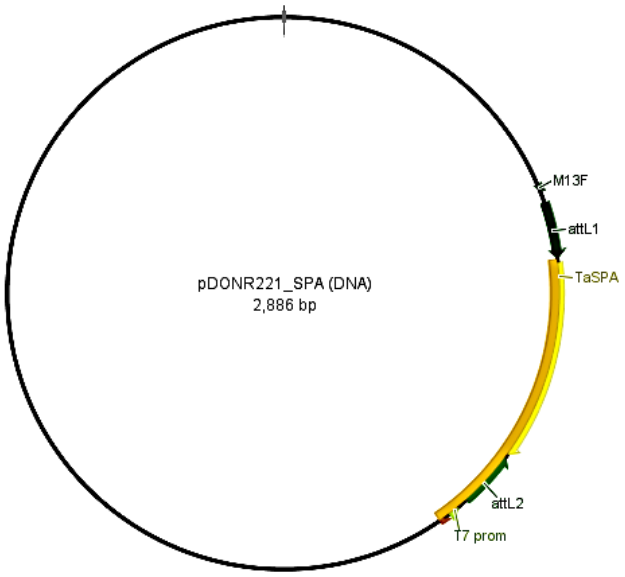


Appendix 7.3.2: Sequence of primers used

RT-PCR	
TaBLZ1/F	ACATGCCTGAGATGGCTTCC
TaBLZ1/R	AAAGTACCCGGCAGCAATGA
Ta-Spa/F	TTAGCTCTTCCTCATTGGAGCC
Ta-Spa/R	GTAACCCTCTTCATAGTTTCC
TaRSR1/F	CATCATCATCGTCAAGCAACC
TaRSR1/R	CTGCTGAATTTCAGTGACAAG
Ta-Q/F	CGCAGTACGGGGTCACCTTC
Ta-Q/R	TACGCTCTTGCAGCTTCAAC
Ta-GAPDH/F	CGCTAGCTGCACCACTAACT
Ta-GAPDH/R	GATCATAGGTTGCTGGCTTC
Hv-BLZ1/F	TCCGTTGTTACCTCCAGCCT
Hv-BLZ1/R	CTGTTAAGGCGGGCTTGCTGA
Hv-BLZ2/F	GGAAGGGGAGGCACAAAACAA
Hv-BLZ2/R	CAAGCATGGCATAGGTTGGG
HvRSR1/F	GTCGTCAAGCAACCAACACC
HvRSR1/R	CGATGCAGAGAGCGCAAAAA
Hv-Actin/F	ACTACAACTCCAGAAGGAG
Hv-Actin/R	TCGCGATAGCTAAAAGAGCA
ScRSR1/F	CATCATCGTCGTCAAGCAACC
ScRSR1/R	GAGGTTAGTATCTGACAGCG
Hv-Q/F	AGATGAGGAACTGGACCAAGG
Hv-Q/R	GCGGCGGTAGAAATCCTGA
poly (dT)	GACTCGAGTCGACATCGA
qRT-PCR	
BLZ1	GCCAGGAAAGTCCCAAGAA
	CTAGAGGTCACAGCCAAAG
Ta-Spa	ATTCCTACATCAGCACCTACTC
	CTGCTCAGGTCTCCCTAA
Ta SPA/IBL	CCAAGCAATGTCTAGCATATCATC
	ACCTTCGCTACATATGCAACCT
Ta SPA/IAL	GTCATGCCTGGATCATCACCCGC
	AACAAGCCGAGAAACCACTATA
Ta SPA/1DL	TCCGCAACGGGAATCTACTA
	AATGTCCACTTAACAAGCCAAACT
Hv-BLZ2	GGAAGGGAGGGAACCAA
	CAAGCATGGCATAGGTTGC
RSR1	GCCTCATTCCGGTTC
	GAGATGAGCTGAGGTTAGT
Q	GCTCGGGAAGAAGTACAT
	CGCAGACCAAGTCA
Ta-GAPDH	CGTAACCCAAAATGCCCTTG
	TTCAACATCATTCCAAGCAGC
Ta-Actin	AACGAAGGATAGCATG
	AGAAGGTGTGACCAAAT

Hv-GAPDH	GTTGTTCTCAGTGGTGGTTCTA
	TCCTTTCAGGTGGTGCAATAA
Hv-Actin	ACCTTCAGTTCCCAGCAAT
	CAGAGTGGAGCACAATACC
mRNA.Ins-hybridization	
Ta-RSR1 RT7	GAATTGTAATACGACTCACTATAGGGCTGCTGAATTTCAGTGACAAG
Sc-RSR1 RT7	GAATTGTAATACGACTCACTATAGGGGAGGTTAGTATCtGACAGCG
Ta-BLZ1 RT7	GAATTGTAATACGACTCACTATAGGGGAAAGTACCCGGCAGCAATGA
Ta-Spa RT7	GAATTGTAATACGACTCACTATAGGGGGTAACCCTCTTCATAGTTTCC
Ta-Q RT7	GAATTGTAATACGACTCACTATAGGGGCGGCGGTAGAAATCCTGA
Hv-QRT7	GAATTGTAATACGACTCACTATAGGGGGCGGCGGTAGAAATCCTGA
TaH4 F	CATCTCGGGGCTCATCTACG
TaH4 RT7	GAATTGTAATACGACTCACTATAGGGTGCGATTAACTTTGTTCCACGA
TaH3 F	AGGAGTGAAGAAGCCTCACC
TaH3 RT7	GAATTGTAATACGACTCACTATAGGGATTCCCACAGGCAAAGCAAG

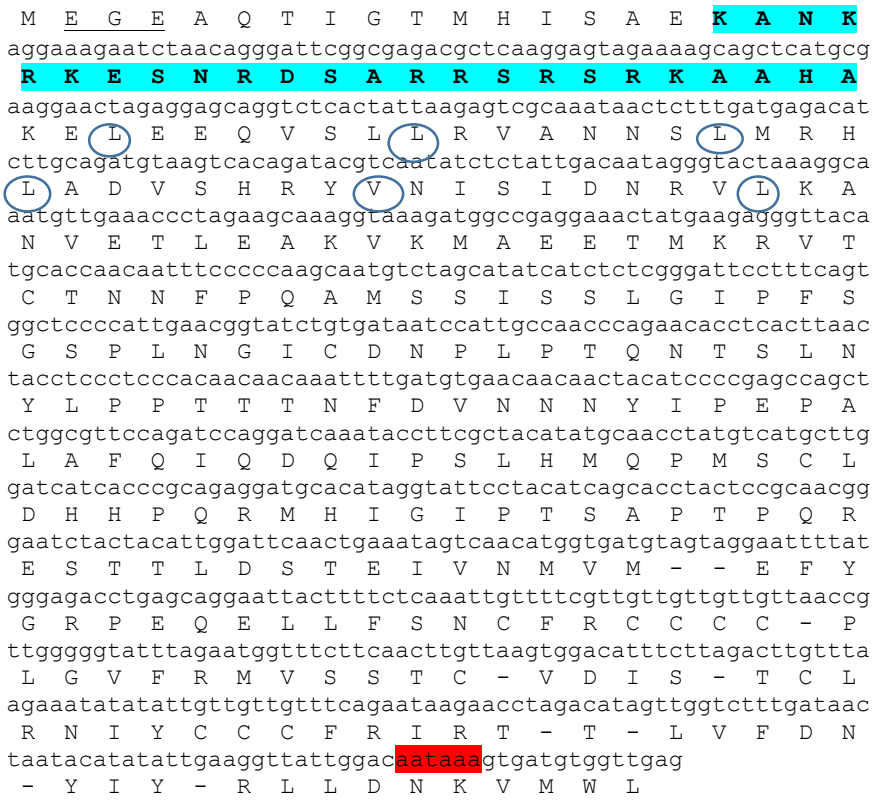
Appendix 7: Dta chapter 4



Appendix 7.4.1: The fragment of SPA which was used for RNAi line to insert in entry clone (Gateway cloning).

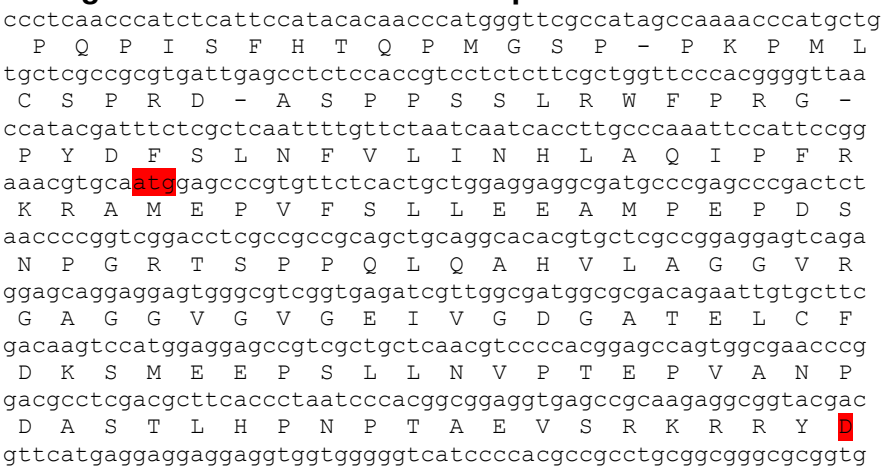
Ta –SPA (B)

acagaaaaaccttgtgcagtagtagtactactgaaataaaatagacatgagaggatcgca
T E K P C A V V V L L K - N R H E R I A
ccctcaacccatctcattcagcacacaacccatgggttagccatagccaaaacccgtggt
P S T H L I Q H T T H G L A I A K T R G
gtgtcacccgcgtcattgggcatctccgtcctctcttcgccggttcccacggggttaaaa
V S P A S L G I S V L S S P V P T G L K
agtttaagccataccattcctcgcccaattttgttctaatcaatcacctttgccccagtt
S L S H T I P R P I L F - S I T F A P V
ccatttcggaaagcttccatggagcccgtgttcttctcactggaggaggcgatgcccgag
P F R K A S M E P V F F S L E E A M P E
cccgactctaacccctgccggacctcgtcgccgccgctggaggcacacatgctcgtcgca
P D S N P C R T S S P P L E A H M L V A
ggactcggaggagtgggcgccggcgaggtcgtcggcgggtgcgcgacgaacgagtgcgcg
G L G G V G A G E V V G G C A T N E C A
acagaatggtgcttccagaagttcgtggacgagccgtggctgctcaacgtccccaccgcg
T E W C F Q K F V D E P W L L N V P T A
ccagtggcgaaccccgaagcttcgacgctttaccctaatcccacggccgaggggagccgc
P V A N P E A S T L Y P N P T A E G S R
aagaggccgtacgacgtccatgagatggtgggcccggaggaggtcatccccacgccgcct
K R P Y D V H E M V G P E E V I P T P P
gcggcgagcccggtggtggaccccgtggcgtacaacgcgatgctcagacggaagttggac
A A S P V V D P V A Y N A M L R R K L D
gcgcatctcgccgccgtcgccatgttgaggaccactcggggaatttgcccacaaagctcc
A H L A A V A M L R T T R G I C P Q S S
catgacaatggagcatcgcaaaattcagattccatccaaggctcagaaaaccacaccgga
H D N G A S Q N S D S I Q G S E N H T G
gatgtcagtttgcatcaacttagctcttcctcattggagccatcaccatcagatggtgat
D V S L H Q L S S S S L E P S P S D G D
atggaaggggaggcacaaacaattggaactatgcatattagtgcagagaaagcgaataag



Appendix 7.4.2. **Ta-SPA nucleotide sequence (cDNA) and amino acid sequence. Amino acid residues of the basic DNA binding domain are in blue, the leucine heptad repeats are circled, and a presumptive serine-rich phosphorylation site is highlighted in red. Acidic amino acid residues, putatively involved in activation, are double underlined. The start codon and stop codon sequences (ATG, AATAAA respectively.**

Hv-BLZ2
H.vulgare mRNA for bZIP transcription factor 2



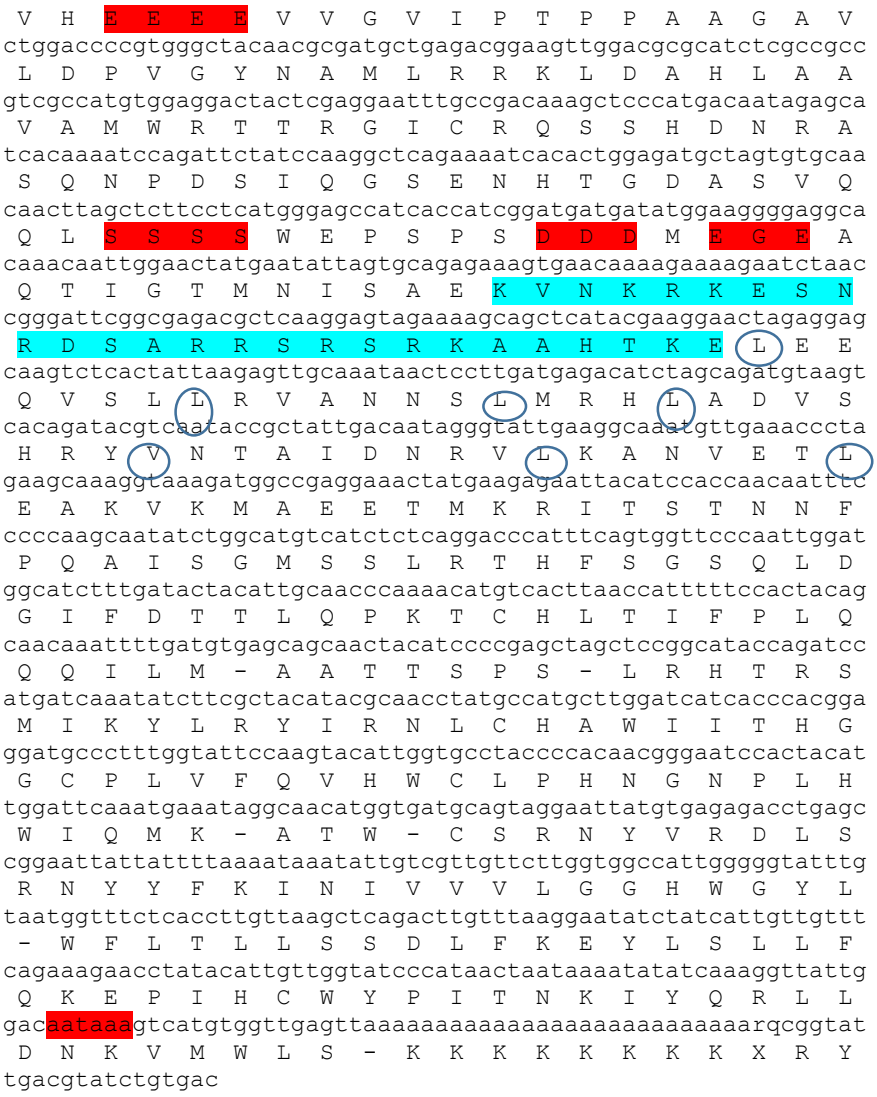


Figure 7.4.3. **Nucleotide sequence and deduced amino acid sequence of the Blz2 cDNA. Amino acid residues of the basic DNA binding domain are in blue, the leucine heptad repeats are circled, and a serine-rich phosphorylation site is boxed. Acidic amino acid residues, putatively involved in activation, are double underlined. Nucleotide sequence numbers refer to the ATG translation initiation codon. The stop codon is in red.**

Appendix 7.4.4. Gene constructs generated in this project for functional analyses.

