Proteini MATH-BTB u pšenice i njihova uloga u ranoj embriogenezi

Škiljaica, Andreja

Master's thesis / Diplomski rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:491285

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

University of Zagreb
Faculty of Science
Division of Biology

Andreja Škiljaica

Wheat MATH-BTB proteins and their role in early embryogenesis

Graduation thesis

Zagreb, 2016.

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Andreja Škiljaica

Proteini MATH-BTB u pšenice i njihova uloga u ranoj embriogenezi

Diplomski rad

Zagreb, 2016.

Ovaj rad, izrađen pri Zavodu za molekularnu biologiju, pod vodstvom izv. prof. dr. sc. Dunje Leljak-Levanić, predan je na ocjenu Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra molekularne biologije.

ACKNOWLEDGMENTS

I wish to thank my parents, for their lifelong support and for instilling in me the desire to learn. I am thankful to Professor Dunja Leljak-Levanić, for being a terrific mentor and teaching me the importance of being imaginative. Last but not least, many thanks to Professor Nataša Bauer for all the tips and tricks and to Ana, for details that make a difference.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek Diplomski rad

Proteini MATH-BTB u pšenice i njihova uloga u ranoj embriogenezi

Andreja Škiljaica
Rooseveltov trg 6, 10 000 Zagreb, Hrvatska

Izolacijom jajne stanice, zigote i dvostaničnih proembrija iz embrionske vreće pšenice (*Triticum aestivum* L.) dobivene su pouzdane biblioteke cDNA, na temelju kojih je uočena prisutnost proteina TaMAB2 tijekom razvoja zigote i dvostraničnog embrija. Protein TaMAB2 pripada porodici proteina s dvije domene: MATH (eng. *Meprin and TRAF homology*) i BTB (eng. *Brick-a-brack/tramtrack/broad complex*), čiji su predstavnici široko rasprostranjeni u eukariotima. Na temelju aminokiselinske sekvence proteina TaMAB2 pretražene su baze podataka NCBI i EnsemblPlants, otkrivajući 46 novih proteina MATH-BTB pšenice. Filogenetička analiza gena *MATH-BTB* pšenice, riže (*Oryza sativa*), kukuruza (*Zea mays*) i uročnjaka (*Arabidopsis thaliana* L.), provedena u programima Clustal X i Seaview otkrila je grupiranje gena *TaMAB2* u subkladu E3. Sravnjenje odabranih aminokiselinskih slijedova u programima MUSCLE i Boxshade pokazalo je veliku sličnost u sekvencama svih proteina subklade E3, ukazujući na moguću funkcionalnu sličnost. Analiza potencijalnih interakcijskih partnera proteina TaMAB2 sustavom dvaju kvašćevih hibrida otkrila je protein katanin kao mogući fiziološki supstrat proteina TaMAB2. Analiza fenotipa transgene linije uročnjaka s prekomjernom ekspresijom gena *TaMAB2* rezultirala je inhibicijom rasta i razvoja odraslih biljaka te smanjenom produkcijom somatskih embrija u kulturi.

(57 stranica, 11 slika, 4 tablice, 3 tablice u prilogu, 56 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici.
Ključne riječi: filogenija proteina MATH-BTB, proteinske interakcije, TaMAB2, katanin, somatska embriogeneza
Voditelj: dr. sc. Dunja Leljak-Levanić, izv. prof.
Ocjenitelji: dr. sc. Dunja Leljak-Levanić, izv. prof.
dr. sc. Željka Vidaković-Cifrek, izv. prof.
dr. sc. Sofia Ana Blažević, doc.
Rad prihvaćen: 21. lipnja 2016.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Division of Biology

Graduation thesis

Wheat MATH-BTB proteins and their role in early embryogenesis

Andreja Škiljaica
Rooseveltov trg 6, 10 000 Zagreb, Croatia

The isolation of egg cells, zygotes and two-celled proembryos from embryo sacs of wheat (*Triticum aestivum* L.) resulted in trustworthy cDNA libraries, based on which the presence of TaMAB2 protein was detected in the zygote and two-celled embryo. TaMAB2 protein belongs to a family of proteins with two domains: MATH (Meprin and TRAF homology) and BTB (Brick-a-brack/tramtrack/broad complex). MATH-BTB family is widely represented among eukaryotes. NCBI and EnsemblPlants databases were searched using the amino acid sequence of TaMAB2 protein as query, revealing 46 novel wheat MATH-BTB proteins. Using Clustal X and Seaview, a phylogenetic analysis of *MATH-BTB* genes of wheat, rice (*Oryza sativa*), maize (*Zea mays*) and *Arabidopsis thaliana* was conducted, with TaMAB2 protein clustering into the E3 subclade. Multiple sequence alignments of selected proteins in MUSCLE and Boxshade indicated a significant sequence similarity between all proteins of the E3 subclade, which suggesed their possible functional similarity. A yeast two-hybrid assay was conducted in order to analyze potential interaction partners of TaMAB2 protein. This revealed Katanin protein as a plausible physiological substrate of TaMAB2. Phenotype analysis of *TaMAB2* overexpression line of arabidopsis indicated inhibited growth and development of adult arabidopsis plants as well as decreased frequency of somatic embryogenesis in in culture.

(58 pages, 11 figures, 4 tables, 3 supplemental tables, 56 references, original in: English)

Thesis deposited in the Central Biological Library.
Key words: MATH-BTB phylogeny, protein - protein interactions, TaMAB2, Katanin, somatic embryogenesis
Supervisor: Dr. Sc. Dunja Leljak-Levanić, Assoc. Prof.
Reviewers: Dr. Sc. Dunja Leljak-Levanić, Assoc. Prof.
Dr. Sc. Željka Vidaković-Cifrek, Assoc. Prof.
Dr. Sc. Sofia Ana Blažević, Asst. Prof.
Thesis accepted: June 21st, 2016

Table of Contents

1 INTRODUCTION 1

1.1 Plant embryogenesis 1

1.1.1 Zygotic embryogenesis 1

1.1.2 Somatic embryogenesis 4

1. 2 MATH-BTB protein family 7

1.2.1 Phylogeny of MATH-BTB protein family 8

1.3 Cullin-based E3 ligases 9

1.4 Functional characterization of animal MATH-BTB proteins 13

1.4.1 MEL-26 protein from *Caenorhabditis elegans* 13

1.4.2 Human SPOP protein 14

1.5 Functional characterization of plant MATH-BTB proteins 15

1.5.1 Maize ZmMAB1 protein 15

1.5.2 Arabidopsis AtBPM proteins 16

1.5.3 Wheat TaMAB2 protein 16

1.6 Thesis objective 17

2 MATERIALS AND METHODS 18

2.1 Materials 18

2.1.1 Plant material and growth conditions 18

2.1.1.1 *Arabidopsis thaliana* (L.) Heynh. 18

2.1.1.2 Wheat, *Triticum aestivum* 18

2.1.2 Yeast, *Saccharomyces cerevisiae* strain 18

2.1.3 Bacteria strains 18

2.1.4 Primer sequences 19

2.2 Methods 19

2.2.1 Bioinformatics and phylogenetic analysis 19

2.2.2 Generation of total RNA and cDNA synthesis 19

2.2.3 Generation of mRNA and cDNA synthesis from minute amount of tissue 20

2.2.4 Standard PCR reaction 20

2.2.5 Generation of plasmid constructs 21

2.2.5.1 Generation of PCR inserts 21

2.2.5.2 Plasmid preparation and cloning procedure 23

2.2.6 Transformation of competent cells 23

2.2.7 Isolation of plasmid DNA, restriction analysis and gel-electrophoresis 24
2.2.7.1 Isolation of plasmid DNA 24
2.2.7.2 Restriction analysis and gel-electrophoresis 24
2.2.8 Yeast-two-hybrid screen 24
2.2.8.1 Co-transformation of yeast cells 24
2.2.8.2 Histidine prototrophy assay 26
2.2.8.3 β-galactosidase assay using X-gal as substrate 27
2.2.8.4 Liquid culture assay using ONPG as substrate 27
2.2.9 Arabidopsis zygotic embryo isolation 28
2.2.10 Induction of somatic embryogenesis 29
2.2.11 Microscopy 30
3 RESULTS 31
3.1 Identification of TaMAB2 homologs 31
3.1.1 TaMAB2 paralogs 31
3.1.2 TaMAB2 plant homologs 33
3.2 Phylogenetic tree of selected plant MATH-BTB proteins 33
3.3 Protein alignment of TaMAB2 and selected homologs 34
3.4 Phenotype analysis of *TaMAB2* overexpression and wild type arabidopsis plants 41
3.5 Isolation of zygotic embryos from *TaMAB2* overexpression and wild type arabidopsis plants 43
3.6 Somatic embryogenesis in *TaMAB2* overexpression and wild type arabidopsis plants 44
3.7 Protein interaction analysis 44
3.7.1 Cloning of selected genes in Y2H system 44
3.7.2 Katanin is a potential substrate of TaMAB2 47
4 DISCUSSION 48
4.1 Phylogeny of wheat MATH-BTB proteins and selected homologs 48
4.2 TaMAB2 interacts with Katanin and itself 50
4.3 *TaMAB2* overexpression has an ambiguous effect on arabidopsis development 51
5 CONCLUSION 53
6 REFERENCES 54
7 APPENDIX i
CURRICULUM VITAE v

1 INTRODUCTION

1.1 Plant embryogenesis

Embryogenesis is a crucial developmental process within the life cycle of plants. In flowering plants, sexual embryogenic development involves double fertilization, resulting in the simultaneous formation of the embryo and the endosperm. Once activated by fertilization, the egg cell undergoes a complex series of morphological, cellular and molecular changes resulting in the formation of a two-celled proembryo (Willemsen & Scheres, 2004). In many vascular plants, sexual reproduction alternates with, or is replaced by different kinds of asexual reproduction. Adventitious embryos can develop from any other type of somatic cells via the process referred to as somatic embryogenesis (SE). SE is considered to be developmental reprogramming of somatic cells toward the embryogenic pathway (Fehér et al., 2003; Fehér, 2005) followed by development through characteristic morphological stages that resemble zygotic embryo development.

1.1.1 Zygotic embryogenesis

Zygotic embryogenesis takes place in the ovule after fusion of male and female gametes and the consequent formation of the unicellular zygote. This single totipotent cell ultimately develops into a mature embryo with the same basic tissue pattern as any post-embryonic plant: ground tissue, vascular tissue (in vascular plants) and epidermis. The intricacies of this process can vary greatly between species. The embryos of most flowering plants species undergo a series of seemingly chaotic cell divisions (reviewed in Hove et al., 2015), whereas some species, such as the *Brassicaceae* family species (including *Arabidopsis thaliana*) display a pattern of highly organised and predictable cells divisions (Mansfield & Briarty, 1991; Yoshida et al., 2014). Nevertheless, the developed seedling will in both cases comprise a stereotypical organisation, with primary shoot and root meristems, one or two cotyledons and a radial tissue pattern (reviewed in Hove et al., 2015). The central characteristic of this process is cellular diversity, i.e. the difference in morphology, function and identity of newly generated cells, achieved almost universally through asymmetric division. To generate unequal daughter cells, the dividing cell needs to have

mechanisms for positioning its division plane and these mechanisms must be closely coordinated with distribution of cell fate-specifying factors (reviewed in Abrash & Bergmann, 2009). Due to its regularity, the embryonic development of *Arabidopsis thaliana* (Fig 1) has been studied in great detail. Already extensive, this knowledge has been further improved by a recent three-dimensional description of cellular patterns and volumes during early arabidopsis embryogenesis (Yoshida et al., 2014).

One of the hallmarks of zygotic embryogenesis of arabidopsis is its observable polarity, which starts even before fertilization occurs. The polarity of the embryo sac is due to a gradient of natural auxin indole-acetic acid along the micropyle-chalaza axis, with the highest levels of accumulation at the micropylar region outside the embryo sac. Following this pattern, the egg cell is also highly polarized, with its nucleus at the chalazal pole (Pagnussat et al., 2009). The first asymmetric division of the zygote gives rise to two daughter cells: the basal and the apical cell. The upper apical cell is smaller, densely cytoplasmic and will eventually give rise to most of the embryo. Through two rounds of longitudinal divisions, the apical cell gives rise to four cells of equal size. Then, a transverse division produces two tiers of four cells each and the octant stage embryo is developed (reviewed in Hove et al., 2015). These tiers make up the apical and central regions of the embryo, which will later give rise to the shoot apical meristem (SAM) and hypocotyl, respectively. At the dermatogen stage, a round of tangential divisions of all cells in both tiers will create two radial layers of cells: the outer protoderm and the inner cell mass (reviewed in DeSmet & Beekman, 2011). Protoderm cells (precursors of the epidermis) continue to divide anticlinally. The inner cells divide longitudinally, so that the four basal cells form the ground tissue precursors and the smaller inner cells form the early vascular precursors. At the globular embryo stage, periclinal cell divisions at the two flanks give rise to the cotyledon primordia. At the late heart stage, the three layers of the shoot meristem – L1, L2, L3 – can be distinguished between the growing cotyledons (reviewed in Boscá et al., 2011; DeSmet & Beekman, 2011; Hove et al., 2015). Of the zygote's two daughter cells, the lower basal cell has a completely different fate. The basal cell is large, highly vacuolated and through a series of transversal divisions, it gives rise to the extra-embryonic suspensor, a column of six to nine cells. During the early to mid-globular stage, the suspensor's uppermost cell (hypophysis) undergoes an asymmetric division which produces two daughter cells. The smaller, lens-shaped cell will subsequently become the organising centre (OC) of the root apical meristem and the larger basal

cell will give rise to distal stem cells of the root meristem (reviewed in Boscá et al., 2011; DeSmet & Beekman, 2011; Hove et al., 2015).