Experiment	Construct	Plasmid/ vector	Strain	Gene	Cloning site	Insert size	Promoter	Antibiotics/ amino acid resistance
RNAi								
Entry clone	TaSPA_pDNR221	pDNR221		Ta-SPA	attB+attP=>attL	340 bp	T7 promoter	Kanamycin (50 µg/ml)
Destination vector	Ta-SPA- pDNR	pDNR		Ta-SPA	attB+attP=>attL	340 bp	Rice- Actin	
Yeast two hybrids								
Bait Clone	TaSPA_pGADT7	pGBKT7	Y2HGold	TaSPA	EcoR1-BamH1	1500 bp	GAL4 DNA-BD	Kanamycin3 (50 µg/ml), Leucine4*
Bait Clone	TaSPA_pGBDT7	pGBKT7	Y2HGold	TaSPA	EcoR1-BamH1	500 bp	GAL4 BD	Kanamycin3 (50 µg/ml), Leucine4*
Prey Clone	TaSPA_pGADT7	pGADT7	Y187	TaSPA	EcoR1-BamH1		GAL4 AD	
Prey Clone	TaBLZ1_pGADT7	pGADT7	Y187	Ta-BLZ1	EcoR1-BamH1	1800 bp	GAL4 AD	Ampicilin1 (100 µg/ml), Tryptophan
	TaBLZ1_pGADT7	pGADT7	Y187	Ta-BLZ1	EcoR1-BamH1	600 bp	GAL4 AD	Ampicilin1 (100 µg/ml), Tryptophan
	TaBLZ1_pGADT7	pGADT7	Y187	Ta-BLZ1	EcoR1-BamH1	320 bp	GAL4 AD	Ampicilin1 (100 µg/ml), Tryptophan
Bait Clone	TaPBF_pGADT7	pGBKT7	Y2HGold	TaPBF	EcoR1-BamH1	600 bp	GAL4 DNA-BD	Kanamycin3 (50 µg/ml), Leucine4*
Prey Clone	TaPBF_pGBDT7	pGADT7	Y187	TaPBF	EcoR1-BamH1	600 bp	GAL4 AD	Ampicilin1 (100 µg/ml), Tryptophan
Bait Clone	TaRSR1_pGADT7	pGBKT7	Y2HGold	TaRSR1	EcoR1-BamH1	500 bp	GAL4 DNA-BD	Kanamycin3 (50 µg/ml), Leucine4*
Prey Clone	TaRSR1_pGBDT7	pGADT7	Y187	TaRSR1	EcoR1-BamH1	500 bp	GAL4 AD	Ampicilin1 (100 µg/ml), Tryptophan
Bait Clone	TaMADS47_pGADT7	pGBKT7	Y2HGold	MADS47	EcoR1-BamH1	500 bp	GAL4 DNA-BD	Kanamycin3 (50 µg/ml), Leucine4*
Prey Clone	TaMADS47_pGBDT7	pGADT7	Y187	MADS47	EcoR1-BamH1	500 bp	GAL4 AD	Ampicilin1 (100 µg/ml), Tryptophan
Transient Dual Luciferase assay								
Entry clone	TaSPA_pDNR221	pDNR221	DH5α	TaSPA	attB+attP=>attL	800 bp	-	Kanamycin (50 µg/ml)
Destination vector			DH5α-gold					Spectinomycin (100 µg/ml)
	TaBLZ1_pDNR221	pDNR221	DH5α	TaBLZ1	attB+attP=>attL	100 bp	-	Kanamycin (50 µg/ml)
Destination vector			DH5α-gold					Spectinomycin (100 µg/ml)
	TaPBF_pDNR221	pDNR221	DH5α	TaPBF	attB+attP=>attL	800 bp	-	Kanamycin (50 µg/ml)
Destination vector			DH5α-gold					Spectinomycin (100 µg/ml)
	TaRSR1_pDNR221	pDNR221	DH5α	TaRSR1	attB+attP=>attL	600 bp	-	Kanamycin (50 µg/ml)
Destination vector			DH5α-gold					Spectinomycin (100 µg/ml)

Effectors								
Entry clone	**TaLMW_pDNR221**	**pDNR221**	**DH5α**	**Pro. LMW**	**attB+attP=>attL**	**1000 bp**	**-**	**Kanamycin (50 µg/ml)**
Destination vector			**DH5α-gold**				**Pro. LMW**	**Spectinomycin (100 µg/ml)**
Entry clone	**TaSIIIA_pDNR221**	**pDNR221**	**DH5α**	**Pro.SIIIA**	**attB+attP=>attL**	**1200bp**	**-**	**Kanamycin (50 µg/ml)**
Destination vector			**DH5α-gold**				**Pro.SIIIA**	**Spectinomycin (100 µg/ml)**
Entry clone	**TaSIIA_pDNR221**	**pDNR221**	**DH5α**	**Pro.SIIA**	**attB+attP=>attL**	**1200bp**	**-**	**Kanamycin (50 µg/ml)**
Destination vector			**DH5α-gold**				**Pro.SIIA**	**Spectinomycin (100 µg/ml)**
Entry clone	**TaSBE_pDNR221**	**pDNR221**	**DH5α**	**Pro.SBE**	**attB+attP=>attL**	**1000 bp**	**-**	**Kanamycin (50 µg/ml)**
Destination vector			**DH5α-gold**				**Pro.SBE**	**Spectinomycin (100 µg/ml)**
Reporter								
Entry clone	**RLuc**	**pDNR221**	**DH5α**	**RLuc**	**attB+attP=>attL**	**700 bp**	**-**	**Kanamycin (50 µg/ml)**
Destination vector			**DH5α**					

Appendix 7.4.5. Sequence of primers used

Yeast-2-Hybrid	
TaSPA(1)-EcoR1	GAGCCATAGCCAAAACCCGT
Ta-SPA(1)-BamH1	CGGATCCATTGAATCCAATGTAGTAGATTCCCGTTGC
TaSPA(2)-EcoR1	GACGAATTCGCGTACAACGCGATGCTC
TaSPA(2)-BamH1	CGGATCCATTGACGTATCTGTGACTTACATCTG
TaBLZ1(1)-EcoR1	GAATTCCATGTGGAGGGCCTCTG
TaBLZ1(1)-BamH1	CGGATCCATTGCTTCCTCCTCCGC
TaBLZ1(2)-EcoR1	GAATTCCATATGCCGAATTCTGGGGTCCCTACTGATC
TaBLZ1(2)-BamH1	CGGATCCATTGCTTCCTCCTCCGC
TaBLZ1(3)-EcoR1	GAATTCATCAGAGGCCACCTCCG
TaBLZ1(3)-BamH1	CGGATCCATTGACAGCCAAAGAACAAATAG
TaPBF - EcoR1	GAATTCCATATGC AAGTGTTTCCGTCAAAC
TaPBF - BamH1	CGGATCCATTGTATTGAGGAGGCTTTGC
TaMADS- EcoR1	GAATTCCATATGCGCTCGGCAAGAAG
TaMADS- BamH1	CGGATCCATTGGCGAAGATCACATGCAGAC
TaRSR1- EcoR1	GAATTCCATATGCAAGCAGATGAAGGGCCT
TaRSR1- BamH1	CGGATCCATTGTGGCTTCTTTTTCAGC
Ta PBF - EcoR1	GAATTCTGTTTCCGTCAAACTCCAAG
Ta PBF - BamH1	CGGATCCATTGAGGAGGCTTTGC
Transient Dual luciferase assay	
LMW/F	TGTATAGAAAAGTTGGATAGAACTGATGATTTATCT
LMW/R	TTTTGTACAAACTTGGGTGGATTGGTGTTAACCGTGT
SIIIA/F	TGTATAGAAAAGTTGTGTGAAATTCTAGTTTTACTTCTTG
SIIIA/R	TTTTGTACAAACTTGTTCCCGGAACCTACTAAA
SIIA/F	TGTATAGAAAAGTTGGGTCAAGAGACATTCTCTGTGATG
SIIA/R	TTTTGTACAAACTTGATCCGACGCGCCAGAAGT
SBE/F	GTATAGAAAAGTTGTGAATTATGGTCTTCGCTAGGAAG
SBE/R	TTTTGTACAAACTTGCGTCGTCGTCCACGGAACTGC
Ta SPA/F	TGTACAAAAAAGCAGGCTCTTAGCCATAGCCAAAACCCG
Ta SPA/R	TTGTACAAGAAAGCTGGGTCGAGAGATGATATGCTAGACATTG
Ta BLZ1/F	TGTACAAAAAAGCAGGCTCTATGTGGAGGGCCTCTG
Ta BLZ1/R	TGTACAAGAAAGCTGGGTCCACCACCAAGATATGTGTCTG
Ta RSR1/F	TGTACAAAAAAGCAGGCTCTAGGAGGACGACATGAAGCAG
Ta RSR1/R	TGTACAAGAAAGCTGGGTCGATTGGTGGATGCTCGGTCA
Ta PBF/F	TTG TAC AAA AAA GCA GGC TCTACGAATGAAGTGTTTCCGTCAAAC
Ta PBF/R	TTGTACAAGAAAGCTGGGTTTGGGATCGCTATTGAGGAGGCTTTGC
attB4-F adapter	GGGGACAACTTTGTATAGAAAAGTTG
attB1-R adapter	GGGGACTGCTTTTTTGTACAAACTTG
attB1-F adapter	GGGGACAAGTTTGTACAAAAAAGCAGGCT
attB2-R adapter	GGGGACCACTTTGTACAAGAAAGCTGGGT
attB2-F adapter	GGGGACAGCTTTCTTGTACAAAGTGG
attB3-R adapter	GGGGACAACTTTGTATAATAAAGTTG
M13 F	GTAAAACGACGGCCAG
M13 R	CAGGAAACAGCTATGAC

REFERENCES

REFERENCES

Adam, E., Szell, M., Szekeres, M., Schaefer, E. and Nagy, F., 1994. The developmental and tissue-specific expression of tobacco phytochrome-A genes. *The Plant Journal,* 6(3), pp. 283-293.

Albani, D., Hammond-Kosack, M.C., Smith, C., Conlan, S., Colot, V., Holdsworth, M. and 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,* 9(2), pp. 171-184.

Alber, T., 1992. Structure of the leucine zipper.*Current opinion in genetics & development,* 2(2), pp. 205-210.

Albert, H.A., Martin, T. and Sun, S.S., 1992. Structure and expression of a sugarcane gene encoding a housekeeping phosphoenolpyruvate carboxylase. *Plant Molecular Biology,* 20(4), pp. 663-671.

AlmlÖf, T., Gustafsson, J. and Wright, A.P. 1997. Role of hydrophobic amino acid clusters in the transactivation activity of the human glucocorticoid receptor. Molecular and cellular biology, 17(2), pp. 934-945.

Andersson, A.A.M., Courtin, C.M., Delcour, J.A., Fredriksson, H., Schofield, J.D., Trogh, I., Tsiami, A.A. and Åman, P., 2003. Milling performance of North European hull-less barleys and characterization of resultant millstreams. *Cereal chemistry*, *80*(6), pp.667-673.

Andersson, A., Andersson, R., Autio, K. and Åman, P., 1999. Chemical composition and microstructure of two naked waxy barleys. *Journal of cereal science,* 30(2), pp. 183-191.

Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M. and Hasebe, M., 2012. AP2-type transcription factors determine stem cell identity in the moss Physcomitrella patens. *Development,* 139(17), pp. 3120-3129

Ao, Z. and Jane, J., 2007. Characterization and modeling of the A-and B-granule starches of wheat, triticale, and barley. *Carbohydrate Polymers,* 67(1), pp. 46-55.

Aukerman, M.J. and Sakai, H., 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell*, *15*(11), pp.2730-2741.

Aukerman, M.J., Schmidt, R.J., Burr, B. and Burr, F.A., 1991. An arginine to lysine substitution in the bZIP domain of an opaque-2 mutant in maize abolishes specific DNA binding. *Genes & development,* 5(2), pp. 310-320.

Ba, A.N.N., Pogoutse, A., Provart, N. and Moses, A.M., 2009. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC bioinformatics,* 10(1), pp. 202.

Bai, C. and Elledge, S.J., 1997. Gene identification using the yeast two-hybrid system. Anonymous *Methods in enzymology.* Elsevier, pp. 141-156.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research,* 37(suppl_2), pp. W208.

Ball, S.G. and Morell, M.K. 2003. From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annual review of plant biology,* 54(1), pp. 207-233.

Bao, X., Franks, R.G., Levin, J.Z. and Liu, Z. 2004. Repression of AGAMOUS by BELLRINGER in floral and inflorescence meristems. *The Plant Cell,* 16(6), pp. 1478-1489.

Bao, J. 2014. Genes and QTLs for rice grain quality improvement. Anonymous *Rice-Germplasm, Genetics and Improvement.* InTech.

Bai, C. and Elledge, S.J. 1997. Gene identification using the yeast two-hybrid system. Anonymous *Methods in enzymology.* Elsevier, pp. 141-156.

Bailey, T.L. and Elkan, C.1995. The value of prior knowledge in discovering motifs with MEME. In *Ismb* (3), pp. 21-29).

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* 116(2), pp. 281-297.

Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M.M., Twardziok, S.O., Hackauf, B., Gordillo, A., Wilde, P. and Schmidt, M., 2017. Towards a whole-genome sequence for rye (Secale cereale L.). *The Plant Journal*, *89*(5), pp.853-869.

Bechtel, D.B. and Wilson, J.D.2003. Amyloplast formation and starch granule development in hard red winter wheat. *Cereal Chemistry,* 80(2), pp. 175-183.

Bechtel, D.B., Zayas, I., Kaleikau, L. and Pomeranz, Y.1990. Size-distribution of wheat starch granules during endosperm development. *Cereal Chemistry,* 67(1), pp. 59-63.

Becraft, P.W. and Asuncion-Crabb, Y. 2000. Positional cues specify and maintain aleurone cell fate in maize endosperm development. *Development,* 127(18), pp. 4039-4048.

Belitz, H., Grosch, W. and Schieberle, P. 2009. Cereals and cereal products. *Food Chemistry*, pp. 670-745.

Beltrano, J., Carbone, A., Montaldi, E.R. and Guiamet, J.J.1994. Ethylene as promoter of wheat grain maturation and ear senescence. *Plant Growth Regulation,* 15(2), pp. 107-112.

Bennetzen, J.L. and Hake, S.C. eds., 2009. *Handbook of maize: genetics and genomics*. Springer Science & Business Media.Bewley, J.D. and Black, M. 1994. Seeds', in Anonymous *Seeds.* Springer, pp. 1-33.

Berger, F. 1999. Endosperm development. *Current opinion in plant biology,* 2(1), pp. 28-32.

Birney, E., Clamp, M. and Durbin, R. 2004. GeneWise and genomewise. *Genome research,* 14(5), pp. 988-995.

Bolot, S., Abrouk, M., Masood-Quraishi, U., Stein, N., Messing, J., Feuillet, C. and Salse, J. 2009. The 'inner circle'of the cereal genomes'. *Current opinion in plant biology.* 12(2), pp. 119-125.

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C., van Lammeren, A.A. and Miki, B.L. 2002. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *The Plant Cell,* 14(8), pp. 1737-1749.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M. and Smyth, D.R. 1993. Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. *Development,* 119(3), pp. 721-743.

Bowman, J.L., Smyth, D.R. and Meyerowitz, E.M. 1989. Genes directing flower development in Arabidopsis. *The Plant Cell,* 1(1), pp. 37-52.

Bremer, K. (2002) Gondwanan evolution of the grass alliance of families (poales). Evolution, 56(7), 1374-87.

Briggs, D.E., 1978. The origin and classification of barleys. In *Barley* (pp. 76-88). Springer, Dordrecht.

Brocklehurst, P.A. 1977. Factors controlling grain weight in wheat. *Nature,* 266(5600), pp. 348.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L. and Voinnet, O. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Science,* 320(5880), pp. 1185-1190.

Brown, A.H., 2010. Variation under domestication in plants: 1859 and today. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1552), pp.2523-2530.Brown, R.C. and Lemmon, B.E. 2007. The developmental biology of cereal endosperm. Anonymous *Endosperm.* Springer, pp. 1-20.

Brown, R.H. and Bregitzer, P. 2011. A Ds insertional mutant of a barley miR172 gene results in indeterminate spikelet development. *Crop Science,* 51(4), pp. 1664-1672.

Brückner, A., Polge, C., Lentze, N., Auerbach, D. and Schlattner, U., 2009. Yeast two-hybrid, a powerful tool for systems biology. *International journal of molecular sciences*, *10*(6), pp.2763-2788.

Bruggmann, R., Bharti, A.K., Gundlach, H., Lai, J., Young, S., Pontaroli, A.C., Wei, F., Haberer, G., Fuks, G., Du, C. and Raymond, C., 2006. Uneven chromosome contraction and expansion in the maize genome. *Genome research*, *16*(10), pp.1241-1251.

Cai, J., Cai, C., Man, J., Yang, Y., Zhang, F. and Wei, C. 2014. Crystalline and structural properties of acid-modified lotus rhizome C-type starch. *Carbohydrate polymers*, *102*, pp.799-807.

Cai, Y., Xie, D., Wang, Z. and Hong, M. 2002. Interaction of rice bZIP protein REB with the 5′-upstream region of both rice sbe1 gene and waxy gene. *Chinese Science Bulletin,* 47(4), pp. 310-314.

Catalan, P., Kellogg, E.A. and Olmstead, R.G. 1997. Phylogeny of Poaceae Subfamily Pooideae Based on ChloroplastndhF Gene Sequences. Molecular phylogenetics and evolution, 8(2), pp. 150-166.

Cernac, A. and Benning, C. 2004. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *The Plant Journal,* 40(4), pp. 575-585.

Chanda, S.V. and Singh, Y.D. 1998. Cell enlargement as an important factor in controlling grain weight in wheat. *Journal of Agronomy and Crop Science,* 181(4), pp. 223-228.

Chang, W.C., Lee, T.Y., Huang, H.D., Huang, H.Y. and Pan, R.L. 2008. PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC genomics*, *9*(1), p.561.

Chen, J., Janssen, B., Williams, A. and Sinha, N. 1997. A gene fusion at a homeobox locus: alterations in leaf shape and implications for morphological evolution. *The Plant Cell,* 9(8), pp. 1289-1304.

Chen, J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A. and Lai, J. 2014. Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant physiology*, pp.pp-114.

Chen, W., Chao, G. and Singh, K.B. 1996. The promoter of a H2O2-inducible, Arabidopsis glutathione S-transferase gene contains closely linked OBF-and OBP1-binding sites. *The Plant Journal,* 10(6), pp. 955-966.

Chen, X. 2004. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science,* 303(5666), pp. 2022-2025.

Chen, Y., Gelfond, J.A., McManus, L.M. and Shireman, P.K. 2009. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC genomics,* 10(1), pp. 407.

Chen, Z.J. 2007.Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu.Rev.Plant Biol.,* 58, pp. 377-406.

Chen, Z.J. and Ni, Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *Bioessays,* 28(3), pp. 240-252.

Chern, M., Eiben, H.G. and Bustos, M.M. 1996. The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. *The Plant Journal,* 10(1), pp. 135-148.

Chiotelli, E. and Le Meste, M. 2002. Effect of small and large wheat starch granules on thermomechanical behavior of starch. *Cereal Chemistry,* 79(2), pp. 286-293.

Chojecki, A., Bayliss, M.W. and Gale, M.D.1986a. Cell production and DNA accumulation in the wheat endosperm, and their association with grain weight. *Annals of Botany,* 58(6), pp. 809-817.

Chojecki, A., Gale, M.D. and Bayliss, M.W.1986b.The number and sizes of starch granules in the wheat endosperm, and their association with grain weight. *Annals of Botany,* 58(6), pp. 819-831.

Chow, C.N., Zheng, H.Q., Wu, N.Y., Chien, C.H., Huang, H.D., Lee, T.Y., Chiang-Hsieh, Y.F., Hou, P.F., Yang, T.Y. and Chang, W.C., 2015. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic acids research*, *44*(D1), pp.D1154-D1160.

Chuck, G., Meeley, R.B. and Hake, S.1998. The control of maize spikelet meristem fate by theAPETALA2-like gene indeterminate spikelet1. *Genes & development,* 12(8), pp. 1145-1154.

Chuck, G., Meeley, R., Irish, E., Sakai, H. and Hake, S. 2007. The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. *Nature genetics,* 39(12), pp. 1517.

Ciceri, P., Gianazza, E., Lazzari, B., Lippoli, G., Genga, A., Hoscheck, G., Schmidt, R.J. and Viotti, A. 1997. Phosphorylation of Opaque2 changes diurnally and impacts its DNA binding activity.*The Plant Cell,* 9(1), pp. 97-108.

Cohen, D.R. and Curran, T. 1990. Analysis of dimerization and DNA binding functions in Fos and Jun by domain-swapping: involvement of residues outside the leucine zipper/basic region. *Oncogene,* 5(6), pp. 929-939.

Cochrane, M.P. and Duffus, C.M.1980. The nucellar projection and modified aleurone in the crease region of developing caryopses of barley Hordeum vulgare L. var. distichum. *Protoplasma,* 103(4), pp. 361-375.

Colonna, P. & Buleon, A.1992. New insights on starch structure and properties. *9. International cereal and bread congress. Paris (France). 1-5 Jun 1992*INRA.

Colonna, P. and Mercier, C. 1985. Gelatinization and melting of maize and pea starches with normal and high-amylose genotypes. *Phytochemistry*, *24*(8), pp.1667-1674.

Commuri, P.D. and Jones, R.J. 1999. Ultrastructural characterization of maize (Zea mays L.) kernels exposed to high temperature during endosperm cell division. *Plant, Cell & Environment,* 22(4), pp. 375-385.

Conant, G.C., Birchler, J.A. and Pires, J.C. 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time', *Current opinion in plant biology,* 19, pp. 91-98.

Coppock, J. 1975. TRITICALE. *Nutrition & Food Science,* 75(2), pp. 15-18.

Costa, L.M., Gutierrez-Marcos, J.F. and Dickinson, H.G. 2004. More than a yolk: the short life and complex times of the plant endosperm. *Trends in plant science,* 9(10), pp. 507-514.

Curaba, J., Spriggs, A., Taylor, J., Li, Z. and Helliwell, C. 2012. miRNA regulation in the early development of barley seed. *BMC plant biology,* 12(1), pp. 120.

Dang, C.V. and Lee, W. 1988. Identification of the human c-myc protein nuclear translocation signal. *Molecular and cellular biology,* 8(10), pp. 4048-4054.

Darlington, H.F., Tecsi, L., Harris, N., Griggs, D.L., Cantrell, I.C. and Shewry, P.R. 2000. Starch granule associated proteins in barley and wheat. *Journal of cereal science,* 32(1), pp. 21-29.

Davis, J.I. and Soreng, R.J. 1993. Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. *American Journal of Botany.* pp. 1444-1454.

Dayhoff, M.O. 1972. A model of evolutionary change in proteins. *Atlas of protein sequence and structure,* 5, pp. 89-99.

Dezulian, T., Remmert, M., Palatnik, J.F., Weigel, D. and Huson, D.H. 2005. Identification of plant microRNA homologs. Bioinformatics, 22(3), pp. 359-360.

de Folter, S., Immink, R.G., Kieffer, M., Pařenicová, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L. and Kater, M.M. 2005. Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *The Plant Cell,* 17(5), pp. 1424-1433.

Dellaporta, S.L. and Calderon-Urrea, A. 1993. Sex determination in flowering plants. *The Plant Cell,* 5(10), pp. 1241-1251.

Deluca, M. and McElroy, W.D. 1978. Purification and properties of firefly luciferase. Anonymous *Methods in enzymology.* Elsevier, pp. 3-15.

Deppmann, C.D., Acharya, A., Rishi, V., Wobbes, B., Smeekens, S., Taparowsky, E.J. and Vinson, C. 2004. Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs. *Nucleic acids research,* 32(11), pp. 3435-3445.

Diamond, J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature,* 418(6898), pp. 700.

Diaz, I. and Carbonero, P. 1998. Barley BLZ1: a bZIP transcriptional activator that interacts with endosperm-specific gene promoters. *Plant J,* 13, pp. 629-640.

Dietz, K., Vogel, M.O. and Viehhauser, A. 2010. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signaling. *Protoplasma,* 245(1-4), pp. 3-14.

Dingwall, C. and Laskey, R.A.1998. Nuclear import: a tale of two sites. *Current biology,* 8(25), pp. R924.

Djamei, A., Pitzschke, A., Nakagami, H., Rajh, I. and Hirt, H. 2007. Trojan horse strategy in Agrobacterium transformation: abusing MAPK defense signaling. *Science,* 318(5849), pp. 453-456.

Doebley, J., Stec, A. and Hubbard, L.1997. The evolution of apical dominance in maize. *Nature,* 386(6624), pp. 485.

Doebley, J. and Lukens, L. 1998. Transcriptional regulators and the evolution of plant form.*The Plant Cell,* 10(7), pp. 1075-1082.

Dombrink-Kurtzman, M.A. and Bietz, J.A. (1993) 'Main content area Zein composition in hard and soft endosperm of maize', *Cereal Chemistry,* 70(1), pp. 105-108.

Döring, E., Schneider, J., Hilu, K.W. and Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex (Pooideae, Poaceae). *Kew Bulletin*, pp.407-424.

Dornez, E., Holopainen, U., Cuyvers, S., Poutanen, K., Delcour, J.A., Courtin, C.M. and Nordlund, E. 2011.Study of grain cell wall structures by microscopic analysis with four different staining techniques. *Journal of cereal science.*

Drea, S., Leader, D.J., Arnold, B.C., Shaw, P., Dolan, L. and Doonan, J.H. (2005) 'Systematic spatial analysis of gene expression during wheat caryopsis development', The Plant Cell, 17(8), pp. 2172-2185.

Dubcovsky, J. and Dvorak, J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science,* 316(5833), pp. 1862-1866.

Egli, D.B. 1998.*Seed biology and the yield of grain crops.*

Ehlert, A., Weltmeier, F., Wang, X., Mayer, C.S., Smeekens, S., Vicente-Carbajosa, J. and Dröge-Laser, W. 2006.Two-hybrid protein–protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. *The Plant Journal,* 46(5), pp. 890-900.

Eilam, T., Anikster, Y., Millet, E., Manisterski, J., Sagi-Assif, O. and Feldman, M. 2007. Genome size and genome evolution in diploid Triticeae species. *Genome,* 50(11), pp. 1029-1037.

Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA complex. *Cell,* 71(7), pp. 1223-1237.

Earp, C.F., McDonough, C.M. and Rooney, L.W. 2004. Microscopy of pericarp development in the caryopsis of Sorghum bicolor (L.) Moench. *Journal of Cereal Science*, *39*(1), pp.21-27.

Evers, A.D. 1971. Scanning electron microscopy of wheat starch. III. Granule development in the endosperm', *Starch-Stärke,* 23(5), pp. 157-162.

Evers, T. and Millar, S. 2002. Cereal grain structure and development: some implications for quality. *Journal of cereal science,* 36(3), pp. 261-284.

Fahlgren, N., Montgomery, T.A., Howell, M.D., Allen, E., Dvorak, S.K., Alexander, A.L. and Carrington, J.C. 2006. Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. *Current biology,* 16(9), pp. 939-944.

Faltermaier, A., Zarnkow, M., Becker, T., Gastl, M. and Arendt, E.K. 2015. Common wheat (Triticum aestivum L.): evaluating microstructural changes during the malting process by using confocal laser scanning microscopy and scanning electron microscopy. *European Food Research and Technology,* 241(2), pp. 239-252.