Beside structural analysis of arabidopsis embryogenesis, genetic studies have shed some additional light on this process. Several genes involved in the regulation of asymmetric divisions during embryonic development have been elucidated. One example is the SHORT SUSPENSOR (*SSP*) gene, important for the first asymmetric division of the zygote. Its mRNA comes from pollen and it encodes an interleukin-1 receptor-associated kinase/Pelle-like kinase, genetically upstream of YODA (a MAPKK kinase) and GROUNDED (downstream transcription factor). The absence of either SSP, YODA or GROUNDED leads to a more symmetric division of the zygote and mis-specification of suspensor fate. Mutations in *WRKY2* gene also lead to a more symmetric division of the zygote, specifically through lack of both the nucleus shifting to the upper half of the zygote and accumulation of the vacuole in the lower half (reviewed in Boscá et al., 2011).

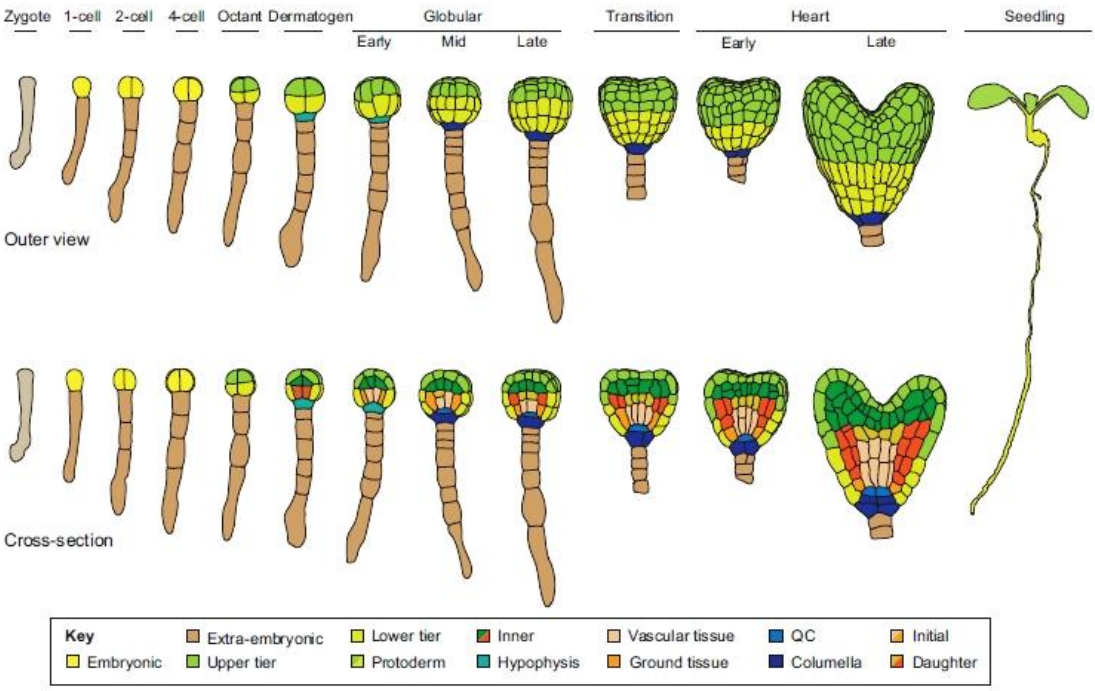


Figure 1. Arabidopsis embryo development. Outer view and cross-section of the arabidopsis embryo through all stages of development. Tissue types are colour coded according to the key. (Image acquired from Hove et al., 2015.)

Furthermore, extensive research has elucidated genes which become unequally expressed between daughter cells, where they help to maintain their separate cell fates. Among the best known are the Wuschel homeobox (*WOX*) genes, a plant-specific subclade of the eukaryotic homeobox transcription factor superfamily, characterized by the presence of a conserved DNA-binding homeodomain (reviewed in Graaf et al., 2009).

The product of the aforementioned *WRKY2* gene directly activates the transcription of *WOX8* and the redundant *WOX9* in the zygote. Post-division, the expression of *WOX8/9* continues in the basal lineage only. *WOX2* gene is expressed in the zygote, but is later restricted to the apical cell and subsequently to the upper half of the dermatogen embryo. Therefore, initially coexpressed in the zygote, *WOX2* and *WOX8* act as complementary cell fate regulators in the apical and basal lineage, respectively (Breuninger et al., 2008). *WOX8/9* also act non-cell autonomously in development of the apical lineage where they act as inducers of *WOX2* expression. In turn, *WOX2*, *WOX1*, *WOX3* and *WOX5* act redundantly in the apical lineage where they induce the expression of auxin efflux carrier *PIN-FORMED (PIN 1)* which helps form the auxin response maxima in the proembryo and consequently a polar body axis (Breuninger et al., 2008; reviewed in Boscá et al., 2011). Another *WOX* gene, *WUSCHEL* (*WUS*) is expressed in a small group of L3 cells called the organising center (OC) of SAM. WUS activates the expression of *CLAVATA3* (*CLV3*) in a group of stem cells adjacent to the OC. In a negative feedback loop, CLV3 binds several receptor-like kinases, including CLAVATA1 (CLV1), which then act inhibitory on *WUS* expression. Consequently, a small pool of stem cells is constantly kept undifferentiated and cells that exit this stem-cell niche become recruited into differentiation pathways which will give rise to different tissue types (reviewed in Boscá et al., 2011; Graaf et al., 2009). Although a detailed genetic background of arabidopsis embryogenesis exceeds the scope of this short introduction, the examples above provide a conceptual framework on how differential distribution of transcription factors and negative feedback loops help to maintain plant cells physiologically and functionally diverse.

1.1.2 Somatic embryogenesis

Somatic embryogenesis (SE) is a process in which differentiated plant cells regain totipotency and subsequently develop into functional embryos. Although this plant-specific

phenomenon is widely used for *in vitro* plant propagation, the underlying biological processes which occur during somatic embryogenesis are still being investigated. Regarding the ability of only certain genotypes and only certain cells to undergo somatic embryogenesis, Fehér (2008) hypothesises that "the genetically determined embryogenic potential may allow the expression of embryogenic competence under appropriate conditions which will result in the initiation of embryo development in response to an appropriate developmental signal in those cells only where the physiological conditions are favourable".

Somatic and zygotic embryogenesis are highly similar in that they both start with a totipotent cell which gives rise to a whole new plant, and the difference lies in the lack of sexual processes during somatic embryogenesis. In theory, somatic embryos have a single-cell origin, just like their zygotic counterparts. For instance, Portillo et al. (2007) reported the unicellular origin of somatic embryos in the monocotyledonous species *Agave tequilana* Weber var. Azul. Moreover, the events following the first asymmetric division of this single somatic cell closely resembled the events in zygotic embryogenesis of *Agave tequilana*. The initial somatic cell was highly polarized with its nucleus positioned to one pole, leaving the other pole highly vacuolated. The first division was transversal and asymmetric, yielding a smaller apical and larger basal cell. The second division of the apical cell produced the two-celled proembryo (embryo proper) and the basal cell was putatively the first cell of the suspensor. Further divisions gave rise to the four-celled embryo head and a two-celled suspensor. Later divisions of the suspensor ultimately produced the hypophyseal region where the plant radicle would be formed. Through a series of coordinated cell divisions, a globular stage of a somatic embryo was formed (Portillo et al., 2007). Generally speaking, once the induction of embryogenic state is complete, the patterning mechanisms which lead to formation of a zygotic-like embryo are common to all other forms of embryogenesis. Thus, both zygotic and somatic embryos will go through the globular, heart-shaped and torpedo-shaped stages in dicots, and globular, scutellar and coleoptilar stages in monocots (reviewed in Jiménez, 2001).

A pattern of distinct events has been observed during somatic embryogenesis. The first phase of SE is the induction (or initiation) stage in which differentiated cells become competent for embryogenesis. These cells have specific characteristics, such as early activation of the division cycle, more alkalic vacuolar pH, an altered auxin metabolism, and a non-functional chloroplast (reviewed in Zavatierri et al., 2010). In practice, cells coming from younger tissues, especially immature zygotic embryos, show the highest levels of competence for SE. However, stems, roots

and leaves have shown to be useful as well (reviewed in Gutiérrez-Mora et al., 2012). Induction of competence happens either directly or indirectly through a dedifferentiation step which usually involves a callus phase (reviewed in Jimenez et al., 2001; Fehér et al., 2003; Zavatierri et al., 2010, among others). There are two major categories of agents used for induction of SE: internal and/or external plant growth regulators (reviewed in Vondráková et al., 2016) and stress factors, such as osmotic shock, culture medium dehydration, water stress, heavy metal ions, alterations of culture medium pH, heat or cool shock treatments, hypoxia, antibiotics, ultraviolet radiation, and mechanical or chemical treatments (reviewed in Zavatierri et al, 2010). The induction phase is followed by expression of SE, i.e. the development of competent cells or proembryos. The last step is maturation, when somatic embryos prepare for germination through accumulation of reserves and desiccation (reviewed in Fehér et al., 2003; Zavatierri et al., 2010). To achieve all this, the pattern of gene expression present in differentiated somatic cells must be replaced with a new embryogenic gene expression pattern. The studies of various arabidopsis mutants revealed some of the genes potentially involved in SE. Genes encoding transcription factors Leafy Cotyledon 1 and 2 (LEC1 and 2) are present during zygotic embryogenesis and their overexpression may induce ectopic embryo development in leaf cells (Lotan et al., 1998). Another example is the *pickle* gene, which encodes a chromatin remodelling ATPase. Its activity is necessary for suppression of embryogenic gene expression program in somatic cells. *Pickle* mutants, therefore, exhibit embryo development in place of some of the root meristems (reviewed in Fehér, 2008). This finding led to the hypothesis that, from a genetic perspective, embryogenic potential in somatic cells comes from a "lack of suppression", rather than "induction" of embryogenesis (reviewed in Fehér, 2008).

Understanding the initiation phase of SE might help to elucidate the processes such as acquirement of totipotency and regulation of developmental switches. As was noted earlier, different inducers can be used to initiate SE. However, the most widely used inducer is auxin, especially 2,4-dichlorophenoxyacetic acid (2,4-D) (reviewed in Fehér et al., 2003). Auxin treatment, as well as the majority of treatments used for induction of SE, affects the auxin balance within the cells. After this phase, most treatments drop out the auxin, which indicates that cells soon become capable of self-supporting auxin synthesis (reviewed in Fehér, 2008). Auxin treatment (or stress treatment) might serve as a strong non-specific signal which helps to release

the embryogenic gene expression program from its usual state of chromatin-mediated repression in somatic cells (reviewed in Fehér, 2008).

1. 2 MATH-BTB protein family

Functional analysis of MATH and BTB domains, in their respective protein families, show that both are capable of a variety of protein-protein interactions (Stogois et al., 2005, reviewed in Zapata et al., 2007). They serve in a wide array of cellular processes where they help regulate development and homeostasis (reviewed in Zapata et al., 2007). Proteins encompassing a MATH domain and a BTB/POZ domain are broadly represented among eukaryotes. They are found in lower eukaryotes, such as Trypanosomatidae (Euglenozoa), and Coccidia (Alveolata), plants (Viridiplantae), including both eudicotyledons and monocotyledons (Liliopsidae), and in metazoa (reviewed in Zapata et al., 2007). However, MATH/BTB proteins have not yet been found in fungi (Juranić & Dresselhaus, 2014). MATH-BTB proteins usually contain an N-terminal MATH domain and a C-terminal BTB domain, but in some cases, the MATH domain appears on the C-terminus, and the BTB domain on the N-terminus. Also, some proteins contain MATH/BTB domains in tandem (reviewed in Zapata et al., 2007).

The BTB (or POZ) domain was first discovered as a conserved motif in the *Drosophila melanogaster* Bric-à-Brac, Tramtrack and Broad Complex transcription regulators (Zollman et al., 1994) and in many Pox Virus Zinc finger proteins (Koonin et al., 1992). BTB proteins have various roles, such as transcription repression, cytoskeleton regulation, protein ubiquitination/degradation and many others (Stogois et al., 2005 and references cited therein). Nevertheless, BTB proteins can be functionally classified into two major groups: those which participate in BTB domain-based protein-protein interactions and those which regulate transcription through DNA binding (Stogois et al., 2005). The BTB domain is around 100 amino acids long and is present in one or two copies in BTB proteins, along with one or two other types of domains, including zinc finger (ZF), Kelch, BTB and C-terminal Kelch (BACK), meprin and TRAF homology (MATH), ankyrin repeats (ANK), PHR, and Ras homology (Rho). Based on the presence of these additional domains, several BTB protein subfamilies can be discerned: BTB-only proteins, BTB-ZF proteins, BTB-Kelch proteins, BTB-BACK proteins, BTB-BACK-Kelch proteins, MATH-BTB proteins, BTB-ANK proteins, BTB-BACK-PHR proteins and Rho-BTB proteins (Stogois et al, 2005; Cheng et al., 2014 and references cited therein). The BTB domain facilitates homodimeric, heteromeric and

homotetrameric protein-protein interactions (Stogois et al., 2005). In their analysis of protein architecture, genomic distribution and sequence conservation of BTB proteins found in 17 fully sequenced eukaryotes, Stogois et al. (2005) reported little similarity in sequence of BTB domains from different protein families, despite the well-conserved tertiary structure. As a result, BTB proteins from different families can participate in a variety of protein-protein interactions.

TNF-Receptor Associated Factors (TRAFs) are proteins originally recognized as regulators and interactive partners of different members of the Tumor Necrosis Factor Receptor (TNFR) family. TRAFs are structurally characterized by a C-terminal region encompassing about 180 amino acids, forming a 7-8 antiparallel β-sheets fold (TRAF-C domain) preceded by a coiled coil (TRAF-N) domain (reviewed in Zapata et al., 2007). Meprins, a family of extracellular metalloproteases, contain a C-terminal domain with a high sequence homology to the TRAF-C domain. Consequently, the TRAF-C domain is commonly known as the MATH domain (Marín, 2015). Proteins encompassing a MATH domain (MATH proteins hereafter) are commonly found in many eukaryotes including some lower protozoa and unicellular fungi, and some iridoviruses, but as of yet not in prokaryotes (reviewed in Zapata et al., 2007). MATH proteins generally interact with other protein domains, such as peptidases, filamin and RluA domains, BTB domain, tripartite motif (TRIM), astacin domain and RING and Zinc finger domains (reviewed in Zapata et al., 2007). Regardless of their respective subfamily, all MATH proteins seem to be involved in the regulation of protein processing. MATH-domain Ubiquitin Proteases (UBPs) and Meprins possess an intrinsic protease activity and TRAF, MATH-TRIM and MATH-BTB protein families function as E3 ubiquitin ligases (reviewed in Zapata et al., 2007). The role of several MATH-BTB proteins in protein degradation will be discussed in later sections.

1.2.1 Phylogeny of MATH-BTB protein family

MATH-BTB proteins are commonly found in animals and plants but are present at disproportionate rates in different species' genomes. For instance, MATH-BTB proteins are presented by 6 members in the arabidopsis genome (Weber & Hellman, 2009), 2 in human (Juranić et al., 2012), 33 in maize (Juranić et al., 2012 and this work), 46 in *C. elegans* (Stogois et al., 2005) and 67 in rice (Gingerich et al., 2005; Gingerich et al., 2007). Gingerich et al. (2005) conducted a phylogenetic analysis of *MATH-BTB* genes from arabidopsis and rice, which showed clustering of

plant MATH-BTB genes into two separate clades: the core clade and the expanded clade. Of special interest was the large expansion of MATH-BTB proteins evident in the rice genome: rice encodes at least 67 MATH-BTB proteins, with five additional BTB proteins with MATH-related domains, and another 41 genomic loci predicted to be MATH-BTB pseudogenes (Gingerich et al., 2007; Juranić et al., 2012). While all arabidopsis proteins clustered in the core clade, rice proteins clustered into both clades, which implied that the MATH-BTB family can be divided into two groups with contrasting evolutionary histories: a small core set that is common to both rice and arabidopsis and a larger expanded set present in rice (Gingerich et al., 2005). Later analysis included a diverse spectrum of land plant species, including monocots (sorghum and wheat; *Triticum monoccum*) and dicots (*Medicago truncatula* and poplar; *Populus trichocarpa*), a gymnosperm (pine; *Pinus taeda*), a moss (*Physcomitrella patens*), and a bryophyte (*Selaginella moellendorffi*). Inclusion of more monocots in the analysis brought out the hypothesis that the expanded clade is monocot specific. Phylogenetic comparisons detected numerous gene duplication and/or loss events since the rice and arabidopsis *BTB* lineages split. Consequently, some functional specialization possibly occurred within individual *BTB* families, which might be reflected in the major expansion and diversification of MATH-BTB proteins in rice and other monocots. This likely occurred after the monocot/dicot split and was followed by a much faster evolution of *MATH-BTB* genes in the expanded clade than the conserved core clade. Rapid diversification of the substrate recognition module of many monocot MATH-BTB E3 ligases might be related to a rapid diversification of their physiological substrates (Gingerich et al., 2007). It was recently hypothesized that the expanded clade is in fact grass-specific. To test this hypothesis, *MATH-BTB* genes of non-grass monocots (banana) and eudicots (grapevine) were included in the phylogenetic analysis of plant MATH-BTB genes. None of these additional genes clustered into the expanded clade, confirming its specificity for grass MATH-BTB genes (Juranić & Dresselhaus, 2014).

1.3 Cullin-based E3 ligases

A remarkable characteristic of biological cells is their ability to rapidly adapt to various internal and external cues. This adaptation often requires quick elimination of the cell's existing proteins, most often by the ubiquitin-dependent mechanism of the 26S proteasome (reviewed in

Willems et al., 2004). Each damaged or otherwise redundant protein is recognized and designated for degradation by covalent attachment of a polyubiquitin chain on its lysine residues. Three enzymes coordinate this process. The ubiquitin-activating enzyme (E1) and the ubiquitin-conjugating enzyme (E2) mediate activation and subsequent transfer of ubiquitin through thioester bond formation. Finally, ubiquitin ligases (E3) recognize the ubiquitin-marked protein and mediate its degradation (reviewed in Pintard et al., 2004; Willems et al., 2004; Genschik et al., 2013, among others). At least four ubiquitin monomers are needed for a protein to be recognized by an E3 ligase and led to the catalytic site of the 26S proteasome, where a set of protease subunits rapidly cleaves the protein into short peptides in an ATP-dependent manner. The specificity of the process is ensured by the hundreds of E3 ligases recognising particular substrates through specific protein-protein interactions (reviewed in Willems et al., 2004). There are two major classes of E3 ligases: HECT-type E3s which possess intrinsic catalytic activity, and RING-H2-type E3s (also called Cullin-RING ligases; CRLs) which promote ubiquitinilation by positioning the activated E2 near the substrate (reviewed in Willems et al., 2004). The first major type of RING-H2 E3 ligases is the APC/C (Anaphase promoting Complex/Cyclosome) which degrades securin and cyclin B and therefore controls the onset of anaphase (reviewed in Peters, 2006). The second type consists of SCF-like complexes which are considered to be the archetype for the RING ubiquitin ligase family (reviewed in Willems et al., 2004). SCF (Skp1/Cullin/F-box) complex was first discovered through research of the budding yeast *Saccharomyces cerevisiae* cell cycle. There, the SCF complex degrades the Cdc28 cyclin-dependent kinase inhibitor Sic1, thereby allowing the cell cycle to transition into the replication (S) phase (Patton et al., 1998). The SCF complex is comprised of four main subunits: the cullin Cul1, the ring-finger protein Hrt1 (alternatively titled Roc1 or Rbx1), the linker Skp1 and a member of the F-box protein (FBP) family (reviewed in Genschik et al., 2013). Evidence suggests that the SCF complex is the prototype for two more types of RING-H2 E3 ligases, namely the ESC (ElonginC/Cul2/SOC2) complex present in animals only and the Cul3-based ligases associated with a BTB-domain protein (Pintard et al., 2003; Xu et al., 2003). In all three cases, there is a common core structure comprised of the Cullin protein as the molecular scaffold which links up the catalytic module (RING-finger domain protein and an E2 ligase) to the substrate recognition module (adaptor module) which physically interacts with target proteins. The catalytic core can associate with various substrate recognition modules which, although sharing a significant structural similarity, recognize different proteins. This creates many different E3 ligases

which can in turn process many different substrates (reviewed in Willems et al., 2004). The Cul3-based E3 ligases which function in association with BTB-domain proteins possess some unique characteristics. In contrast to Skp1 of the SCF complex and ElonginC of the ESC complex, which are both small proteins, most BTB proteins contain an additional protein-protein interaction domain (Fig 2, discussed in section 1.2.1). For instance, out of 11 BTB-proteins isolated from *C. elegans* and reported as Cul3 interactors, five contain a MATH domain which binds the target protein, while the BTB domain binds the Cullin 3 scaffold protein. Together, these two domains act as a single-polypeptide bridge between Cul3 and the substrate (Fig 2). This led to the conclusion that a single BTB protein serves as a substrate-specific adaptor which on its own performs the role of an Skp1/F-box dimer (reviewed in Pintard et al., 2004; Willems et al., 2004; Genschik et al., 2013), which was additionally confirmed by biochemical studies (Xu et al., 2003).

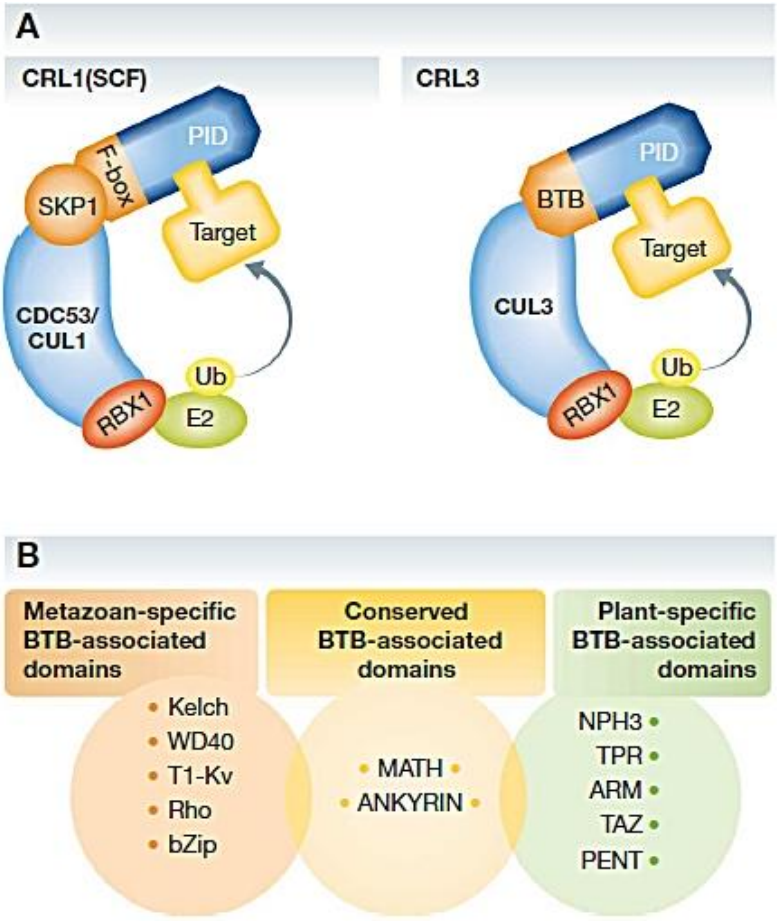


Figure 2. Structural organization of SCF/CRL1 and the CRL3 complexes. **(A)** The SCF/CRL1 and the CRL3 complexes share a similar catalytic core module composed of the scaffold proteins CUL1 and CUL3, respectively, and the RING finger protein RBX1. Single-subunit BTB domain proteins bridge CUL3-RBX1 to substrates, while this function requires an SKP1/FBP heterodimer in SCF/CRL1. Substrate recognition is governed by an independent protein–protein interaction domain (PID) found in most of the FBPs and CUL3-interacting BTB domain proteins. **(B)** A brief overview of protein domains commonly found associated with the BTB domain in CRL3 adaptors. Some domains occur in both metazoans and higher plants (BTB-MATH and BTB-ANKYRIN repeat), while others are specific to either the animal kingdom: BTB-KELCH; BTB-WD40; BTB-T1-Kv (voltage-gated potassium channel T1); BTBRho (Ras homology); BTB-bZip (basic leucine Zipper); or plant kingdom: BTB-NPH3 (non-phototropic hypocotyl 3); BTB-TPR (Tetratrico Peptide Repeat); BTB-ARM (Armadillo); BTB-TAZ (Transcriptional Adaptor Zinc finger); BTB-PENT (Pentapeptide). (Image acquired from Genschik et al., 2013.)

1. 4 Functional characterization of animal MATH-BTB proteins

To date, only two animal MATH-BTB proteins have been functionally described: MEL-26 protein from *Caenorhabditis elegans* and the human SPOP protein. Both proteins interact with Cullin 3, where they act as a component of a ubiquitin ligase complex. The BTB domain directly interacts with CUL3 and mediates dimerization of the entire E3 complex, whereas the MATH domain is responsible for substrate specificity (Gingerich et al., 2005; Stogios et al., 2005). Animal MATH-BTB proteins have various other functions, depending on protein substrates of the MATH domain.