Fang, Y., You, J., Xie, K., Xie, W. and Xiong, L. 2008. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Molecular Genetics and Genomics*, *280*(6), pp.547-563.

Fans, J.D., Simons, K.J., Zhang, Z. and Gill, B.S. 2005. The wheat super domestication gene Q. *Frontiers of Wheat Bioscience: Memorial Issue, Wheat Information Service,* (100), pp. 129-148.

Faris, J.D., Fellers, J.P., Brooks, S.A. and Gill, B.S. 2003. A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene.*Genetics,* 164(1), pp. 311-321.

Ferrante, A., Savin, R. and Slafer, G.A. 2012.Floret development and grain setting differences between modern durum wheats under contrasting nitrogen availability.*Journal of experimental botany,* 64(1), pp. 169-184.

Fields, S. and Song, O. 1989. A novel genetic system to detect protein–protein interactions. *Nature,* 340(6230), pp. 245.

Finnie, C. and Svensson, B., 2009. Barley seed proteomics from spots to structures. *Journal of proteomics*, *72*(3), pp.315-324.

Fischer, R.A. 2011.Wheat physiology: a review of recent developments. *Crop and Pasture Science,* 62(2), pp. 95-114.

Flemington, E. and Speck, S.H. 1990. Autoregulation of Epstein-Barr virus putative lytic switch gene BZLF1.*Journal of virology,* 64(3), pp. 1227-1232.

Flores, R. A., Hicks, K. B., Eustace, D. W., & Phillips, J. 2005. Highstarch and high b-glucan fractions milled with experimental mills. *Cereal Chemistry,* 82(6), 727e733.

Floyd, S.K. and Bowman, J.L., 2004. Gene regulation: ancient microRNA target sequences in plants. *Nature*, *428* (6982), p.485.

Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P., Pywell, J., Shewry, P.R., Bunce, N. and Miflin, B.J. 1985. Short tandem repeats shared by B-and C-hordein cDNAs suggest a common evolutionary origin for two groups of cereal storage protein genes.*The EMBO journal,* 4(1), pp. 9-15.

Foster, R., Izawa, T. and Chua, N. 1994. Plant bZIP proteins gather at ACGT elements.*The FASEB Journal,* 8(2), pp. 192-200.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P. and Solano, R., 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. Proceedings of the National Academy of Sciences, 111(6), pp.2367-2372.

Fu, F. and Xue, H. 2010.Co-expression analysis identifies Rice Starch Regulator1 (RSR1), a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. *Plant Physiology,* , pp. pp. 110.159517.

Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H. and Ohme-Takagi, M. 2000. Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box–mediated gene expression. *The Plant Cell,* 12(3), pp. 393-404.

Furtado, A., Henry, R.J. and Takaiwa, F. 2008. Comparison of promoters in transgenic rice. *Plant biotechnology journal,* 6(7), pp. 679-693.

Gaines, C.S., Finney, P.L. and Andrews, L.C. 1997.Influence of kernel size and shriveling on soft wheat milling and baking quality. *Cereal Chemistry,* 74(6), pp. 700-704.

Gaut, B.S. 2002.Evolutionary dynamics of grass genomes. *New Phytologist,* 154(1), pp. 15-28.

Geera, B.P., Nelson, J.E., Souza, E. and Huber, K.C. 2006. Composition and properties of A-and B-type starch granules of wild-type, partial waxy, and waxy soft wheat. *Cereal Chemistry*, *83*(5), pp.551-557.

Gibson, D.J. (2009) *Grasses and grassland ecology.* Oxford University Press.

Gill, B.S., Appels, R., Botha-Oberholster, A., Buell, C.R., Bennetzen, J.L., Chalhoub, B., Chumley, F., Dvořák, J., Iwanaga, M. and Keller, B. 2004. A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium', *Genetics,* 168(2), pp. 1087-1096.

Gil-Humanes, J., Pistón, F., Shewry, P.R., Tosi, P. and Barro, F. 2011. Suppression of gliadins results in altered protein body morphology in wheat. *Journal of experimental botany*, *62*(12), pp.4203-4213.

Gil-Humanes, J., Pistón, F., Hernando, A., Alvarez, J.B., Shewry, P.R. and Barro, F. 2008. Silencing of γ-gliadins by RNA interference (RNAi) in bread wheat. *Journal of cereal science,* 48(3), pp. 565-568.

Gil-Humanes, J., Pistón, F., Tollefsen, S., Sollid, L.M. and Barro, F. 2010. Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. *Proceedings of the National Academy of Sciences*, *107*(39), pp.17023-17028.

Gill, G. & Ptashne, M. 1987. Mutants of GAL4 protein altered in an activation function. Cell, 51, 121-6.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. and Toulmin, C. 2010.Food security: the challenge of feeding 9 billion people', *Science,* , pp. 1185383.

Goff, S.A., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P. and Varma, H. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science,* 296(5565), pp. 92-100.

Goldfarb, A.N. and Lewandowska, K. 1994. Nuclear redirection of a cytoplasmic helix-loop-helix protein via heterodimerization with a nuclear localizing partner', *Experimental cell research,* 214(2), pp. 481-485.

Golldack, D., Lüking, I. and Yang, O.2011. Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. Plant cell reports, 30(8), pp.1383-1391.

Gu, Y.Z., Hogenesch, J.B. and Bradfield, C.A. 2000. The PAS superfamily: sensors of environmental and developmental signals. *Annual review of pharmacology and toxicology*, *40*(1), pp.519-561.

Grass Phylogeny Working Group, Barker, N.P., Clark, L.G., Davis, J.I., Duvall, M.R., Guala, G.F., Hsiao, C., Kellogg, E.A., Linder, H.P. and Mason-Gamer, R.J. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden,*pp. 373-457.

Greenwood, J.R., Finnegan, E.J., Watanabe, N., Trevaskis, B. and Swain, S.M. 2017. New alleles of the wheat domestication gene Q reveal multiple roles in growth and reproductive development', Development, 144(11), pp. 1959-1965.

Guo, A.Y., Zhu, Q.H., Chen, X. and Luo, J.C., 2007. GSDS: a gene structure display server. *Yi chuan= Hereditas*, *29*(8), pp.1023-1026.

Gustafson, J.P. and Flavell, R.B.1996. Control of nucleolar expression in triticale. Anonymous *Triticale: today and tomorrow.* Springer, pp. 119-125.

Guzman, C., Peña, R.J., Singh, R., Autrique, E., Dreisigacker, S., Crossa, J., Rutkoski, J., Poland, J. and Battenfield, S., 2016. Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Applied & translational genomics*, *11*, pp.3-8.Gronemeyer, H., Gustafsson, J.Å. and Laudet, V., 2004. Principles for modulation of the nuclear receptor superfamily. *Nature reviews Drug discovery*, *3*(11), p.950.

Haasen, D., Köhler, C., Neuhaus, G. and Merkle, T. 1999. Nuclear export of proteins in plants: AtXPO1 is the export receptor for leucine-rich nuclear export signals in Arabidopsis thaliana.*The Plant Journal,* 20(6), pp. 695-705.

Halford, N.G. and Shewry, P.R. 2007.The structure and expression of cereal storage protein genes. Anonymous *Endosperm.* Springer, pp. 195-218.

Hammond-Kosack, M.C., Holdsworth, M.J. and Bevan, M.W. 1993. In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. *The EMBO journal,* 12(2), pp. 545-554.

Hands, P. and Drea, S. 2012.A comparative view of grain development in Brachypodium distachyon. *Journal of cereal science,* 56(1), pp. 2-8.

Hannah, L.C. 2005. Starch synthesis in the maize endosperm, *Maydica,* 50(3/4), pp. 497.

Hao, M., Luo, J., Zhang, L., Yuan, Z., Yang, Y., Wu, M., Chen, W., Zheng, Y., Zhang, H. and Liu, D. 2013. Production of hexaploid triticale by a synthetic hexaploid wheat-rye hybrid method. *Euphytica*, *193*(3), pp.347-357.

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. and Thompson, R. 1989. The O2 gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *The EMBO journal,* 8(10), pp. 2795-2801.

Haseneyer, G., Stracke, S., Piepho, H.P., Sauer, S., Geiger, H.H. and Graner, A., 2010. DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. *BMC plant biology*, *10*(1), p.5.

Haseneyer, G., Schmutzer, T., Seidel, M., Zhou, R., Mascher, M., Schön, C.C., Taudien, S., Scholz, U., Stein, N., Mayer, K.F. and Bauer, E., 2011. From RNA-

seq to large-scale genotyping-genomics resources for rye (Secale cereale L.). *BMC Plant Biology*, *11*(1), p.131.

Haughn, G.W. and Somerville, C.R. 1988.Genetic control of morphogenesis in Arabidopsis, *genesis,* 9(2), pp. 73-89.

Hennen-Bierwagen, T.A., Lin, Q., Grimaud, F., Planchot, V., Keeling, P.L., James, M.G. and Myers, A.M. 2009.Proteins from multiple metabolic pathways associate with starch biosynthetic enzymes in high molecular weight complexes: a model for regulation of carbon allocation in maize amyloplasts. *Plant Physiology,* 149(3), pp. 1541-1559.

Hizukuri, S., Takeda, Y., Yasuda, M. and Suzuki, A. 1981.Multi-branched nature of amylose and the action of debranching enzymes. *Carbohydrate research,* 94(2), pp. 205-213.

Hope, I. A. & Struhl, K. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell, 46, 885-94.

Holopainen, U.R., Wilhelmson, A., Salmenkallio-Marttila, M., Peltonen-Sainio, P., Rajala, A., Reinikainen, P., Kotaviita, E., Simolin, H. and Home, S. 2005. Endosperm structure affects the malting quality of barley (Hordeum vulgare L.). *Journal of Agricultural and Food Chemistry,* 53(18), pp. 7279-7287.

Hood, E.E., Bailey, M.R., Beifuss, K., Magallanes-Lundback, M., Horn, M.E., Callaway, E., Drees, C., Delaney, D.E., Clough, R. and Howard, J.A. 2003. Criteria for high-level expression of a fungal laccase gene in transgenic maize', *Plant Biotechnology Journal,* 1(2), pp. 129-140.

Hoshikawa, K. 1993. Rice grain texture and grading standard. In: T. Matsuo & K. Hoshilawa (Eds), 'Science of the Rice Plant, pp. 383–389. Vol.1. Morphology, Food and Agriculture Policy Research Center, Tokyo, Japan.

Hu, P.P., Liu, H., Zhan, L., Zheng, L.L. and Huang, C.Z. 2015.Coomassie brilliant blue R-250 as a new surface-enhanced Raman scattering probe for prion protein through a dual-aptamer mechanism', *Talanta,* 139, pp. 35-39.

Huijser, P. and Schmid, M. 2011.The control of developmental phase transitions in plants. *Development,* 138(19), pp. 4117-4129.

Hung, F., Luo, D., Sauvé, D.M., Muller, M.T. and Roberge, M. 1996. Characterization of topoisomerase II-DNA interaction and identification of a DNA-binding domain by ultraviolet laser crosslinking. *FEBS letters,* 380(1-2), pp. 127-132.

Huo, N., Vogel, J.P., Lazo, G.R., You, F.M., Ma, Y., McMahon, S., Dvorak, J., Anderson, O.D., Luo, M. and Gu, Y.Q. 2009.Structural characterization of Brachypodium genome and its syntenic relationship with rice and wheat. *Plant Molecular Biology,* 70(1-2), pp. 47-61.

Hurkman, W.J. and Wood, D.F. 2011.High temperature during grain fill alters the morphology of protein and starch deposits in the starchy endosperm cells of

developing wheat (Triticum aestivum L.) grain. *Journal of Agricultural and Food Chemistry,* 59(9), pp. 4938-4946.

International Brachypodium Initiative.2010. Genome sequencing and analysis of the model grass Brachypodium distachyon.*Nature,* 463(7282), pp. 763.

Irish, V.F. and Sussex, I.M. 1990.Function of the apetala-1 gene during Arabidopsis floral development. *The Plant Cell,* 2(8), pp. 741-753.

Izawa, T., Foster, R., Nakajima, M., Shimamoto, K. and Chua, N. 1994.The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development.*The Plant Cell,* 6(9), pp. 1277-1287.

Jääskeläinen, A.S., Holopainen-Mantila, U., Tamminen, T. and Vuorinen, T., 2013. Endosperm and aleurone cell structure in barley and wheat as studied by optical and Raman microscopy. Journal of cereal science, 57(3), pp.543-550.

Jane, J.L., 2006. Current understanding on starch granule structures. *Journal of Applied Glycoscience*, *53*(3), pp.205-213.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. and Parcy, F. 2002. bZIP transcription factors in Arabidopsis. Trends in plant science, 7(3), pp.106-111.

Jeon, J.S., Ryoo, N., Hahn, T.R., Walia, H. and Nakamura, Y. 2010. Starch biosynthesis in cereal endosperm. *Plant physiology and Biochemistry*, *48*(6), pp.383-392.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H. and Soltis, P.S. 2011.Ancestral polyploidy in seed plants and angiosperms. *Nature,* 473(7345), pp. 97.

Jin, D., Wang, Y., Zhao, Y. and Chen, M. 2013. MicroRNAs and their cross-talks in plant development. *Journal of Genetics and Genomics,* 40(4), pp. 161-170.

Jing, Y.P., Liu, D.T., Yu, X.R., Xiong, F., Li, D.L., Zheng, Y.K., Hao, Y.F., Gu, Y.J. and Wang, Z. 2014. Development of endosperm cells and starch granules in common wheat', *Cereal research communications,* 42(3), pp. 514-524.

Jofuku, K.D., Den Boer, B.G., Van Montagu, M. and Okamuro, J.K. 1994. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *The Plant Cell,* 6(9), pp. 1211-1225.

Jones, D.T., Taylor, W.R. and Thornton, J.M. 1992.The rapid generation of mutation data matrices from protein sequences. *Bioinformatics,* 8(3), pp. 275-282.

Jones, R.J., Schreiber, B. and Roessler, J.A. 1996.Kernel sink capacity in maize: genotypic and maternal regulation.*Crop Science,* 36(2), pp. 301-306.

Jones, R. 1969.The fine structure of barley aleurone cells. *Planta; An International Journal of Plant Biology,* 85(4), pp. 359-375.

Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants.*Annu.Rev.Plant Biol.,* 57, pp. 19-53.

Jung, J., Seo, Y., Seo, P.J., Reyes, J.L., Yun, J., Chua, N. and Park, C. 2007.The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. *The Plant Cell,* 19(9), pp. 2736-2748.

Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. 1984. A short amino acid sequence able to specify nuclear location. *Cell,* 39(3), pp. 499-509.

Kang, G., Xu, W., Liu, G., Peng, X. and Guo, T. 2012. Comprehensive analysis of the transcription of starch synthesis genes and the transcription factor RSR1 in wheat (Triticum aestivum) endosperm. *Genome,* 56(2), pp. 115-122.

Kapushesky, M., Adamusiak, T., Burdett, T., Culhane, A., Farne, A., Filippov, A., Holloway, E., Klebanov, A., Kryvych, N. and Kurbatova, N. 2011.Gene Expression Atlas update—a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic acids research,* 40(D1), pp. D1081.

Karlsson, R., Olered, R. and Eliasson, A. 1983. Changes in starch granule size distribution and starch gelatinization properties during development and maturation of wheat, barley and rye. *Starch-Stärke,* 35(10), pp. 335-340.

Katagiri, F., Seipel, K. and Chua, N.H. 1992. Identification of a novel dimer stabilization region in a plant bZIP transcription activator. *Molecular and cellular biology,* 12(11), pp. 4809-4816.

Kater, M.M., Colombo, L., Franken, J., Busscher, M., Masiero, S., Campagne, Michiel M Van Lookeren and Angenent, G.C.1998.Multiple AGAMOUS homologs from cucumber and petunia differ in their ability to induce reproductive organ fate. *The Plant Cell,* 10(2), pp. 171-182.

Kawakatsu, T., Yamamoto, M.P., Touno, S.M., Yasuda, H. and Takaiwa, F. 2009. Compensation and interaction between RISBZ1 and RPBF during grain filling in rice. *The Plant Journal,* 59(6), pp. 908-920.

Kawakatsu, T. and Takaiwa, F. 2010. Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. *Plant biotechnology journal,* 8(9), pp. 939-953.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. and Duran, C. 2012.Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28(12), pp. 1647-1649.

Keeling, P.L. and Myers, A.M. 2010.Biochemistry and genetics of starch synthesis.

Kellogg, E.A. 2001.Evolutionary history of the grasses. *Plant Physiology,* 125(3), pp. 1198-1205.

Kellogg, E.A. 1998. Relationships of cereal crops and other grasses. *Proceedings of the National Academy of Sciences,* 95(5), pp. 2005-2010.

Kellogg, E.A., Camara, P.E., Rudall, P.J., Ladd, P., Malcomber, S.T., Whipple, C. and Doust, A.N. 2013. Early inflorescence development in the grasses (Poaceae). *Frontiers in plant science,* 4, pp. 250.

Kersey, P.J., Allen, J.E., Allot, A., Barba, M., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P. and Grabmueller, C. 2017. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic acids research,* 46(D1), pp. D808.

Khraiwesh, B., Arif, M.A., Seumel, G.I., Ossowski, S., Weigel, D., Reski, R. and Frank, W. 2010.Transcriptional control of gene expression by microRNAs. *Cell,* 140(1), pp. 111-122.

Kidner, C.A. and Martienssen, R.A. 2005. The developmental role of microRNA in plants. *Current opinion in plant biology,* 8(1), pp. 38-44.

Kim, S., Soltis, P.S., Wall, K. and Soltis, D.E. 2005. Phylogeny and domain evolution in the APETALA2-like gene family. *Molecular biology and evolution,* 23(1), pp. 107-120.

Kim, H. and Huber, K.C. 2008. Channels within soft wheat starch A-and B-type granules. *Journal of cereal science,* 48(1), pp. 159-172.

Kirchler, T., Briesemeister, S., Singer, M., Schütze, K., Keinath, M., Kohlbacher, O., Vicente-Carbajosa, J., Teige, M., Harter, K. and Chaban, C. 2010. The role of phosphorylatable serine residues in the DNA-binding domain of Arabidopsis bZIP transcription factors. European journal of cell biology, 89(2-3), pp.175-183.

Klinge, B., Überlacker, B., Korfhage, C. and Werr, W. 1996. ZmHox: a novel class of maize homeobox genes. Plant molecular biology, 30(3), pp.439-453.

Komaki, M.K., Okada, K., Nishino, E. and Shimura, Y. 1988. Isolation and characterization of novel mutants of Arabidopsis thaliana defective in flower development. *Development,* 104(2), pp. 195-203.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., Stein, N., Graner, A., Wicker, T. and Tagiri, A. 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proceedings of the National Academy of Sciences,* 104(4), pp. 1424-1429.

Koncz, C., Langridge, W.H., Olsson, O., Schell, J. and Szalay, A.A. (1990) 'Bacterial and firefly luciferase genes in transgenic plants: advantages and disadvantages of a reporter gene', *Developmental genetics,* 11(3), pp. 224-232.

König, P. and Richmond, T.J. 1993.The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *Journal of Molecular Biology,* 233(1), pp. 139-154.

Sabelli, P.A., 2012. Replicate and die for your own good: endoreduplication and cell death in the cereal endosperm. *Journal of cereal science*, *56*(1), pp.9-20.Kozomara, A. and Griffiths-Jones, S. 2013.

miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research,* 42(D1), pp. D73.

Kreis, M., Forde, B.G., Rahman, S., Miflin, B.J. and Shewry, P.R. 1985. Molecular evolution of the seed storage proteins of barley, rye and wheat. *Journal of Molecular Biology,* 183(3), pp. 499-502.

Krizek, B.A. 2003. AINTEGUMENTA utilizes a mode of DNA recognition distinct from that used by proteins containing a single AP2 domain. *Nucleic acids research,* 31(7), pp. 1859-1868.

Krizek, B.A., Prost, V. and Macias, A. (2000) 'AINTEGUMENTA promotes petal identity and acts as a negative regulator of AGAMOUS', *The Plant Cell,* 12(8), pp. 1357-1366.

Kunst, L., Klenz, J.E., Martinez-Zapater, J. and Haughn, G.W. 1989.AP2 gene determines the identity of perianth organs in flowers of Arabidopsis thaliana.*The Plant Cell,* 1(12), pp. 1195-1208.

Lacadena, J.R., Cermeno, M., Orellana, J. and Santos, J.L. 1984. Evidence for wheat-rye nucleolar competition (amphiplasty) in triticale by silver-staining procedure. *Theoretical and Applied Genetics,* 67(2-3), pp. 207-213.

Lackey, J.A. 2010. Endosperm size diversity in domesticated, wild, and semiwild soybean. *Crop Science,* 50(1), pp. 168-176.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E. and Corbett, A.H. 2007. Classical nuclear localization signals: definition, function, and interaction with importin α. *Journal of Biological Chemistry,* 282(8), pp. 5101-5105.

Lara, P., Oñate-Sánchez, L., Abraham, Z., Ferrándiz, C., Díaz, I., Carbonero, P. and Vicente-Carbajosa, J. 2003.Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2', Journal of Biological Chemistry, 278(23), pp. 21003-21011.

Lata, C., Mishra, A.K., Muthamilarasan, M., Bonthala, V.S., Khan, Y. and Prasad, M. 2014. Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (Setaria italica L.). PLoS One, 9(11), pp. e113092.

Lauter, N., Kampani, A., Carlson, S., Goebel, M. and Moose, S.P. 2005. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *Proceedings of the National Academy of Sciences,* 102(26), pp. 9412-9417.

Le, S.Q., Lartillot, N. and Gascuel, O. (2008) 'Phylogenetic mixture models for proteins', *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 363(1512), pp. 3965-3976.

Lee, D. and An, G.2012.Two AP2 family genes, supernumerary bract (SNB) and Osindeterminate spikelet 1 (OsIDS1), synergistically control inflorescence architecture and floral meristem establishment in rice. *The Plant Journal for cell and molecular biology,* 69(3), pp. 445.

Lee, H., Yoo, S.J., Lee, J.H., Kim, W., Yoo, S.K., Fitzgerald, H., Carrington, J.C. and Ahn, J.H. 2010. Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. *Nucleic acids research,* 38(9), pp. 3081-3093.

Lehmann, U. and Robin, F. 2007. Slowly digestible starch–its structure and health implications: a review. *Trends in Food Science & Technology,* 18(7), pp. 346-355.

Leon-Kloosterziel, K.M., Keijzer, C.J. and Koornneef, M. 1994. A seed shape mutant of Arabidopsis that is affected in integument development. *The Plant Cell,* 6(3), pp. 385-392.

Leroux, B.M., Goodyke, A.J., Schumacher, K.I., Abbott, C.P., Clore, A.M., Yadegari, R., Larkins, B.A. and Dannenhoffer, J.M. 2014. Maize early endosperm growth and development: from fertilization through cell type differentiation. *American Journal of Botany,* 101(8), pp. 1259-1274.

Li, C.Y., Li, C., Lu, Z.X., Li, W.H. and Cao, L.P. 2012. Morphological changes of starch granules during grain filling and seed germination in wheat. *Starch-Stärke*, *64*(2), pp.166-170.

Li, H., Xiao, Q., Zhang, C., Du, J., Li, X., Huang, H., Wei, B., Li, Y., Yu, G. and Liu, H. 2017. Identification and characterization of transcription factor ZmEREB94 involved in starch synthesis in maize', *Journal of Plant Physiology,* 216, pp. 11-16.

Li, J.H., Vasanthan, T., Hoover, R. and Rossnagel, B.G. 2004.Starch from hull-less barley: V. In-vitro susceptibility of waxy, normal, and high-amylose starches towards hydrolysis by alpha-amylases and amyloglucosidase. *Food Chemistry,* 84(4), pp. 621-632.

Licausi, F., Ohme-Takagi, M. and Perata, P. 2013. APETALA 2/Ethylene Responsive Factor (AP 2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist,* 199(3), pp. 639-649.

Lima-Brito, J., Guedes-Pinto, H., Harrison, G.E. and Heslop-Harrison, J.S. 1996. Chromosome identification and nuclear architecture in triticale× tritordeum F1 hybrids. *Journal of experimental botany,* 47(4), pp. 583-588.

Lindeboom, N., Chang, P.R. and Tyler, R.T. 2004. Analytical, biochemical and physicochemical aspects of starch granule size, with emphasis on small granule starches: a review', *Starch-Stärke,* 56(3-4), pp. 89-99.

Liu, C., Mao, B., Ou, S., Wang, W., Liu, L., Wu, Y., Chu, C. and Wang, X. 2014. OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. *Plant molecular biology*, *84*(1-2), pp.19-36.

Liu, C. and Zhang, T. 2017.Expansion and stress responses of the AP2/EREBP superfamily in cotton. *BMC genomics,* 18(1), pp. 118.

Liu, C., Wu, Y. and Wang, X. 2012.bZIP transcription factor OsbZIP52/RISBZ5: a potential negative regulator of cold and drought stress response in rice. *Planta,* 235(6), pp. 1157-1169.

Liu, H., Liu, Y., Yang, M. and Shen, S. 2009. A comparative analysis of embryo and endosperm proteome from seeds of Jatropha curcas. *Journal of integrative plant biology,* 51(9), pp. 850-857.

Liu, J. and Coaker, G. 2008. Nuclear trafficking during plant innate immunity. *Molecular plant,* 1(3), pp. 411-422.

Liu, L., White, M.J. and MacRae, T.H. 1999.Transcription factors and their genes in higher plants. *The FEBS Journal,* 262(2), pp. 247-257.

Liu, P., Liu, J., Dong, H. and Sun, J. 2018. Functional regulation of Q by microRNA172 and transcriptional co-repressor TOPLESS in controlling bread wheat spikelet density. *Plant biotechnology journal*, *16*(2), pp.495-506.

Liu, Q., Gu, Z., Donner, E., Tetlow, I. and Emes, M. 2007. Investigation of digestibility in vitro and physicochemical properties of A-and B-type starch from soft and hard wheat flour. *Cereal Chemistry,* 84(1), pp. 15-21.

Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science,* 297(5589), pp. 2053-2056.

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N., Hartings, H., Salamini, F. and Thompson, R.D. 1991. The maize regulatory locus Opaque-2 encodes a DNA-binding protein which activates the transcription of the b-32 gene. *The EMBO journal,* 10(3), pp. 617-624.

Lopes, M.A. and Larkins, B.A. 1993. Endosperm origin, development, and function', *The Plant Cell,* 5(10), pp. 1383.

Lüscher, B. 2001. Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene,* 277(1), pp. 1-14.

Lyck, R., Harmening, U., Höhfeld, I., Treuter, E., Scharf, K.D. and Nover, L., 1997. Intracellular distribution and identification of the nuclear localization signals of two plant heat-stress transcription factors. Planta, 202(1), pp.117-125.

Ma, X. and Gustafson, J.P. 2008.Allopolyploidization-accommodated genomic sequence changes in triticale', *Annals of Botany,* 101(6), pp. 825-832.

Magnani, E., Sjölander, K. and Hake, S., 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. The Plant Cell, 16(9), pp.2265-2277.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M. and Hurwitz, D.I. 2014. CDD: NCBI's conserved domain database. *Nucleic acids research,* 43(D1), pp. D226.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B.B., Steuernagel, B., Mayer, K.F. and Olsen, O. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science,* 345(6194), pp. 1250092.

Massari, M.E. and Murre, C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and cellular biology,* 20(2), pp. 429-440.

Mathieu, J., Yant, L.J., Mürdter, F., Küttner, F. and Schmid, M. 2009. Repression of flowering by the miR172 target SMZ. PLoS biology, 7(7), p.e1000148.

Matthews,J.C.,Hori,K.andCormier,M.J. 1977. PurificationandpropertiesofRenilla reniformis luciferase. Biochemistry, 16(1), 85–91. doi: 10.1021/bi00620a014 99

Meier, I. and Gruissem, W. 1994. Novel conserved sequence motifs in plant G-box binding proteins and implications for interactive domains. *Nucleic acids research,* 22(3), pp. 470-478.

. Martinez-Perez, E., Shaw, P. and Moore, G. 2001. The Ph1 locus is needed to ensure specific somatic and meiotic centromere association. *Nature,* 411(6834), pp. 204.

Meng, F., Liu, H., Wang, K., Liu, L., Wang, S., Zhao, Y., Yin, J. and Li, Y. 2013 'Development-associated microRNAs in grains of wheat (Triticum aestivum L.). *BMC plant biology,* 13(1), pp. 140.