1.4.1 MEL-26 protein from *Caenorhabditis elegans*

MEL-26 (*m*aternal *e*ffect *l*ethal-26, Dow & Mains, 1998) is a member of the MATH-BTB protein family and functions as a substrate-specific adaptor of the Cul3-based E3 ligase in *C. elegans*. The ligase is responsible for ubiquitin-dependent degradation of the microtubule-severing katanin protein MEI-1 after meiosis, a process essential for assembly of the mitotic spindle. The MATH domain of MEL-26 serves as a substrate-binding domain, while the BTB domain associates with CUL 3 (reviewed in Pintard et al., 2004). Research based on mutants, namely the mel-26(RNAi), cul-3(RNAi) and mel-26 temperature sensitive mutant, revealed failure to degrade MEI-1 and consequently a lack of mitotic spindle assembly (Pintard et al., 2003). Reconstitution of CUL-3/MEL-26 *in vitro* enabled polyubiquitination of MEI-1 (reviewed in Pintard et al., 2004). Beside the role of MEL-26 in mitotic spindle assembly, an additional role of MEL-26 was reported during cytokinesis in *C. elegans* embryos. Here, MEL-26 binds the actin binding protein POD-1 and positively regulates formation of the cleavage furrow which physically separates two daughter cells after mitosis. Not only that this function of MEL-26 is independent of its role as a substrate-specific adaptor in CUL-3 E3 ligase complexes, but MEL-26 is itself a substrate of the CUL-3 complex (Luke-Glaser et al., 2005).

1.4.2 Human SPOP protein

The human speckle-type POZ protein (SPOP) is a substrate-specific adaptor of a Cul3-based E3 ligase. SPOP belongs to the MATH-BTB subfamily and appears to be well conserved: both its sequence and function show high similarity with orthologs MEL-26 in *C. elegans* and HIB in *Drosophila melanogaster* (reviewed in Mani, 2014). Structurally, this 42 kDa protein comprises an N-terminal MATH domain, a BTB/POZ domain, a 3-box domain and a C-terminal nuclear localization sequence. Substrate proteins are recognized by characteristic amino acid residues (Y87, F102, Y123, W131 and F133) of the MATH domain. For this protein-protein interaction to occur, there needs to be a SPOP-binding consensus (SBC) motif in the substrate protein (Zhuang et al., 2009). Signature SBC motifs have been reported in several SPOP substrates, such as Macro H2A, Puc, Daxx, Gli and others (reviewed in Mani, 2014). SPOP binds the Cullin 3 scaffold protein of the E3 ligase complex via a conserved hydrophobic BTB domain, specifically an α3-β4 loop consisting of ten amino acid residues. An additional pair of α-helices stretching beyond the BTB domain, called 3-box, seems to enhance the SPOP-Cullin 3 interaction. Beside Cullin 3 binding, the BTB domain is involved in dimerization of SPOP. Thus, the functional SPOP-Cullin 3-RING box 1 ubiquitin ligase complex contains two substrate-binding sites from SPOP and two catalytic cores from Cullin 3-RING box 1 (Zhuang et al., 2009). Similar to other BTB-containing substrate-specific adaptors of Cul3-based E3 ligases, the main function of SPOP is to target various proteins designated for degradation by the 26S proteasome (reviewed in Mani, 2014). One of these targets is Daxx, a multifunctional protein involved in regulation of various cellular processes, such as transcription, cell cycle and apoptosis (Kwon et al., 2006). However, several specialized functions which do not involve targeting for proteasomal degradation have been attributed to SPOP, for example, X-inactivation, an essential physiological process of silencing one of the two X chromosomes present in cells of XX females. The MATH domain of SPOP binds a leucine zipper region of the histone variant Macro H2A and localizes it to the X-chromosome marked for inactivation. This step is necessary for X-inactivation to occur (reviewed in Mani, 2014). Similarly, SPOP binds phosphatidylinositol 4,5-bisphosphate and the polycomb group protein Bmi1 and this interaction does not appear to involve their subsequent degradation (reviewed in Mani, 2014). Extensive research is being made on the structure and function of SPOP protein, due to its connection with cancer development. SPOP is frequently mutated in prostate and endometrial

cancer cells and has been linked to several other types of cancers. All the SPOP mutations identified to date in prostate and endometrial cancers cluster in the MATH domain, presumably affecting substrate binding (reviewed in Mani, 2014).

1.5 Functional characterization of plant MATH-BTB proteins

The first detailed analyses of plant MATH-BTB proteins were published for maize ZmMAB1 protein (Juranić et al., 2012) and for arabidopsis MATH-BTB proteins, which were shown to be involved in proteasomal degradation of transcription regulators representing at least three protein families (Chen et al. 2013; 2015; Lechner et al. 2011).

1.5.1 Maize ZmMAB1 protein

Plant MATH-BTB proteins still make a fairly unexplored family of proteins. The maize (*Zea mays* L.) genome encodes 31 MATH-BTB proteins and one of them, namely the ZmMAB1 (for *Z. mays* MATH-BTB domain 1), was recently studied. *MAB1* gene is expressed solely in the male and female germ lineages and the zygote of maize. Research based on mab1(RNAi) mutant plants elucidated the role of ZmMAB1 protein during meiosis II and the first mitotic division. Specifically, downregulation of *MAB1* in maize appeared to affect nuclei positioning and identity during the meiosis-to-mitosis transition of both the male and the female gametophyte. Another observed phenotype involved chromosome segregation defects and short spindles during meiosis, which proposed a role of ZmMAB1 in spindle apparatus formation and function. ZmMAB1 interacts with itself and with one of three known Cullin 3 proteins of maize (specifically, CUL3a), suggesting a dimeric E3 ligase interaction model in which ZmMAB1 binds both the CUL3a protein and another ZmMAB1 protein via its BTB domain. Furthermore, ZmMAB1 shows weak interaction with p60 subunit of the microtubule-severing protein Katanin from arabidopsis (AtKTN1). Together, these results suggest a functional similarity between the maize ZmMAB1 protein and *C. elegans* MEI-26 protein (Juranić et al., 2012).

1.5.2 Arabidopsis AtBPM proteins

In *Arabidopsis thaliana*, there are six MATH-BTB proteins, AtBPM1-6 (for *A. thaliana* BTB/POZ-MATH), localized in either the nucleus (e.g. BPM2), the cytoplasm (e.g. BPM4) or both compartments (Weber & Hellman, 2009). Via their BTB domain, they interact with CUL3a and CUL3b proteins, presumably as substrate-specific adaptors of a Cul3-based ubiquitin E3 ligase (Weber et al., 2005; Gingerich et al., 2005, among others). AtBPMs have been connected with the regulation of abscisic acid (ABA) response, a mechanism of crucial importance for both biotic and abiotic stress responses. The class I homeoboxleucine zipper (HD-ZIP) transcription factor ATHB6 is a negative regulator of ABA responses such as sensitivity toward ABA during seed germination and stomatal closure. It has been demonstrated that the MATH domain of AtBPM proteins directly interacts with the leucine-zipper domain of ATHB6 protein, which is then targeted for degradation (Lechner et al., 2011). AtBPM proteins have also been linked to ethylene response. Specifically, AtBPMs interact with members of the ethylene response factor/Apetala2 (ERF/AP2) transcription factor family via their MATH domain. A working model has been proposed in which AtBPMs affect transcriptional activities of ERF⁄AP2 by interfering with their DNA-binding ability but ultimately target them for degradation (Weber & Hellman, 2009). A recent study reported a role of AtBPMs in regulation of flowering. Transcription factor MYB56 was indicated as a negative regulator of *FLOWERING LOCUS T (FT)*, a key activator of flowering in arabidopsis. Interaction of AtBPMs and MYB56 results in instability of MYB56 and therefore promotes flowering (Chen et al., 2014).

1.5.3 Wheat TaMAB2 protein

Analysis of genes specifically induced in the zygote and proembryos of wheat (*Triticum aestivum* L.) revealed the existence of two genes encoding MATH-BTB proteins, namely *TaMAB1* and *TaMAB2*. While *TaMAB1* gene is expressed in egg cells only, *TaMAB2* expression was detected in both the zygote and the two-celled embryo stage, followed by a downregulation in later stages (Leljak-Levanić et al., 2013). TaMAB2 protein appears to accumulate in the nucleus but more abundantly around the nucleus unilaterally in both BMS (maize Black Mexican sweet) and BY-2 (*Nicotiana tabacum* cv. Bright Yellow 2) cells (Leljak-Levanić et al., 2013). However, its

exact position within wheat cells *in vivo* is yet to be unequivocally determined. The amino acid sequence of TaMAB1 shows significant similarity with ZmMAB1 protein of maize, AtBPM3 protein of *Arabidopsis thaliana* and several other MATH-BTB proteins from sorghum, rice, *Brachypodium* and barley. It has been proposed that TaMAB2 functions as part of a Cul-3 based ligase complex and, by binding target proteins via its MATH domain, helps regulate processes such as zygotic division, progression of early embryogenesis or the establishment of polarity or cell fate in the developing embryo (Leljak-Levanić et al., 2013).

1.6 Thesis objective

In a recent study, wheat zygotes and early embryos were isolated and *de novo* transcribed genes were identified (Leljak-Levanić et al., 2013). One of these genes was *TaMAB2*, encoding a MATH-BTB protein presumably acting as a substrate-specific adaptor of a Cul3-based E3 ligase. One aim of this thesis was to determine potential interactive partners of TaMAB2 protein, between the candidates reproducibly identified by mass spectroscopy (Leljak-Levanić, unpublished). This would shed light on possible functions of TaMAB2 during the zygote and early embryo development, as well as molecular mechanisms involved in the process. Since viable *TaMAB2* knockout mutants could not be obtained (Leljak-Levanić, unpublished), an aim of this work was to study the effect of *TaMAB2* overexpression on zygotic embryogenesis, somatic embryogenesis and finally on growth and development of adult arabidopsis plants. Lastly, a goal was to study the phylogeny of TaMAB2 within the wheat MATH-BTB family of proteins as well as MATH-BTB proteins of several selected land plant species. Together, these results would broaden the currently limited knowledge on MATH-BTB proteins of wheat.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Plant material and growth conditions

2.1.1.1 *Arabidopsis thaliana* (L.) Heynh.

Seeds of *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) and transgenic plants 35S:TaMAB2:GFP were grown under standard growth chamber conditions at 21 °C with 16 h of supplementary light during the day period and a relative air humidity of 40 to 60%.

2.1.1.2 Wheat, *Triticum aestivum*

Seeds of wheat ecotype Magdalena were grown on cotton wool under standard growth chamber conditions at 21 °C with 16 h of supplementary light during the day period and a relative air humidity of 40 to 60%.

2.1.2 Yeast, *Saccharomyces cerevisiae* strain

For yeast-two-hybrid screens the yeast host Hfc7 [MATα ura3-52 his3-200 ade2-101 lys2-801 trp1-901, leu2-3112 gal4-542 gal 80-538 LYS2::$GAL1_{UAS}$-G-$AL1_{TATA}$-HIS3 URA3::$GAL4_{17mers(x3)}$-$CyC1_{TATA}$-lacZ)] was used. The Hfc7 strain contains reporter genes (HIS3 and lacZ) integrated into the genome (Feilotter et al., 1994).

2.1.3 Bacteria strains

Escherichia coli strain HST04 (Stellar^{TM} Competent Cells, Clonetech, # 636763) was used in this research.

2.1.4 Primer sequences

All primer sequences used in this work were designed by using Clonetech InFusion Primer Design Tool (http://www.clontech.com/US/Products/Cloning_and_Competent_Cells/Cloning_Resources/Online_In-Fusion_Tools) and ordered from Macrogen. Primers were prepared in ultrapure water (Mili-Q®) as a 100 μM stock solution and diluted to 10 μM before use. The stocks and working solutions were stored at –20 °C. For oligonucleotide sequences (Table 1) and in-text details of primer design protocol, see section 2.2.5.1.

2.2 Methods

2.2.1 Bioinformatics and phylogenetic analysis

The databases of NCBI GenBank and Ensembl Plants were searched using BLASTp and the TaMAB2 amino acid sequence as query. The non-redundant full-length sequences obtained from databases were aligned with Clustal X. Sequences were aligned and compared with known MATH-BTB proteins from maize (Juranic & Dresselhaus, 2014), rice (Gingerich et al., 2007; Juranic et al., 2013), and arabidopsis (Figueroa et al., 2005; Gingerich et al., 2005; Weber et al., 2005). The phylogenetic tree of MATH-BTB sequences was generated by neighbor-joining method of Saitou and Nei (1987) implemented in SeaView v.4 and the result was visualised in FigTree v.1.4.2 (Andrew Rambaut, http://tree.bio.ed.ac.uk/). Following generation of phylogernetic tree, selected non-redundant full-length sequences of wheat and rice MATH-BTB proteins were aligned with MUSCLE v3.8.31 (http://www.ebi.ac.uk/Tools/msa/muscle/, Edgar, 2004). The resulting alignments were displayed using Boxshade 3.21 (http://www.ch.embnet.org/software/BOX_form.html).