Miflin, B.J., Field, J.M. and Shewry, P.R. 1983.Cereal storage proteins and their effect on technological properties. *Seed proteins/edited by J.Daussant, J.Mosse, J.Vaughan,* .

Miflin, B.J., Burgess, S.R. and Shewry, P.R. 1981.The Development of Protein Bodies in the Storage Tissues1. *Journal of experimental botany,* 32(1), pp. 199-219.

Miki, D., Itoh, R. and Shimamoto, K. 2005. RNA silencing of single and multiple members in a gene family of rice. *Plant Physiology,* 138(4), pp. 1903-1913.

Minami, M., Huh, G.H., Yang, P. and Iwabuchi, M. 1993.Coordinate gene expression of five subclass histones and the putative transcription factors, HBP-1a and HBP-1b, of histone genes in wheat. *Plant Molecular Biology,* 23(2), pp. 429-434.

Miralles, D.J., Richards, R.A. and Slafer, G.A. 2000.Duration of the stem elongation period influences the number of fertile florets in wheat and barley.*Functional Plant Biology,* 27(10), pp. 931-940.

Mizukami, Y. and Fischer, R.L. 2000. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proceedings of the National Academy of Sciences,* 97(2), pp. 942-947.

Modrusan, Z., Reiser, L., Feldmann, K.A., Fischer, R.L. and Haughn, G.W. 1994. Homeotic transformation of ovules into carpel-like structures in Arabidopsis.*The Plant Cell,* 6(3), pp. 333-349.

Moose, S.P. and Sisco, P.H. 1996. Glossy15, an APETALA2-like gene from maize that regulates leaf epidermal cell identity. *Genes & development,* 10(23), pp. 3018-3027.

Moose, S.P. and Sisco, P.H. 1994. Glossy15 controls the epidermal juvenile-to-adult phase transition in maize.*The Plant Cell,* 6(10), pp. 1343-1355.

Morell, M.K. and Myers, A.M.2005. Towards the rational design of cereal starches. *Current Opinion in Plant Biology*, *8*(2), pp.204-210.

Morell, M.K., Kosar-Hashemi, B., Cmiel, M., Samuel, M.S., Chandler, P., Rahman, S., Buleon, A., Batey, I.L. and Li, Z. 2003. Barley sex6 mutants lack starch synthase IIa activity and contain a starch with novel properties. *The Plant Journal,* 34(2), pp. 173-185.

Morris, P.C. and Bryce, J.H. (2000) *Cereal biotechnology.* CRC Press.

Morrison, I.N., Kuo, J. and O'Brien, T.P. 1975. Histochemistry and fine structure of developing wheat aleurone cells. *Planta,* 123(2), pp. 105-116.

Morrison, I.N. and O'brien, T.P. 1976. Cytokinesis in the developing wheat grain; division with and without a phragmoplast. *Planta,* 130(1), pp. 57-67.

Müller, M. and Knudsen, S. 1993. The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box. *The Plant Journal,* 4(2), pp. 343-355.

Murphy, D.J. 2007.*People, plants & genes: the story of crops and humanity.* Oxford University Press on Demand.

Näär, A.M., Lemon, B.D. and Tjian, R. 2001. Transcriptional coactivator complexes', Annual Review of Biochemistry, 70(1), pp. 475-501.

Nair, S.K., Wang, N., Turuspekov, Y., Pourkheirandish, M., Sinsuwongwat, S., Chen, G., Sameri, M., Tagiri, A., Honda, I. and Watanabe, Y. 2010.Cleistogamous flowering in barley arises from the suppression of

microRNA-guided HvAP2 mRNA cleavage. *Proceedings of the National Academy of Sciences,* 107(1), pp. 490-495.

Nakamura, Y. 2002. Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. *Plant and Cell Physiology,* 43(7), pp. 718-725.

Nakano, T., Suzuki, K., Fujimura, T. and Shinshi, H. 2006. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiology,* 140(2), pp. 411-432.

Nakase, M., Aoki, N., Matsuda, T. and Adachi, T. 1997. Characterization of a novel rice bZIP protein which binds to the α-globulin promoter. *Plant Molecular Biology,* 33(3), pp. 513-522.

Nantel, A. and Quatrano, R.S. 1996.Characterization of three rice basic/leucine zipper factors, including two inhibitors of EmBP-1 DNA binding activity. *Journal of Biological Chemistry,* 271(49), pp. 31296-31305.

Neto, G.C., Yunes, J.A., da Silva, M.J., Vettore, A.L., Arruda, P. and Leite, A. 1995. The involvement of Opaque 2 on β-prolamin gene regulation in maize and Coix suggests a more general role for this transcriptional activator. *Plant Molecular Biology,* 27(5), pp. 1015-1029.

Nguyen, H.N., Sabelli, P.A. and Larkins, B.A. 2007.Endoreduplication and programmed cell death in the cereal endosperm. Anonymous *Endosperm.* Springer, pp. 21-43.

Nevo, E. 2011. 10.1 Domestication as a Gigantic Human Evolutionary Experiment. *Wild Crop Relatives: Genomic and Breeding Resources: Cereals,* 1, pp. 407.

Nijhawan, A., Jain, M., Tyagi, A.K. and Khurana, J.P. 2008. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiology,* 146(2), pp. 333-350.

Nodine, M.D. and Bartel, D.P. 2010.MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes & development,* 24(23), pp. 2678-2692.

Nonogaki, H., Gee, O.H. and Bradford, K.J. 2000.A germination-specific endo-β-mannanase gene is expressed in the micropylar endosperm cap of tomato seeds. *Plant Physiology,* 123(4), pp. 1235-1246.

Nuin, P.A., Wang, Z. and Tillier, E.R. 2006. The accuracy of several multiple sequence alignment programs for proteins. *BMC bioinformatics,* 7(1), pp. 471.

Ober, E.S., Setter, T.L., Madison, J.T., Thompson, J.F. and Shapiro, P.S. 1991. Influence of water deficit on maize endosperm development: enzyme activities and RNA transcripts of starch and zein synthesis, abscisic acid, and cell division. *Plant Physiology,* 97(1), pp. 154-164.

Ohdan, T., Francisco Jr, P.B., Sawada, T., Hirose, T., Terao, T., Satoh, H. and Nakamura, Y. 2005. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. Journal of experimental botany, 56(422), pp. 3229-3244.

Ohler, U. and Niemann, H. 2001. Identification and analysis of eukaryotic promoters: recent computational approaches', TRENDS in Genetics, 17(2), pp. 56-60.

Ohta, M., Matsui, K., Hiratsu, K., Shinshi, H. and Ohme-Takagi, M. 2001.Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The Plant Cell,* 13(8), pp. 1959-1968.

Ohto, M., Fischer, R.L., Goldberg, R.B., Nakamura, K. and Harada, J.J. 2005.Control of seed mass by APETALA2. *Proceedings of the National Academy of Sciences of the United States of America,* 102(8), pp. 3123-3128.

Olsen, O., Potter, R.H. and Kalla, R. 1992.Histo-differentiation and molecular biology of developing cereal endosperm. *Seed Science Research,* 2(3), pp. 117-131.

Olsen, O. 2004. Nuclear endosperm development in cereals and Arabidopsis thaliana. The Plant Cell, 16(suppl 1), pp. S227.

Olsen, O. 2001.Endosperm development: cellularization and cell fate specification', Annual review of plant biology, 52(1), pp. 233-267.

Oñate, L., Vicente-Carbajosa, J., Lara, P., Díaz, I. and 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', *Journal of Biological Chemistry,* 274(14), pp. 9175-9182.

Onodera, Y., Suzuki, A., Wu, C., Washida, H. and Takaiwa, F. 2001.A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. *Journal of Biological Chemistry,* 276(17), pp. 14139-14152.

Okamuro, J.K., Caster, B., Villarroel, R., Van Montagu, M. and Jofuku, K.D. 1997. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proceedings of the National Academy of Sciences,* 94(13), pp. 7076-7081.

Opanowicz, M., Vain, P., Draper, J., Parker, D. and Doonan, J.H. 2008. Brachypodium distachyon: making hay with a wild grass. *Trends in plant science,* 13(4), pp. 172-177.

Osborne, T.B., 1907. *The proteins of the wheat kernel* (No. 84). Carnegie institution of Washington.

Oscarsson, M., 1997. *New barley cultivars: chemical, microstructural and nutritional aspects focusing on carbohydrates*.

O'Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil', *Science,* 254(5031), pp. 539-544.

Păcurar, D.I., Thordal-Christensen, H., Păcurar, M.L., Pamfil, D., Botez, C. and Bellini, C. 2011. Agrobacterium tumefaciens: From crown gall tumors to genetic transformation. *Physiological and Molecular Plant Pathology,* 76(2), pp. 76-81.

Paolis, A., Sabatini, S., Pascalis, L., Costantino, P. and Capone, I. 1996. A rolB regulatory factor belongs to a new class of single zinc finger plant proteins. *The Plant Journal,* 10(2), pp. 215-223.

Park, S.H., Wilson, J.D., Chung, O.K. and Seib, P.A. 2004.Size distribution and properties of wheat starch granules in relation to crumb grain score of pup-loaf bread. *Cereal Chemistry,* 81(6), pp. 699-704.

Park, W., Li, J., Song, R., Messing, J. and Chen, X. 2002.CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Current Biology,* 12(17), pp. 1484-1495.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T. and Poliakov, A. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature,* 457(7229), pp. 551.

Pawson, T. and Nash, P. 2003. Assembly of cell regulatory systems through protein interaction domains. *Science,* 300(5618), pp. 445-452.

Pedro, A., Savin, R., Parry, M.A. and Slafer, G.A. 2012.Selection for high grain number per unit stem length through four generations from mutants in a durum wheat population to increase yields of individual plants and crops. *Field Crops Research,* 129, pp. 59-70.

Peng, F.Y. and Weselake, R.J. 2011. Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *BMC genomics,* 12(1), pp. 286

Penfield, S., Li, Y., Gilday, A.D., Graham, S. and Graham, I.A. 2006. Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *The Plant Cell,* 18(8), pp. 1887-1899.

Percival, J., 1921. The wheat plant. 463 pp. *London: Duckworth and Co*.

Pilpel, Y., Sudarsanam, P. and Church, G.M. 2001. Identifying regulatory networks by combinatorial analysis of promoter elements. Nature genetics, 29(2), pp. 153.

Pirovano, L., Lanzini, S., Hartings, H., Lazzaroni, N., Rossi, V., Joshi, R., Thompson, R.D., Salamini, F. and Motto, M. 1994. Structural and functional analysis of an Opaque-2-related gene from sorghum. *Plant Molecular Biology,* 24(3), pp. 515-523.

Piston, F., Uauy, C., Fu, L., Langston, J., Labavitch, J. and Dubcovsky, J. 2010. Down-regulation of four putative arabinoxylan feruloyl transferase genes from family PF02458 reduces ester-linked ferulate content in rice cell walls. *Planta,* 231(3), pp. 677-691.

Preston, J.C. and Kellogg, E.A. 2008. Discrete developmental roles for temperate cereal grass VERNALIZATION1/FRUITFULL-like genes in flowering competency and the transition to flowering. *Plant Physiology*, *146*(1), pp.265-276.

Pysh, L.D., Aukerman, M.J. and Schmidt, R.J. 1993. OHP1: a maize basic domain/leucine zipper protein that interacts with opaque2. *The Plant Cell,* 5(2), pp. 227-236.

Qiao, Z., Qi, W., Wang, Q., Feng, Y., Yang, Q., Zhang, N., Wang, S., Tang, Y. and Song, R. 2016. ZmMADS47 regulates zein gene transcription through interaction with Opaque2. PLoS genetics, 12(4), pp. e1005991.

Qu, L.Q. and Takaiwa, F. 2004. Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. *Plant Biotechnology Journal,* 2(2), pp. 113-125.

Quaglia, G.B. (1988) 'Other durum wheat products'.

Rabinovich, S.V. 1998. Importance of wheat-rye translocations for breeding modern cultivar of Triticum aestivum L. *Euphytica,* 100(1-3), pp. 323-340.

Radchuk, V.V., Borisjuk, L., Sreenivasulu, N., Merx, K., Mock, H., Rolletschek, H., Wobus, U. and Weschke, W. 2009. Spatiotemporal profiling of starch biosynthesis and degradation in the developing barley grain. *Plant Physiology,* 150(1), pp. 190-204.

Raeker, M.Ö., Gaines, C.S., Finney, P.L. and Donelson, T., 1998. Granule size distribution and chemical composition of starches from 12 soft wheat cultivars. *Cereal Chemistry*, *75*(5), pp.721-728.

Rashid, M., Guangyuan, H., Guangxiao, Y., Hussain, J. and Xu, Y. 2012. AP2/ERF transcription factor in rice: genome-wide canvas and syntenic relationships between monocots and eudicots. *Evolutionary Bioinformatics,* 8, pp. EBO. S9369.

Raikhel, N. 1992. Nuclear targeting in plants. *Plant Physiology,* 100(4), pp. 1627.

Ramsay, L., Comadran, J., Druka, A., Marshall, D.F., Thomas, W.T., Macaulay, M., MacKenzie, K., Simpson, C., Fuller, J. and Bonar, N. 2011. INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nature genetics,* 43(2), pp. 169.

Ravel, C., Martre, P., Romeuf, I., Dardevet, M., El-Malki, R., Bordes, J., Duchateau, N., Brunel, D., Balfourier, F. and Charmet, G. 2009. Nucleotide polymorphism in the wheat transcriptional activator Spa influences its pattern of expression and

has pleiotropic effects on grain protein composition, dough viscoelasticity, and grain hardness. *Plant Physiology,* 151(4), pp. 2133-2144.

Reddy, V.M. and Daynard, T.B. 1983. Endosperm characteristics associated with rate of grain filling and kernel size in corn.' *Maydica.*

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. 2002. MicroRNAs in plants, *Genes & development,* 16(13), pp. 1616-1626.

Regier, J. L., Shen, F. & Triezenberg, S. J. 1993. Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator. Proc Natl Acad Sci U S A, 90, 883-7.

Reynolds, M., Foulkes, M.J., Slafer, G.A., Berry, P., Parry, M.A., Snape, J.W. and Angus, W.J. 2009. Raising yield potential in wheat. *Journal of experimental botany,* 60(7), pp. 1899-1918.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. 2002. Prediction of plant microRNA targets. *Cell,* 110(4), pp. 513-520.