2.2.2 Generation of total RNA and cDNA synthesis

To isolate total RNA from wild type *Triticum aestivum* seedlings, RNeasy Plant Mini Kit (Qiagen) was used, following instructions provided with the kit. To obtain complementary DNA (cDNA) from extracted RNA samples, a polymerase chain reaction using reverse transcriptase

(RT-PCR) was performed. To 200 – 350 ng of RNA template, 0.5 μg of Oligo $(dT)_{18}$ primer and DEPC water in a total volume of 11 μL was added. The reaction was incubated at 70 °C for 5 minutes. This was followed by addition of 2 μL 10mM deoxyribonucleotide mix, 4 μL of 5X Reaction Buffer (MBI Fermentas) and 40 U of RiboLock Ribonuclease Inhibitor (MBI Fermentas). The sample was incubated at 37 °C for 5 minutes. Finally, 200 U of RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI Fermentas) was added. The mixture was incubated at 45 °C for 1 hour and then at 70 °C for 10 minutes. Obtained cDNA was stored at –20 °C until further use. Afterwards, cDNA samples were applied for a standard PCR (35 cycles) using gene-specific primers pairs.

2.2.3 Generation of mRNA and cDNA synthesis from minute amount of tissue

To isolate total messenger RNA (mRNA) from wild type arabidopsis somatic embryos (3 per sample) Dynabeads® mRNA DIRECT™ Kit (Thermo Fisher Scientific) was used, following manufacturer's instructions. The end product of this procedure was total extracted mRNA bound to oligo $(dT)_{25}$ residues covalently coupled to magnetic beads. To obtain cDNA from this sample, an RT-PCR was performed by adding 0.5 μg Oligo $(dT)_{18}$ primer and 1 μL 10 mM deoxyribonucleotide mix and incubation for 5 minutes at 65 °C. This was followed by addition of 200 U of RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI Fermentas), 4 μL of provided 5X Reaction Buffer (MBI Fermentas) and 40 U of RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific). The sample was incubated at 42 °C for 50 minutes and then at 70 °C for 15 minutes. cDNA was stored at –20 °C until further use for standard PCR reactions.

2.2.4 Standard PCR reaction

All reactions were performed in a gradient thermocycler (Eppendorf Mastercycler) in a 20 μL reaction volume containing 0,4 U GoTaq® DNA Polymerase (Promega), 4 μL 5X Green GoTaq® Reaction Buffer (Promega), 8 pmol of each primer, 1 μL 10 mM deoxyribonucleotide mix and 100 - 200 ng of cDNA as template. Initial denaturation step was performed at 95 °C for 2 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, extension

at 72 °C for 1 min/kb and a final extension step at 72 °C for 5 minutes. The reactions were stored at 4 °C and were later used as templates for amplification of the *Katanin* and *Tudor-SN2* gene inserts.

2.2.5 Generation of plasmid constructs

2.2.5.1 Generation of PCR inserts

Gene-specific primers (Table 1) were designed by using Clonetech InFusion Primer Design Tool. The 5' end of each primer contained 15 bases that were homologous to 15 bases at the end of DNA with which the PCR product would be joined. Specifically, a pair of primers was designed in a way that it generated PCR products containing 5' ends homologous to ends of a BamHI-linearized vector. Plasmids used in this work (pGAG424 and pGBT9) share the same sequence surrounding the BamHI restriction site. Therefore, a pair of primers could be used interchangeably to clone into both plasmids. The 3' end of the primer contained a sequence specific to the gene that was being amplified. This part of the primer was 18-25 nucleotides long, had GC content of 40-60% and had a Ta between 58 and 65°C. Each gene was amplified using the 1X InFusion CloneAmp™ HiFi PCR Premix (Clonetech, #639298) along with 7.5 pmol of each gene-specific primer and 25 ng of template DNA (Table 2). The PCR reaction was carried out in a gradient thermocycler (Eppendorf Mastercycler) with the following cycling conditions: denaturation at 98 °C for 10 seconds, annealing at 55 °C for 15 seconds and elongation at 72 °C for 10 or 15 seconds. Total number of cycles was 35 and the reaction volume was 20 µL. Following amplification, gene inserts were purified using the NucleoSpin® Gel and PCR Clean-up (Clonetech) per protocol. Concentrations of inserts were measured using NanoVue™ spectrophotometer (GE Healthcare, Life Sciences) and purified inserts were stored at – 20 °C.

Table 1. Oligonucleotide sequences used for cloning. All primers were designed with Clonetech InFusion Primer Design Tool (for web page, see section 2.1.4). Gene name, gene or transcript ID from the EnsemblPlants, TAIR or GenBank database, primer name and nucleotide sequence, and size of product in base pairs (bp) with cDNA as template are indicated in the table. Each PCR product contained 15 bases of ss DNA at both 5' ends (not included in the final product size). Primer ends homologous to gene of interest are underlined. 15 bases of flanking ends homologous to ends of BamHI linearized plasmid are left unmarked. Start and stop codons are shown in bold.

Gene name	Gene or transcript ID	primer name	nucleotide sequence 5' – 3'	Product size (bp)
TaMAB2	Traes_2BL_7DE6E520A (Ensembl Plants)	FWTaMAB2	GAATTCCCGGGGATC**ATG**GCCAGCAACTCCACC	1074
		REVTaMAB2	GCAGGTCGACGGATC**CTA**GTCAGTAAGAGAAATTTTT	
TUDOR-SN2	AT5G61780.1 (TAIR)	FWTUD	GAATTCCCGGGGATC**ATG**GCGACTGGGGCAGCA	2958
		REVTUD	GCAGGTCGACGGATC**TTA**CCCGCGACCCGGTTT	
Actin11	AT3G12110.1 (TAIR)	FWACT11	GAATTCCCGGGGATC**ATG**GCAGATGGTGAAGACA	1134
		REVACT11	GCAGGTCGACGGATC**TTA**GAAGCACTTCCTGTGG	
Katanin	HG670306.1 (GenBank)	FWKAT	GAATTCCCGGGGATC**ATG**GCGAACCCCCTAGCC	1566
		REVKAT	GCAGGTCGACGGATC**TTA**AGCAGATCCAAACTCGG	

Table 2. Genes used for generation of plasmid constructs. Listed in the table are each gene's name, open reading frame (ORF) size in base pairs (bp) with STOP codon included, source of template DNA and forward and reverse oligonucleotide sequences used as primers for amplification.

Gene name	ORF size (bp)	template DNA	Primers
TaMAB2	1074	pLNU-TaMAB2-GFP	**FWTaMAB2**
			REVTaMAB2
TUDOR-SN2	2958	cDNA of wild type *Arabidopsis thaliana* plants	**FWTUD**
			REVTUD
Actin11	1134	pGEX5x2-actin11	**FWACT11**
			REVACT11
Katanin	1566	cDNA of wild type *Triticum aestivum* plants	**FWKAT**
			REVKAT

2.2.5.2 Plasmid preparation and cloning procedure

Plasmids pGAD424 and pGBT9 were digested with BamHI restriction enzyme for 4 hours at 37 °C. The enzyme was then inactivated by incubation at 65 °C for 20 min. Linearized plasmids were run through 0,8% agarose gels on low voltage (50 V) and stained with ethidium bromide. Bands containing the linearized vector were manually excised from gels under UV light. Plasmid DNA was purified from gels using QIAquick Gel Extraction Kit (Qiagen) as advised by the manufacturer. Concentrations of linearized vectors were measured using NanoVue™ spectrophotometer (GE Healthcare, Life Sciences). The previously generated PCR inserts (section 2.2.5.1) contained 15 additional bases at their ends. These sequences were homologous to the ends of BamHI-digested plasmids (both pGAD424 and pGBT9) which enabled homologous end-joining and subsequent cloning of each gene insert into the linearized vector by means of homologous recombination. To clone PCR-generated gene inserts into BamHI restriction sites of plasmids, InFusion® HD Cloning Kit (Clonetech) was used as follows. Each InFusion reaction mixture was prepared by mixing 50 ng of purified PCR fragments, 100 ng of linearized vector, 2 μL of provided Enzyme Premix and deionized water to make up a total volume of 10 μL. The reaction was incubated for 15 minutes at 50 °C and placed on ice. Thus, seven plasmid constructs were prepared: pGAD424-TaMAB2, pGBT9-TaMAB2, pGAD424-Tudor, pGBT9-Tudor, pGAD424-Actin11, pGBT9-Actin11 and pGBT9-Katanin. In addition, two control reactions were prepared in which the PCR fragment was omitted and only the linearized plasmid (pGAD424 or pGBT9) was added to the mixture. Intensive growth of bacterial cells transformed with these reaction mixtures would imply spontaneous recircularization of the vector and therefore less effective cloning reaction.

2.2.6 Transformation of competent cells

Chemically competent Stellar cells were thawed in an ice bath just before use and 46 μL (per transformation) was transferred to a round-bottom Falcon tube. 2.5 μL of each InFusion reaction mixture containing a plasmid construct (approximately 10 ng of DNA) was added to the bacteria and cells were left to incubate on ice for 30 minutes. Cells were heat-shocked at 42 °C for exactly 45 seconds and then placed on ice for 1-2 minutes. SOC medium (Clonetech) was warmed

to 37 °C and 500 μL was added to each tube. Cells were incubated for 1 hour at 37 °C with shaking (225 rpm). An appropriate amount was plated on LB medium containing bacto-tryptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L in dH_2O, pH 7.0 and agar 15 g/L, supplemented with ampicillin (100 μg/mL). Cells were grown upside-down at 37 °C overnight.

2.2.7 Isolation of plasmid DNA, restriction analysis and gel-electrophoresis

2.2.7.1 Isolation of plasmid DNA

For isolation of plasmid DNA from E. coli strains, a single colony was picked and cultured at 37 °C overnight in 3 mL of LB medium supplemented with 3 μL of ampicillin (stock concentration 100 mg/mL). Cells were harvested by one-step centrifugation (5 min, 14 000 rpm) and plasmid DNA was isolated using the Wizard® Plus SV Minipreps DNA Purification System (Promega) according to manufacturer's instructions. Concentrations of plasmid DNA were subsequently measured using the NanoVue™ spectrophotometer (GE Healthcare, Life Sciences). Purified plasmid DNA was stored at -20 °C.

2.2.7.2 Restriction analysis and gel-electrophoresis

Purified plasmid DNA was cut with appropriate restriction enzymes, in reaction conditions advised by the manufacturer. Following the restriction of plasmid DNA, enzymes were inactivated by heating. DNA fragments were separated on 1% agarose gels (100 V) and stained with ethidium bromide.

2.2.8 Yeast-two-hybrid screen

2.2.8.1 Co-transformation of yeast cells

A GAL4-based yeast two hybrid (Y2H) assay was conducted using a Hfc7 yeast reporter strain. Beside the seven plasmid constructs described in section 2.2.5., two additional plasmid constructs

carrying the ORF of the histone variant H3.3 were used: pSG10-H3.3 and pSG11-H3.3 (by courtesy of Zdravko Lorković, Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Science, Vienna). In general, the ORFs of all plasmid constructs coded for proteins of interest N-terminally fused with either the activation domain (AD) or DNA-binding domain (BD) of Gal4 transcription factor. Nucleotide sequence fidelity of all fusion products was verified by sequencing of plasmid DNA, by service (Macrogen Europe, The Netherlands). In the Y2H assay, plasmid constructs containing the TaMAB2 CDS were combined with each of its potential interactors in both (*Actin11, Tudor, H3.3*) or just one (*Katanin*) orientation, depending on the availability of constructs. Hfc7 yeast cells were co-transformed using a standard lithium-acetate (LiAc) technique (Agatep et al., 1998) as follows. Three colonies were diluted in 5 mL of YPD medium and grown overnight at 28 °C with shaking (225 rpm). Overnight culture was diluted 20X in fresh YPD medium and incubated for 4 hours at 28 °C with shaking (225 rpm). 12 mL of yeast suspension was transferred to a sterile Falcon tube. This volume was sufficient for up to ten individual transformations. The cells were harvested by two-step centrifugation (2,000 g for 3 minutes at RT). The pellet was resuspended in the transformation mixture containing 240 μL 50 % polyethylene glycol (PEG), 36 μL 1M lithium acetate (LiAc) and 5 μL of single stranded carrier DNA (salmon sperm DNA, 2 mg/mL) per single transformation. This step was done on ice to avoid renaturation of carrier DNA, which was denatured prior to this step by heating the tube to 95 °C for 5 minutes. In sterile microcentrifuge tubes, combinations of two different plasmid constructs were prepared. In addition to experimental plasmid combinations, three negative and one positive control combination was used. Proteins DMS3 (DEFECTIVE IN MERISTEM SILENCING3) and RDM1 (RNA-DIRECTED DNA METHYLATION1) are known interactors (Sasaki et al. 2014) (Table 3). Each tube contained 300 ng of each plasmid, in volumes calculated according to plasmid concentration. Transformation mixture was added onto each plasmid combination in a volume of 281 μL and the cells were incubated for 30 minutes at 28 °C with shaking (225 rpm). The cells were then heat-shocked for 15 minutes in a 42 °C water bath. Immediately following the heat-shock, cells were kept on ice for 1-2 minutes and then harvested by centrifugation for 1 minute at 3000 g. The pellet was resuspended in 1 mL of YPD medium and the cells were incubated for 1 hour at 28 °C with shaking (225 rpm). After another centrifugation step (1 min, 3000 g), supernatant was decanted and pellet was resuspended in the remains of YPD medium. The yeast suspensions were then plated on selective YC-ZW^{-} medium lacking leucine

and tryptophan. The plates were sealed with plastic paraffin film and the colonies were grown right-side-up for 3 days at 28 °C.

Table 3. Plasmid combinations used for co-transformation of yeast cells in the Y2H assay. In the left column are prey constructs encoding a protein of interest (POI) fused with the activating domain (AD) of Gal4 transcription factor. In the right column are bait constructs encoding a POI fused with the binding domain (BD) of Gal4. Combination 9 is positive control and combinations 10-12 are negative controls.

	Plasmid combination	
	PREY CONSTRUCT	**BAIT CONSTRUCT**
1	pGAD424-TaMAB2	pGBT9-TaMAB2
2	pGAD424-TaMAB2	pGBT9-Tudor
3	pGAD424-TaMAB2	pGBT9-Katanin
4	pGAD424-TaMAB2	pSG11-H3.3
5	pGAD424-TaMAB2	pGBT9-Actin11
6	pGAD424-Tudor	pGBT9-TaMAB2
7	pGAD424-Actin11	pGBT9-TaMAB2
8	pSG10-H3.3	pGBT9-TaMAB2
9	pGAD424-DMS3	pGBT9-RDM1
10	pGAD424	pGBT9-TaMAB2
11	pGAD424-TaMAB2	pGBT9
12	pGAD424	pGBT9

2.2.8.2 Histidine prototrophy assay

Individual colonies were diluted in sterile distilled water in a volume of 250 μL (for fresh colonies) or 500 μL (for older colonies). To assay for His prototrophy, 20 μL of each dilution was spotted in a grid-like pattern on selective YC-ZWH$^-$ 3-AT medium which lacked leucine, tryptophan and histidine. The medium contained 13 mM 3-amino-1,2,4-triazole (3-AT) for elimination of non-specific protein-protein interactions. Additional 20 μL of each dilution was spotted on a selective YC-ZW$^-$ medium (“master” plate). Two master plates were spotted in each

experiment. The plates were sealed with plastic paraffin film and the colonies were grown right-side-up for 3 days at 28 °C.

2.2.8.3 β-galactosidase assay using X-gal as substrate

A "master" plate was used for β-galactosidase assay. In 10 mL of Z buffer (16.1 g/L $Na_2HPO4 \cdot 7H_2O$, 5.5 g/L $NaH_2PO_4 \cdot H_2O$, 0.75 g/L KCl, 0.246 g/L $MgSO_4 \cdot 7H_2O$), 0.7 g of agarose was melted by boiling. Once it cooled, 167 μL of X-gal (40 mg/mL) and 27 μL of β-mercaptoethanol was added to the tube. Yeast colonies were replica plated on a filter paper, which was then dipped a few times into liquid nitrogen. The filter paper was placed in a clean Petri dish with replicated colonies facing up. Agarose was poured over it in a smooth layer and the Petri dish was sealed with plastic paraffin film. The assay was incubated for a maximum of 5 hours at 28 °C or room temperature.

2.2.8.4 Liquid culture assay using ONPG as substrate

Liquid culture assay using ortho-Nitrophenyl-β-galactoside (ONPG) as substrate was conducted to further examine only the protein-protein interactions implied by previous assays. For this assay, yeast co-transformants were picked directly from master plates, diluted in 5 mL of liquid selection medium (YC-ZW^-) and grown overnight at 28 °C with shaking (225 rpm). Overnight cultures were resuspended and 2 mL of culture was transferred to 8 mL of YPD medium. The cultures were incubated at 28 °C with shaking (230 rpm) until the cells entered mid-log phase of growth. This was confirmed spectrophotometrically (OD_{600} of 1 mL suspension was 0.5-0.8). 1 or 1.5 mL (V_1) of each yeast suspension was transferred to a clean microcentrifuge tube (2 replicas) and centrifuged at 14,000 rpm for 30 seconds at room temperature. Supernatants were removed and pellets were resuspended in 1.5 mL of Z buffer. After another centrifugation step, cell pellets were resuspended in 300 μL (V_2) of Z buffer. 100 μL of each cell suspension was transferred to a new microcentrifuge or plastic Falcon tube and the cells were alternately frozen in liquid nitrogen (30-60 sec) and thawed in a 37 °C water bath (30-60 sec) a few times. For each sample, a blank tube containing 100 μL of Z buffer was prepared. 700 μL of solution A (0.27 mL of β-mercaptoethanol in 100 mL Z buffer) was added to the reaction and blank tubes. Time was

recorded from this point on. 160 μL of solution B (ONPG at 4 mg/mL in Z buffer) was immediately added to each tube. As soon as yellow colour developed, 400 μL of 1 M Na_2CO_3 was added to the tube. The time of this step was recorded for each sample. Reaction tubes were then centrifuged for 10 minutes at 14,000 rpm and supernatants were transferred to clean cuvettes. The spectrophotometer was calibrated against the blank at A_{420} and the OD_{420} of samples was measured. For each sample, β−galactosidase units were calculated according to the following formula:

$$\textbf{β–galactosidase units} = 1{,}000 \times OD_{420} / (t \times V \times OD_{600}),$$

where **t = elapsed time (in min) of incubation**

V = 0.1 mL × concentration factor (calculated as V_1/V_2)

OD_{600} = A_{600} of 1 mL of culture

One β−galactosidase unit is defined as the amount of β−galactosidase which hydrolyses 1 μmol of ONPG to *o*-nitrophenol and D-galactose per minute per cell (Miller, 1972**).**

2.2.9 Arabidopsis zygotic embryo isolation

Arabidopsis seeds of wild type and transgenic line (*35S:TaMAB2-GFP*) were sterilized as follows. The seeds were rinsed for 1 minute in 75% ethanol, centrifuged for a few seconds, then washed for 10 minutes in an aqueous solution containing 1% NaOCl (Izosan®-G, Pliva) and 0.1% Mucasol® (Merz Consumer Care GmbH). Seeds were centrifuged for a few seconds, washed by adding 1 mL of sterile H2O, vortexed and centrifuged again. This step was repeated four more times. Sterilized seeds were stratified overnight at 4 °C in the dark. Seeds were plated on MS medium containing macro- and micronutrients (Murashige & Skoog, 1962), 20 g/L sucrose, 100 mg/L myo-Inositol, 0.1 mg/L thiamine, 0.5 mg/L niacin, 0.5 mg/L pyridoxine, 2 mg/L glycine (pH 5.8) and solidified with 0.8% agar type A (Sigma) The medium for transgenic plants was supplemented with 10 μg/L glufosinate-ammonium (BASTA). Seeds were grown under standard greenhouse conditions at 26 °C with 16 h of supplementary light during the day period and a relative air humidity of 40 to 60%. After two weeks, seedlings were planted in a mix of soil (75%)

and sand (25%) and were grown in the same conditions. The plants were then transferred to a chamber with long-day conditions (16 h light, 4,500 lx, 22 °C) to induce flowering.

For isolation of immature zygotic embryos (IZE) at the late cotyledonary stage, around 10 mm-long siliques positioned lower on the plant stem were cut from wild type and transgenic plants. Siliques were sterilized by the same procedure as described earlier in this section. Using sterile scalpels and pincers, each silique was carefully cut open and the ovules were scraped into a drop of autoclaved distilled water. A glass coverslip was heat-sterilized on an open flame and, once cooled, placed on the drop of water containing the ovules. Applying light pressure to the coverslip, immature zygotic embryos with undamaged cotyledons were pushed out of their ovules. The embryos were used in further procedures (described in section 2.2.10) immediately after isolation.

An alternative enzymatic method was used to isolate less advanced stages of zygotic arabidopsis embryos. The siliques used in this procedure were shorter, thinner and placed higher up on the plant stem. Isolation was performed on special microscopic slides prepared as follows. Using hot wax, clean glass coverslips were fixed onto metal microscopic slides with round holes in the centre. Consequently, a glass-bottomed pool was created. Inside this pool, immature siliques were immersed in 50 μL of 13% mannitol solution and the ovules were isolated by a similar procedure as described above. Another 75 μL of mannitol was added, along with 75 μL of an aqueous enzymatic solution containing 0.5% Pectinase, 0.3% Hemicellulase, 0.5% Cellulase and 0.5% Driselase. The slides were placed in a humid container and incubated at 37 °C for a minimum of 1 hour before microscopy was performed (procedure described in section 2.2.11).

2.2.10 Induction of somatic embryogenesis

Following isolation, late cotyledonary-stage zygotic embryos were plated onto E5 induction medium containing macro- and micronutrients (Gaj, 2011.), 20 g/L sucrose, 100 mg/L myo-inositol, 10 mg/L thiamine, 1 mg/L niacin, 1 mg/L nicotinic acid and 3.5 g/L phytagel (Sigma) (pH 5.8). The medium was supplemented with 5 μM 2,4-dichlorophenoxyacetic acid (2,4-D) for induction of somatic embryogenesis (SE). Plates were kept at 22 °C with 16 h photoperiod. After five days, embryos were transferred to E5 medium without 2,4-D to enable embryo maturation and were kept in the same conditions. The growth and development of embryos was monitored until wild type explants developed somatic embryos. The frequency of SE was

calculated for both wild type and *TaMAB2* overexpression lines as a ratio of explants carrying fully developed somatic embryos and total number of explants.

2.2.11 Microscopy

Images presented in this work were obtained with Axiovert 200M (Zeiss) inverted fluorescence microscope equipped with AxioCam MRc microscope camera and Zeiss binocular magnifier (STEMI 2000-C). Image acquisition on both microscopes was controlled by AxioVision imaging software version 4.5 (Zeiss). Propidium iodide was visualised with Filter 14 (Zeiss), ekscitation BP 510-560 nm, emision LP590.

3 RESULTS

3.1 Identification of TaMAB2 homologs

3.1.1 TaMAB2 paralogs

Three MATH-BTB proteins were previously reported in wheat plants, namely TaMAB1, TaMAB2 and TaMAB3. The expression of wheat *TaMAB2* gene was detected in the zygote and two-celled proembryos of wheat plants (Leljak-Levanić et al., 2013). In order to identify MATH-BTB proteins related to TaMAB2, a BLASTp search of the wheat genome was performed using TaMAB2 amino acid sequence as query. This revealed 49 proteins in total, TaMAB1-49, all of which were named according to the existing nomenclature of the three known TaMAB proteins (for *T. aestivum* MATH-BTB). Pseudogenes were not included in the phylogenetic analysis. All 49 proteins (listed in Table 4) contain one MATH and one BTB domain. With the exception of TaMAB9, TaMAB21 and TaMAB36, all proteins contain an SPOP_C_like_plant domain. None of the 49 genes encoding TaMAB proteins appear to have splicing variants.

Table 4. *MATH-BTB* genes identified in the wheat genome. Forty-nine new genes have been given the same name as *TaMAB1-3* (*T. aestivum* MATH-BTB). Listed in the table are each gene's sequence identifier from the EnsemblPlants database, number of splicing variants, length of protein product as the number of amino acids (aa) and number of coding exons. Highlighted in red are the *TaMAB* genes of the core group.

gene product name	sequence identifier (Locus ID)	number of splicing variants per gene	translation length of transcript (aa)	number of exons
TaMAB1	Traes_7DL_429AF48FB	1	362	1
TaMAB2	Traes_2BL_7DE6E520A	1	357	1
TaMAB3	Traes_2BL_7373D6D88	1	492	2
TaMAB4	Traes_4BL_1C26E7509	1	355	1
TaMAB5	Traes_4BL_8C691D4D8	1	358	1
TaMAB6	Traes_5BL_8D0586A28	1	489	2
TaMAB7	Traes_5BL_3B3D8AFB0	1	358	1
TaMAB8	Traes_5BL_81EDE2391	1	342	1
TaMAB9	TRAES3BF092000180CFD_g	1	293	1
TaMAB10	TRAES3BF026700040CFD_g	1	344	1
TaMAB11	TRAES3BF051700020CFD_g	1	345	1

Table continued on page 32.