Rice Annotation Project .2007. The rice annotation project database (RAP-DB): 2008 update. *Nucleic acids research,* 36(suppl_1), pp. D1033.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J. and Samaha, R.R. 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science,* 290(5499), pp. 2105-2110.

Riechmann, J.L. and Meyerowitz, E.M. 1998.The AP2/EREBP family of plant transcription factors. *Biological chemistry,* 379, pp. 633-646.

Robbins, J., Dilwortht, S.M., Laskey, R.A. and Dingwall, C. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell,* 64(3), pp. 615-623.

Rodriguez-Uribe, L. and O'connell, M.A. 2006.A root-specific bZIP transcription factor is responsive to water deficit stress in tepary bean (Phaseolus acutifolius) and common bean (P. vulgaris). *Journal of experimental botany,* 57(6), pp. 1391-1398.

Robinson-Rechavi, M., Garcia, H.E. and Laudet, V. 2003. The nuclear receptor superfamily. *Journal of cell science*, *116*(4), pp.585-586.

Rogers, S.O. and Quatrano, R.S. 1983. Morphological staging of wheat caryopsis development. *American Journal of Botany,* 70(2), pp. 308-311.

Rügner, A., Frohnmeyer, H., Näke, C., Wellmer, F., Kircher, S., Schäfer, E. and Harter, K. 2001.Isolation and characterization of four novel parsley proteins that interact with the transcriptional regulators CPRF1 and CPRF2. *Molecular Genetics and Genomics,* 265(6), pp. 964-976.

Sabelli, P.A. and Larkins, B.A. 2009. The development of endosperm in grasses. *Plant Physiology,* 149(1), pp. 14-26.

Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A. and Quackenbush, J. 2006.TM4 microarray software suite', *Methods in enzymology,* 411, pp. 134-193.

Sainz, M.B., Grotewold, E. and Chandler, V.L. 1997.Evidence for direct activation of an anthocyanin promoter by the maize C1 protein and comparison of DNA binding by related Myb domain proteins.*The Plant Cell,* 9(4), pp. 611-625.

Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2002. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. *Biochemical and biophysical research communications,* 290(3), pp. 998-1009.

Saleh, A., 2003. Plant AP2/ERF transcription factors. *Genetika*, *35*(1), pp.37-50.

Salman, H., Blazek, J., Lopez-Rubio, A., Gilbert, E.P., Hanley, T. and Copeland, L. 2009. Structure–function relationships in A and B granules from wheat starches of similar amylose content. *Carbohydrate Polymers,* 75(3), pp. 420-427.

Sakamoto, T. and Matsuoka, M. 2008. Identifying and exploiting grain yield genes in rice. *Current opinion in plant biology,* 11(2), pp. 209-214.

Satoh, R., Fujita, Y., Nakashima, K., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2004.A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the ProDH gene in Arabidopsis. *Plant and Cell Physiology,* 45(3), pp. 309-317.

Schindler, U., Menkens, A.E., Beckmann, H., Ecker, J.R. and Cashmore, A.R. 1992. 'Heterodimerization between light-regulated and ubiquitously expressed Arabidopsis GBF bZIP proteins. *The EMBO journal,* 11(4), pp. 1261-1273.

Schlegel1, R. and Korzun, V., 1997. About the origin of 1RS. 1BL wheat-rye chromosome translocations from Germany. *Plant Breeding*, *116*(6), pp.537-540.

Schmidt, R.J., Burr, F.A. and Burr, B. 1987. Transposon tagging and molecular analysis of the maize regulatory locus opaque-2. *Science,* 238(4829), pp. 960-963.

Schmidt, R.J., Ketudat, M., Aukerman, M.J. and Hoschek, G. 1992. Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes.*The Plant Cell,* 4(6), pp. 689-700.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L. and Graves, T.A. 2009.The B73 maize genome: complexity, diversity, and dynamics', *Science,* 326(5956), pp. 1112-1115.

Schultz, E.A. and Haughn, G.W. 1993. Genetic analysis of the floral initiation process (FLIP) in Arabidopsis. *Development,* 119(3), pp. 745-765.

Sestili, F., Janni, M., Doherty, A., Botticella, E., D'Ovidio, R., Masci, S., Jones, H.D. and Lafiandra, D. 2010. Increasing the amylose content of durum wheat through silencing of the SBEIIa genes. *BMC Plant Biology,* 10(1), pp. 144.

Shannon, S. and Meeks-Wagner, D.R. 1993. Genetic interactions that regulate inflorescence development in Arabidopsis. *The Plant Cell,* 5(6), pp. 639-655.

She, K., Kusano, H., Koizumi, K., Yamakawa, H., Hakata, M., Imamura, T., Fukuda, M., Naito, N., Tsurumaki, Y. and Yaeshima, M. 2010. A novel factor FLOURY ENDOSPERM2 is involved in regulation of rice grain size and starch quality. *The Plant Cell,* pp. tpc. 109.070821.

Sherf, B.A., Navarro, S.L., Hannah, R.R. and Wood, K.V. 1996. Dual-luciferase reporter assay: an advanced co-reporter technology integrating firefly and Renilla luciferase assays.*Promega Notes,* 57(2).

Shewry, P.R. 2007.Improving the protein content and composition of cereal grain. *Journal of cereal science,* 46(3), pp. 239-250.

Shewry, P. R. 2009. Wheat. Journal Experimental Botany, 60(6), 1537-53.

Shewry, P.R. and Halford, N.G. 2002. Cereal seed storage proteins: structures, properties and role in grain utilization', *Journal of experimental botany,* 53(370), pp. 947-958.

Shewry, P.R., Halford, N.G. and Lafiandra, D. 2003.Genetics of wheat gluten proteins. Anonymous *Advances in genetics.* Elsevier, pp. 111-184

Shewry, P.R., Mitchell, R.A.C., Tosi, P., Wan, Y., Underwood, C., Lovegrove, A., Freeman, J., Toole, G.A., Mills, E.N.C. and Ward, J.L. 2012. An integrated study of grain development of wheat (cv. Hereward). *Journal of cereal science,* 56(1), pp. 21-30.

Shewry, P.R., Parmar, S. and Miflin, B.J.1983. Extraction, separation, and polymorphism of the prolamin storage proteins (secalins) of rye. *Cereal Chemistry (USA).*

Shewry, P.R. and Tatham, A.S. 1990.The prolamin storage proteins of cereal seeds: structure and evolution. *Biochemical journal,* 267(1), pp. 1.

Shibanuma, K., Takeda, Y., Hizukuri, S. and Shibata, S. 1994. Molecular structures of some wheat starches. *Carbohydrate Polymers,* 25(2), pp. 111-116.

Shigyo, M., Hasebe, M. and Ito, M. 2006.Molecular evolution of the AP2 subfamily. *Gene,* 366(2), pp. 256-265.

Shinde, S.V., Nelson, J.E. and Huber, K.C. 2003.Soft wheat starch pasting behavior in relation to A-and B-type granule content and composition. *Cereal Chemistry,* 80(1), pp. 91-98.

Sievert, D. and Pomeranz, Y. 1989.Enzyme-resistant starch. I. Characterization and evaluation by enzymatic, thermoanalytical, and microscopic methods. *Cereal Chem,* 66(4), pp. 342-347.

Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y., Gill, B.S. and Faris, J.D. 2006. 'Molecular characterization of the major wheat domestication gene Q. *Genetics,* 172(1), pp. 547-555.

Singh, A., Kumar, P., Sharma, M., Tuli, R., Dhaliwal, H.S., Chaudhury, A., Pal, D. and Roy, J. 2015. Expression patterns of genes involved in starch biosynthesis during seed development in bread wheat (Triticum aestivum). Molecular Breeding, 35(9), pp. 184.

Singh, K.B. 1998. Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiology,* 118(4), pp. 1111-1120.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R. and Ruvkun, G. 2000.The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Molecular cell,* 5(4), pp. 659-669.

Somers, D.J. and Gustafson, J.P. 1994.A review of wheat/rye genomic interactions and gene expression in triticale. *Genetics (Life Science Advances),* 13, pp. 57-63.

Song, C., Agarwal, M., Ohta, M., Guo, Y., Halfter, U., Wang, P. and Zhu, J. 2005. Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *The Plant Cell,* 17(8), pp. 2384-2396.

South, J.B. and Morrison, W.R., 1990. Isolation and analysis of starch from single kernels of wheat and barley. *journal of Cereal Science*, *12*(1), pp.43-51.

Sorrells, M.E., La Rota, M., Bermudez-Kandianis, C.E., Greene, R.A., Kantety, R., Munkvold, J.D., Mahmoud, A., Ma, X., Gustafson, P.J. and Qi, L.L. 2003. Comparative DNA sequence analysis of wheat and rice genomes. *Genome research,* 13(8), pp. 1818-1827.

Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.S., Gill, B.S., and Faris, J.D. 2006. Molecular characterization of the major wheat domestication gene q. Genetics 172, 547–555.

Sparkes, I.A., Runions, J., Kearns, A. and Hawes, C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature protocols,* 1(4), pp. 2019.

Srinivas, T. 1975. Pattern of crack formation in rice grain as influenced by shape and orientation of cells. *Journal of the science of food and agriculture,* 26(10), pp. 1479-1482.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* 22(21), pp. 2688-2690.

Stone, B., Morell, M.K., Khan, K. and Shewry, P.R. 2009. Wheat: Chemistry and Technology, Ed 4.

Strathmann, A., Kuhlmann, M., Heinekamp, T. and Dröge-Laser, W.2001. BZI-1 specifically heterodimerises with the tobacco bZIP transcription factors BZI-2, BZI-3/TBZF and BZI-4, and is functionally involved in flower development', *The Plant Journal,* 28(4), pp. 397-408.

Suckow, M., Schwamborn, K., Kisters-Woike, B., von Wilcken-Bergmann, B. and Müller-Hill, B., 1994. Replacement of invariant bzip residuces within the basic region of the yeast transcriptional activator GCN4 can Change its DNA binding specificity. *Nucleic acids research*, *22*(21), pp.4395-4404.

Suckow, M., Wilcken-Bergmann, B.v. and Müller-Hill, B. 1993.Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. *The EMBO journal,* 12(3), pp. 1193-1200.

Sullivan, P., Arendt, E. and Gallagher, E. 2012.The increasing use of barley and barley by- products in the production of healthier baked goods. *Trends in Food Science & Technology,* 29(2).

Sullivan, S. M., Horn, P. J., Olson, V. A., Koop, A. H., Niu, W., Ebright, R. H. & Triezenberg, S. J. 1998. Mutational analysis of a transcriptional activation region of the VP16 protein of herpes simplex virus. Nucleic Acids Res, 26, 4487-96.

Sun, G. 2012.MicroRNAs and their diverse functions in plants. *Plant Molecular Biology,* 80(1), pp. 17-36.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W. and Zhu, J. 2008.Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC plant biology,* 8(1), pp. 25.

Takaiwa, F. and Onodera, Y., National Institute of Agrobiological Sciences, National Agriculture and Bio-oriented Research Organization NARO, 2007. *Bzip type transcription factors regulating the expression of rice storage protein*. U.S. Patent 7,214,851.

Tang, H., Mitsunaga, T. and Kawamura, Y. 2006. Molecular arrangement in blocklets and starch granule architecture. *Carbohydrate Polymers,* 63(4), pp. 555-560.

Tang, M., Li, G. and Chen, M. 2007.The phylogeny and expression pattern of APETALA2-like genes in rice. *Journal of Genetics and Genomics,* 34(10), pp. 930-938.

Tester, R.F. 1997. Influence of growth conditions on barley starch properties. *International journal of biological macromolecules,* 21(1-2), pp. 37-45.

Tetlow, I.J., Morell, M.K. and Emes, M.J. 2004. Recent developments in understanding the regulation of starch metabolism in higher plants. *Journal of experimental botany,* 55(406), pp. 2131-2145.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997.The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research,* 25(24), pp. 4876-4882.

Tomlinson, K. and Denyer, K. 2003. Starch synthesis in cereal grains. *Advances in Botanical Research,* 40, pp. 1-61.

Tonoike, H., Han, I., Jongewaard, I., Doyle, M., Guiltinan, M. and Fosket, D.E. 1994. Hypocotyl expression and light downregulation of the soybean tubulin gene, tubB1. *The Plant Journal,* 5(3), pp. 343-351.

Toole, G.A., Wilson, R.H., Parker, M.L., Wellner, N.K., Wheeler, T.R., Shewry, P.R. and Mills, E. 2007.The effect of environment on endosperm cell-wall development in Triticum aestivum during grain filling: an infrared spectroscopic imaging study. *Planta,* 225(6), pp. 1393-1403.

Trafford, K., Haleux, P., Henderson, M., Parker, M., Shirley, N.J., Tucker, M.R., Fincher, G.B. and Burton, R.A. 2013. Grain development in Brachypodium and other grasses: possible interactions between cell expansion, starch deposition, and cell-wall synthesis. *Journal of experimental botany*, *64*(16), pp.5033-5047.

Travella, S., Klimm, T.E. and Keller, B. 2006.RNA interference-based gene silencing as an efficient tool for functional genomics in hexaploid bread wheat. *Plant Physiology,* 142(1), pp. 6-20.

Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A. and Nakamura, K. 2005.Analysis of a sugar response mutant of Arabidopsis identified a novel B3 domain protein that functions as an active transcriptional repressor. *Plant Physiology,* 138(2), pp. 675-685.

Ueda, T., Wang, Z., Pham, N. and Messing, J. 1994.Identification of a transcriptional activator-binding element in the 27-kilodalton zein promoter, the-300 element. *Molecular and cellular biology,* 14(7), pp. 4350-4359.

Ufaz, S. and Galili, G. 2008. Improving the content of essential amino acids in crop plants: goals and opportunities', *Plant Physiology,* 147(3), pp. 954-961.

Van Amelsvoort, J.M. and Weststrate, J.A., 1992. Amylose-amylopectin ratio in a meal affects postprandial variables in male volunteers. *The American journal of clinical nutrition*, *55*(3), pp.712-718.

Van Barneveld, R.J. & Cooper, K.V. 2002.Nutritional quality of triticale for pigs and poultry. *Proceedings of the 5th International Triticale Symposium, Radzików, Poland, 30 June-5 July, 2002. Volume I: oral presentations*Plant Breeding and Acclimatization Institute, pp. 277.

van der Krol, Alexander R and Chua, N. 1991.The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. *The Plant Cell,* 3(7), pp. 667-675.