TaMAB12	TRAES3BF032500020CFD_g	1	362	1
TaMAB13	TRAES3BF102200050CFD_g	1	351	2
TaMAB14	TRAES3BF109300050CFD_g	1	387	2
TaMAB15	TRAES3BF061400020CFD_g	1	354	1
TaMAB16	TRAES3BF061400010CFD_g	1	354	1
TaMAB17	Traes_7BL_ED4758643	1	345	1
TaMAB18	Traes_7BL_AB14BD6B8	1	387	2
TaMAB19	Traes_7BL_F1DBCFA41	1	373	4
TaMAB20	Traes_7BL_4815ACB6E	1	325	2
TaMAB21	Traes_2BS_CCD960BA7	1	300	5
TaMAB22	Traes_5BL_D693571C1	1	368	4
TaMAB23	Traes_2BL_3E12FF58C	1	239	4
TaMAB24	Traes_2BS_507E9764D	1	268	5
TaMAB25	Traes_3B_B777F2656	1	357	1
TaMAB26	TRAES3BF062000010CFD_g	1	328	2
TaMAB27	TRAES3BF186700010CFD_g	1	325	2
TaMAB28	TRAES3BF061700080CFD_g	1	380	2
TaMAB29	TRAES3BF061700090CFD_g	1	256	1
TaMAB30	TRAES3BF066100020CFD_g	1	365	1
TaMAB31	Traes_5BL_DF35BE5AD	1	367	1
TaMAB32	Traes_5BL_A2F8DBD92	1	369	1
TaMAB33	Traes_5BL_9D7ABCE22	1	335	1
TaMAB34	Traes_5BL_87E53DDDB	1	367	1
TaMAB35	Traes_5BL_FF509A8DD	1	264	1
TaMAB36	TRAES3BF043700190CFD_g	1	309	1
TaMAB37	TRAES3BF043700210CFD_g	1	360	1
TaMAB38	TRAES3BF043700200CFD_g	1	359	1
TaMAB39	TRAES3BF043700230CFD_g	1	341	1
TaMAB40	TRAES3BF044600010CFD_g	1	306	1
TaMAB41	TRAES3BF043700250CFD_g	1	351	1
TaMAB42	TRAES3BF043700240CFD_g	1	362	1
TaMAB43	TRAES3BF044600020CFD_g	1	315	2
TaMAB44	TRAES3BF175800010CFD_g	1	408	1
TaMAB45	Traes_5BL_444B794F2	1	389	1
TaMAB46	TRAES3BF109300040CFD_g	1	382	1
TaMAB47	Traes_2BL_BB28B12CA	1	371	1
TaMAB48	Traes_5BL_CB4EED072	1	353	1
TaMAB49	TRAES3BF075100020CFD_g	1	356	1

3.1.2 TaMAB2 plant homologs

To study the phylogenetic relationship of wheat MATH-BTB proteins with MATH-BTB proteins of other plant species, specifically the two monocots, maize and rice, and a dicot *Arabidopsis thaliana*, another BLASTp search was performed against their genomes, using TaMAB2 amino acid sequence as query. This revealed 31 maize MATH-BTB proteins reported in Juranić et al., 2012, along with two additional maize proteins, named ZmMAB15like and ZmMAB17like (Supplemental Table 1), 67 rice MATH-BTB proteins reported in Gingerich et al., 2007 and Juranić et al., 2012 (Supplemental Table 2) and 6 arabidopsis MATH-BTB proteins reported in Figueroa et al., 2005; Gingerich et al., 2005; Weber et al., 2005 and others (Supplemental Table 3). Therefore, the input for phylogenetic analysis included 155 proteins in total.

3.2 Phylogenetic tree of selected plant MATH-BTB proteins

Nucleotide sequences (open reading frames) encoding all previously known and newly recognized MATH-BTB proteins of wheat, maize, rice and arabidopsis (Table 4, Supplemental Tables 1-3) were used to obtain a phylogenetic tree (Fig 3). Nucleotide sequences of 155 genes (49 of wheat, 33 of maize, 67 of rice and 6 of arabidopsis) were analysed using the neighbour-joining method for inferring molecular phylogeny. Two phylogenetic trees were constructed: an unrooted tree, in which the relations between the nodes are shown without assumed ancestry (Fig 3) and a rooted one (Fig 4). It was previously reported that plant *MATH-BTB* genes cluster into a core clade and an expanded clade (Gingerich et al., 2007; Juranić & Dresselhaus, 2014) and similar results were obtained in this analysis. *MATH-BTB* genes of the expanded clade additionally clustered into five subclades, named E1-E5 to follow the nomenclature of Juranić & Dresselhaus (2014). Most monocots clustered into the monocot-specific expanded clade, with the exception of 4 rice, 4 wheat and 8 maize genes which clustered into core clade. All arabidopsis genes clustered into core clade. Of the expanded group subclades, only E2 did not contain any wheat genes. *TaMAB2* clustered into E3 subclade, along with three other wheat genes (*TaMAB3*, *4* and *5*) and five rice genes (*OsMBTB6, 7, 8, 26* and *27*). All four *TaMAB* genes (*TaMAB21-24*) which clustered into the core clade have the highest number of exons (4 or 5). Out of 45 *TaMAB* genes clustering

into the expanded clade, most have 1 or 2 exons, with the exception of *TaMAB19* which has 4. Two novel *MATH-BTB* genes from the maize genome, namely *ZmMAB15like* and *ZmMAB17like*, clustered into core clade along with previously reported *ZmMAB15* and *ZmMAB17*, with which they share high sequence similarity, respectively. Interestingly, a single rice gene, *OsMBTB61*, formed its own subclade instead of clustering into the E4 subclade as was previously reported (Juranić & Dresselhaus, 2014).

3.3 Protein alignment of TaMAB2 and selected homologs

From a total of 155 MATH-BTB sequences used to generate the phylogenetic tree shown in Figure 3, several representatives were further selected for multiple protein alignment. Two alignments were made, the first, broad-scope alignment in which proteins were selected from all five clades and subclades containing wheat MATH-BTB proteins (the core clade, E1, E3, E4 and E5 subclade). For the second, E3-specific alignment, all proteins clustering into the E3 subclade along with TaMAB2 were selected. When viewed comparatively, higher overall sequence similarity (in number of black or grey boxes) could be observed in the E3-specific alignment (Fig 6) than in the broad-scope alignment (Fig 5). This was especially prominent within the MATH domain. Additionally, in the broad-scope alignment, there were no identical amino acids within the MATH domain in all selected sequences (Fig 5). However, in the E3-specific alignment, 11 amino acids within the MATH domain were identical in all 9 sequences analyzed (Fig 6). Within the BTB domain, 17 identical amino acids were present in all sequences of the broad-scope alignment (Fig 5), compared to 26 identical amino acids in the E3-specific alignment (Fig 6).

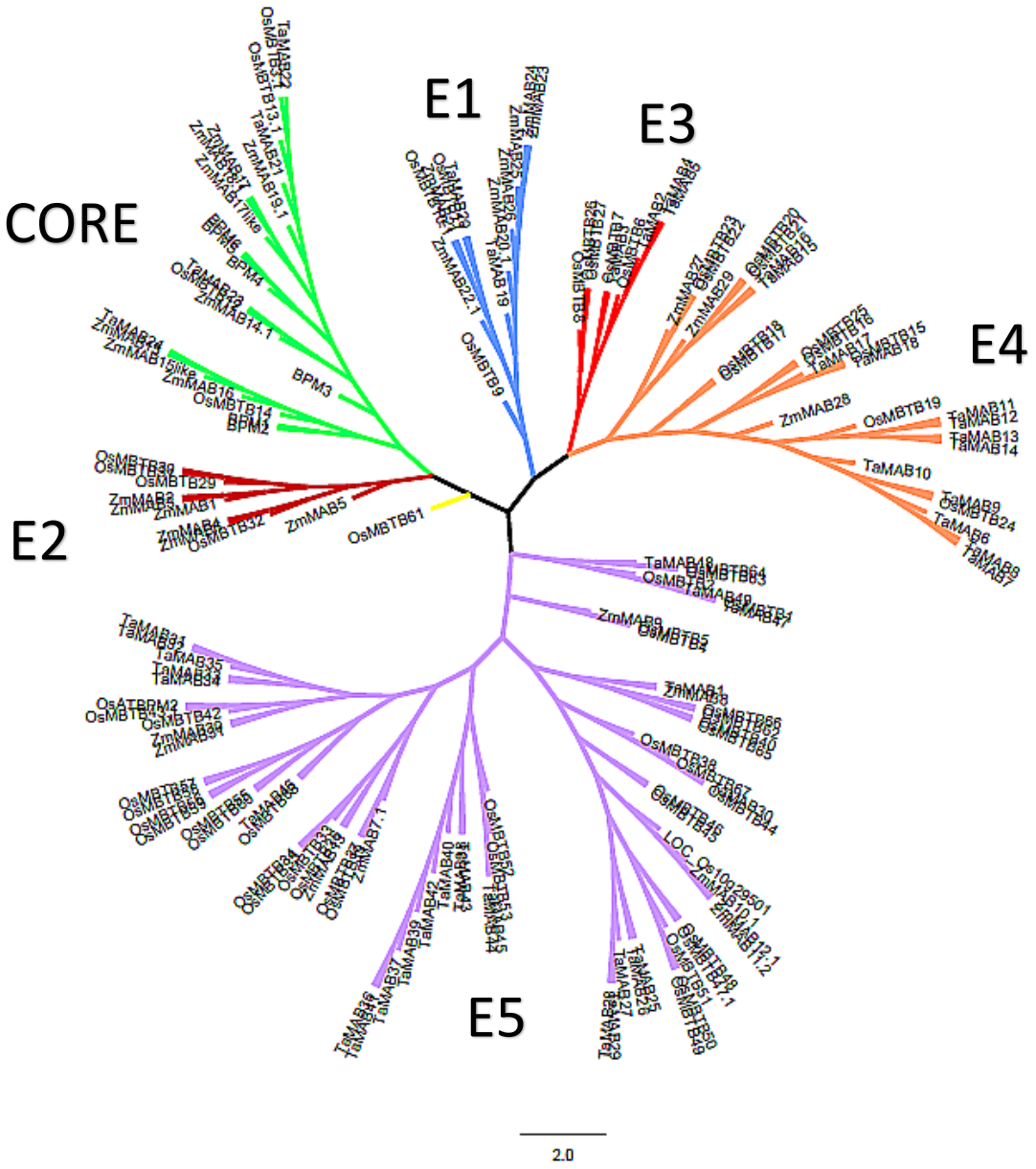


Figure 3. Unrooted phylogenetic tree of 155 *MATH-BTB* genes from four representative land plant species: *T. aestivum, Z. mays, A. thaliana, O. sativa*. Phylogenetic analysis was conducted in SeaView v.4 using nucleotide sequence regions of full-length MATH-BTB proteins. The tree was visualised in FigTree v.1.4.2. Six major subclades are denoted by six different colours. The green-coloured subclade represents the core clade of *MATH-BTB* genes and the remaining five subclades (randomly denominated E1-E5) represent the expanded clade of *MATH-BTB* genes. The yellow-coloured node represents the *OsMBTB61* gene, which did not cluster into any clade. The number after decimal point represents the splicing variant used for phylogenetic analysis. The bar of 2.0 is branch length which represents nucleotide substitutions per site. For sequence identifiers, see Supplemental Tables 1 - 3.

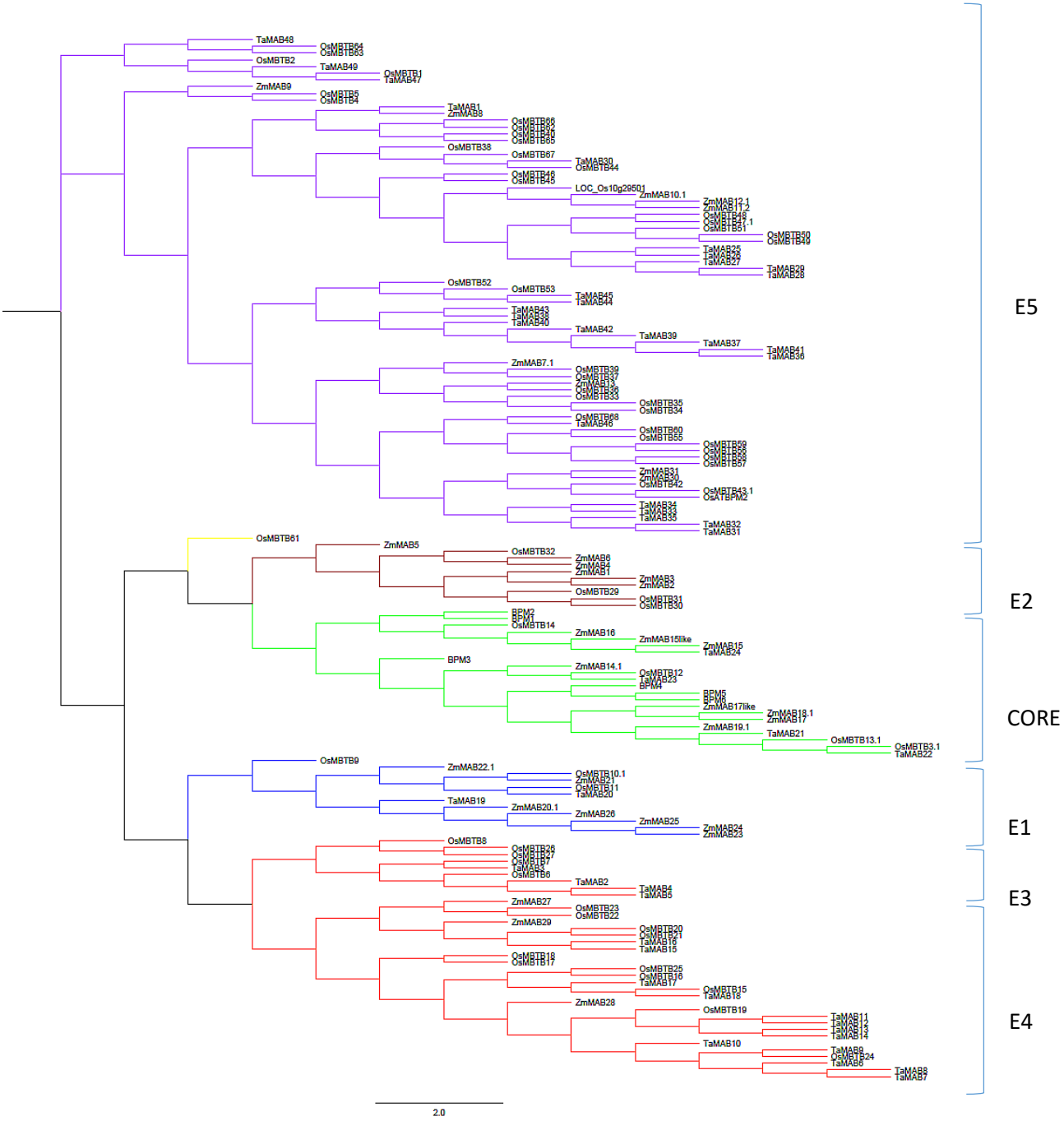


Figure 4. Phylogram of 155 *MATH-BTB* genes from four representative land plant species: *T. aestivum, Z. mays, A. thaliana, O. sativa*. The rooted tree is based on the same nucleotide alignment shown in Figure 3. The green-coloured subclade represents the core clade *MATH-BTB* genes and the remaining five subclades (E1-E5) represent the expanded clade. The number after decimal point represents the splicing variant used for phylogenetic analysis. The bar of 2.0 is branch length which represents nucleotide substitutions per site.