Van Slageren, M.W., 1994. *Wild Wheats: A Monograph of Aegilops L. and Amblypyrum (Jaub. & Spach) Eig (Poaceae)*. Wageningen, the Netherlands: Agricultural University.

Varagona, M.J. and Raikhel, N.V. 1994.The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. *The Plant Journal,* 5(2), pp. 207-214.

Varagona, M.J., Schmidt, R.J. and Raikhel, N.V. 1992. Nuclear localization signal (s) required for nuclear targeting of the maize regulatory protein Opaque-2. *The Plant Cell,* 4(10), pp. 1213-1227.

Vetten, N.C. and Ferl, R.J. 1995.Characterization of a maize G-box binding factor that is induced by hypoxia. *The Plant Journal,* 7(4), pp. 589-601.

Vettore, A.L., Yunes, J.A., Neto, G.C., Da Silva, M.J., Arruda, P. and Leite, A. 1998.The molecular and functional characterization of an Opaque2 homologue gene from Coix and a new classification of plant bZIP proteins. *Plant Molecular Biology,* 36(2), pp. 249-263.

Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. and Schmidt, R.J. 1997.A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences,* 94(14), pp. 7685-7690.

Vincentz, M., Bandeira-Kobarg, C., Gauer, L., Schlögl, P. and Leite, A. 2003. Evolutionary pattern of angiosperm bZIP factors homologous to the maize Opaque2 regulatory protein. *Journal of Molecular Evolution,* 56(1), pp. 105-116.

Vinson, C.R., Hai, T. and Boyd, S.M. 1993. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes & development,* 7(6), pp. 1047-1058.

Vinson, C., Acharya, A. and Taparowsky, E.J. 2006.Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression,* 1759(1-2), pp. 4-12.

Vinson, C., Myakishev, M., Acharya, A., Mir, A.A., Moll, J.R. and Bonovich, M. 2002. Classification of human B-ZIP proteins based on dimerization properties. *Molecular and cellular biology,* 22(18), pp. 6321-6335.

Vollbrecht, E., Springer, P.S., Goh, L., Buckler IV, E.S. and Martienssen, R. 2005.Architecture of floral branch systems in maize and related grasses. *Nature,* 436(7054), pp. 1119.

Wagner, E.F., 2001. AP-1–Introductory remarks. *Oncogene*, *20*(19), p.2334.

Walther, D., Brunnemann, R. and Selbig, J. 2007. The regulatory code for transcriptional response diversity and its relation to genome structural properties in A. thaliana. *PLoS genetics*, *3*(2), p.e11.

Walsh, J.B. 1995.How often do duplicated genes evolve new functions?.*Genetics,* 139(1), pp. 421-428.

Wang, C., Tian, Z., Chen, L., Temelli, F., Liu, H. and Wang, Y., 2010. Functionality of barley proteins extracted and fractionated by alkaline and alcohol methods. *Cereal Chemistry*, *87*(6), pp.597-606.

Wang, H.L., Offler, C.E. and Patrick, J.W. 1994a.Nucellar projection transfer cells in the developing wheat grain.*Protoplasma,* 182(1-2), pp. 39-52.

Wang, H.L., Offler, C.E., Patrick, J.W. and Ugalde, T.D. 1994b.The cellular pathway of photosynthate transfer in the developing wheat grain. I. Delineation of a potential transfer pathway using fluorescent dyes. *Plant, Cell & Environment,* 17(3), pp. 257-266.

Wang, H. and Murphy, P.A. 1994.Isoflavone content in commercial soybean foods. *Journal of Agricultural and Food Chemistry,* 42(8), pp. 1666-1673.

Wang, J., Xu, H., Zhu, Y., Liu, Q. and Cai, X. 2013. OsbZIP58, a basic leucine zipper transcription factor, regulates starch biosynthesis in rice endosperm. *Journal of experimental botany,* 64(11), pp. 3453-3466.

Wang, K., Hasjim, J., Wu, A.C., Henry, R.J. and Gilbert, R.G. 2014. Variation in amylose fine structure of starches from different botanical sources. *Journal of Agricultural and Food Chemistry,* 62(19), pp. 4443-4453.

Wang, N. and Fisher, D.B. 1994c. The use of fluorescent tracers to characterize the post-phloem transport pathway in maternal tissues of developing wheat grains. *Plant Physiology,* 104(1), pp. 17-27.

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., Wang, Y., Chen, X., Zhang, Y. and Gao, C. 2015.The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nature genetics,* 47(8), pp. 949.

Wang, X., Reyes, J.L., Chua, N. and Gaasterland, T. 2004.Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome biology,* 5(9), pp. R65.

Wang, X. and Larkins, B.A. 2001.Genetic Analysis of Amino Acid Accumulation inopaque-2 Maize Endosperm. *Plant Physiology,* 125(4), pp. 1766-1777.

Waniska, R.D. 2000.Structure, phenolic compounds, and antifungal proteins of sorghum caryopses. *Technical and institutional options for sorghum grain mold management: proceedings of an international consultation*Citeseer, pp. 18.

Wei, C., Qin, F., Zhu, L., Zhou, W., Chen, Y., Wang, Y., Gu, M. and Liu, Q. 2009a. Microstructure and ultrastructure of high-amylose rice resistant starch granules modified by antisense RNA inhibition of starch branching enzyme. *Journal of Agricultural and Food Chemistry,* 58(2), pp. 1224-1232.

Wei, C., Zhang, J., Zhou, W., Chen, Y. and Xu, R. 2009b. Accumulation of storage protein and formation of protein body during wheat endosperm development. *Journal of Triticeae Crops,* 29(1), pp. 73-78.

Wei, K., Chen, J., Wang, Y., Chen, Y., Chen, S., Lin, Y., Pan, S., Zhong, X. and Xie, D. 2012. Genome- Wide Analysis of bZIP- Encoding Genes in Maize', *DNA Research,* 19(6), pp. 463-476.

Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schütze, K., Wang, X., Chaban, C., Hanson, J., Teige, M. and Harter, K. 2009a. Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Molecular Biology,* 69(1-2), pp. 107-119.

Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schütze, K., Wang, X., Chaban, C., Hanson, J., Teige, M. and Harter, K. 2009b.Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Molecular Biology,* 69(1-2), pp. 107-119.

Weltmeier, F., Rahmani, F., Ehlert, A., Dietrich, K., Schütze, K., Wang, X., Chaban, C., Hanson, J., Teige, M. and Harter, K. 2009c. Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Molecular Biology,* 69(1-2), pp. 107-119.

Wendel, J.F. 2000. Genome evolution in polyploids. Anonymous *Plant molecular evolution.* Springer, pp. 225-249.

Wicker, T., Mayer, K.F., Gundlach, H., Martis, M., Steuernagel, B., Scholz, U., Šimková, H., Kubaláková, M., Choulet, F. and Taudien, S. 2011. Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *The Plant Cell,* 23(5), pp. 1706-1718.

Wieser, H. and Koehler, P., 2008. The biochemical basis of celiac disease. *Cereal Chemistry*, *85*(1), pp.1-13.

Wightman, B., Ha, I. and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell,* 75(5), pp. 855-862.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Prüß, M., Reuter, I. and Schacherer, F., 2000. TRANSFAC: an integrated system for gene expression regulation. *Nucleic acids research*, *28*(1), pp.316-

319.Williams, M.E., Foster, R. and Chua, N. (1992) 'Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding.', *The Plant Cell,* 4(4), pp. 485-496.

Wolfe, S.A., Nekludova, L. and Pabo, C.O. 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annual Review of Biophysics and Biomolecular Structure,* 29(1), pp. 183-212.

Woo, Y., Hu, D.W., Larkins, B.A. and Jung, R. 2001.Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *The Plant Cell,* 13(10), pp. 2297-2317.

Wu, G., Park, M.Y., Conway, S.R., Wang, J., Weigel, D. and Poethig, R.S. 2009.The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. Cell, 138(4), pp. 750-759.

Wu, Y.V., Stringfellow, A.C., Anderson, R.A., Sexson, K.R. and Wall, J.S. 1978.Triticale for food uses. *Journal of Agricultural and Food Chemistry,* 26(5), pp. 1039-1048.

Wu, Y. and Messing, J. 2012.RNA interference can rebalance the nitrogen sink of maize seeds without losing hard endosperm.*PLoS One,* 7(2), pp. e32850.

Wu, Z., Li, X., Liu, Z., Li, H., Wang, Y. and Zhuang, J. 2015.Transcriptome-based discovery of AP2/ERF transcription factors related to temperature stress in tea plant (Camellia sinensis). *Functional & integrative genomics,* 15(6), pp. 741-752.

Xu, C., Gao, Y., Tian, B., Ren, J., Meng, Q. and Wang, P. 2017.Effects of EDAH, a novel plant growth regulator, on mechanical strength, stalk vascular bundles and grain yield of summer maize at high densities.*Field Crops Research,* 200, pp. 71-79.

Xu, J. and Messing, J. 2008. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proceedings of the National Academy of Sciences,* 105(38), pp. 14330-14335.

Yalçın, E., Çelik, S. and İbanoğlu, E., 2008. Foaming properties of barley protein isolates and hydrolysates. *European Food Research and Technology*, *226*(5), p.967.

Yamamori, M., Fujita, S., Hayakawa, K., Matsuki, J. and Yasui, T. 2000. Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. *Theoretical and Applied Genetics,* 101(1-2), pp. 21-29.

Yamamoto, M.P., Onodera, Y., Touno, S.M. and Takaiwa, F. 2006.Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiology,* 141(4), pp. 1694-1707.

Yamaguchi, A. and Abe, M. 2012.Regulation of reproductive development by non-coding RNA in Arabidopsis: to flower or not to flower. Journal of Plant Research, 125(6), pp. 693-704.

Yan, H., Pan, X., Jiang, H. and Wu, G. 2009.Comparison of the starch synthesis genes between maize and rice: copies, chromosome location and expression divergence. *Theoretical and Applied Genetics,* 119(5), pp. 815-825.

Yanagisawa, S. 1996. Dof DNA-binding proteins contain a novel zinc finger motif. *Trends in plant science,* 1(7), pp. 213-214.

Yang, J., Ji, C. and Wu, Y., 2016. Divergent transactivation of maize storage protein zein genes by the transcription factors Opaque2 and OHPs. *Genetics*, pp.genetics-116.

Yang, Y., Lai, K., Tai, P. and Li, W. 1999.Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between Brassica and other angiosperm lineages. *Journal of Molecular Evolution,* 48(5), pp. 597-604.

Yoshikawa, M. 2013. Biogenesis of trans-acting siRNAs, endogenous secondary siRNAs in plants. *Genes & genetic systems,* 88(2), pp. 77-84.

Young, T.E. and Gallie, D.R. 1999. Analysis of programmed cell death in wheat endosperm reveals differences in endosperm development between cereals. *Plant Molecular Biology,* 39(5), pp. 915-926.

Zadoks, J.C., Chang, T.T. and Konzak, C.F., 1974. A decimal code for the growth stages of cereals. *Weed research*, *14*(6), pp.415-421.

Zhang, L., Li, Z., Quan, R., Li, G., Wang, R. and Huang, R. 2011. An AP2 domain-containing gene, ESE1, targeted by the ethylene signaling component EIN3 is important for the salt response in Arabidopsis. *Plant Physiology,* 157(2), pp. 854-865.

Zhang, Q., Chen, Q., Wang, S., Hong, Y. and Wang, Z. 2014. Rice and cold stress: methods for its evaluation and summary of cold tolerance-related quantitative trait loci. *Rice,* 7(1), pp. 24.

Zhang, Z., Yang, J. and Wu, Y. 2015. Transcriptional regulation of zein gene expression in maize through the additive and synergistic action of opaque2, prolamine-box binding factor, and O2 heterodimerizing proteins. *The Plant Cell,* pp. tpc. 15.00035.

Zhang, Z., Zheng, X., Yang, J., Messing, J. and Wu, Y. 2016.Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. *Proceedings of the National Academy of Sciences,* 113(39), pp. 10842-10847.

Zhao, L., Pan, T., Cai, C., Wang, J. and Wei, C. 2016. Application of whole sections of mature cereal seeds to visualize the morphology of endosperm cell and starch and the distribution of storage protein. *Journal of Cereal Science*, *71*, pp.19-27.

Zhao, W., Gatehouse, J.A. and Boulter, D. 1983. The purification and partial amino acid sequence of a polypeptide from the glutelin fraction of rice grains; homology to pea legumin. *FEBS letters,* 162(1), pp. 96-102.

Zhou, C., Guo, J., Feng, Z., Cui, X. and Zhu, J. 2012. Molecular characterization of a novel AP2 transcription factor ThWIND1-L from Thellungiella halophile. *Plant Cell, Tissue and Organ Culture (PCTOC),* 110(3), pp. 423-433.

Zhou, M.X., 2009. Barley production and consumption. In *Genetics and Improvement of Barley Malt Quality* (pp. 1-17). Springer, Berlin, Heidelberg.

Zhou, Z., Wang, L., Li, J., Song, X. and Yang, C. 2009. Study on programmed cell death and dynamic changes of starch accumulation in pericarp cells of Triticum aestivum L. *Protoplasma,* 236(1-4), pp. 49-58.

Zhu, G., Ye, N., Yang, J., Peng, X. and Zhang, J. 2011. Regulation of expression of starch synthesis genes by ethylene and ABA in relation to the development of rice inferior and superior spikelets. Journal of experimental botany, 62(11), pp. 3907-3916.

Zhu, Q. and Helliwell, C.A. 2010. Regulation of flowering time and floral patterning by miR172. Journal of experimental botany, 62(2), pp. 487-495.

Zhu, Q., Upadhyaya, N.M., Gubler, F. and Helliwell, C.A. 2009. Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (Oryza sativa). *BMC Plant Biology,* 9(1), pp. 149.

Zhu, Y., Cai, X., Wang, Z. and Hong, M. 2003. An interaction between a MYC protein and an EREBP protein is involved in transcriptional regulation of the rice Wx gene. *Journal of Biological Chemistry,* 278(48), pp. 47803-47811.

Zobel, H.F. 1988. Molecules to granules: a comprehensive starch review. *Starch-Stärke,* 40(2), pp. 44-50.

Zumajo-Cardona, C. and Pabón-Mora, N. 2016. Evolution of the APETALA2 gene lineage in seed plants. *Molecular biology and evolution*, *33*(7).