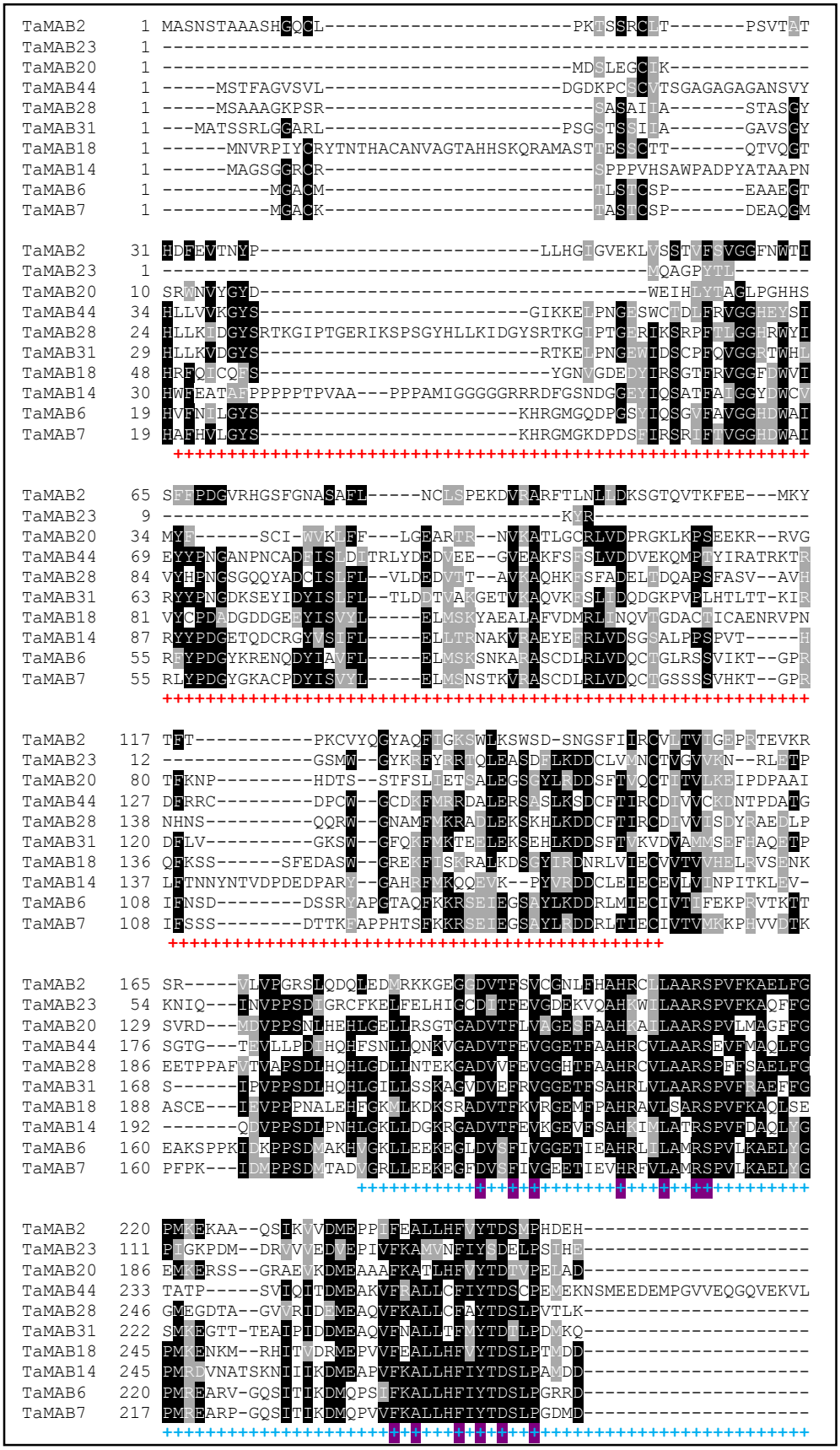


Figure continued on page 38.

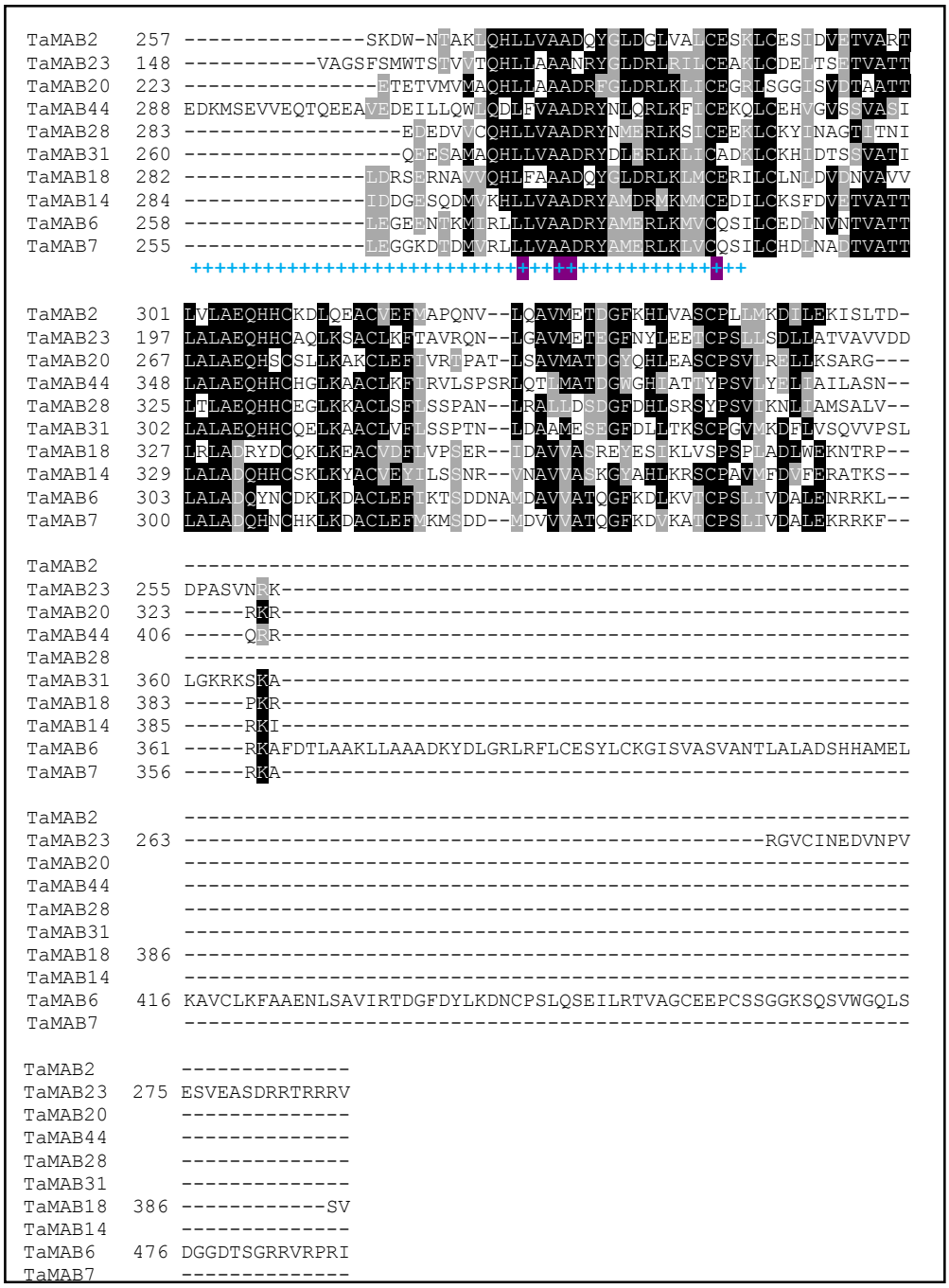


Figure 5. Sequence alignment of 10 wheat MATH-BTB proteins from 5 subclades. Entire amino acid sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) and Boxshade (http://www.ch.embnet.org/software/BOX_form.html) using a threshold of 50% sequence identity. Identical and similar amino acids are shown in black and grey boxes, respectively. Red plus marks (+) encompass the MATH domain and blue plus marks (+) encompass the BTB domain, as predicted by the NCBI database for TaMAB2 protein. Highlighted in purple are the amino acids of the BTB domain present in all 10 proteins.

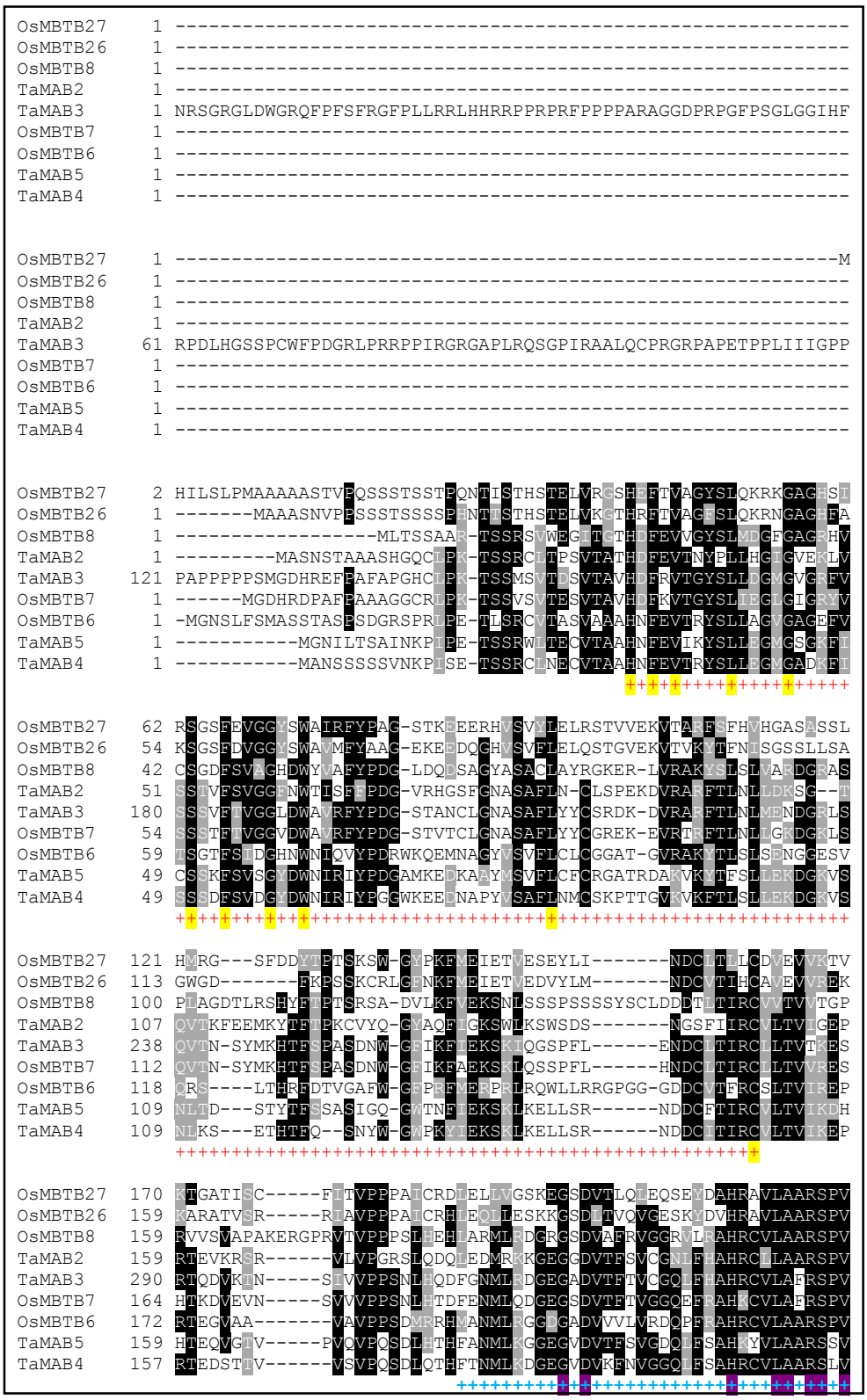


Figure continued on page 40.

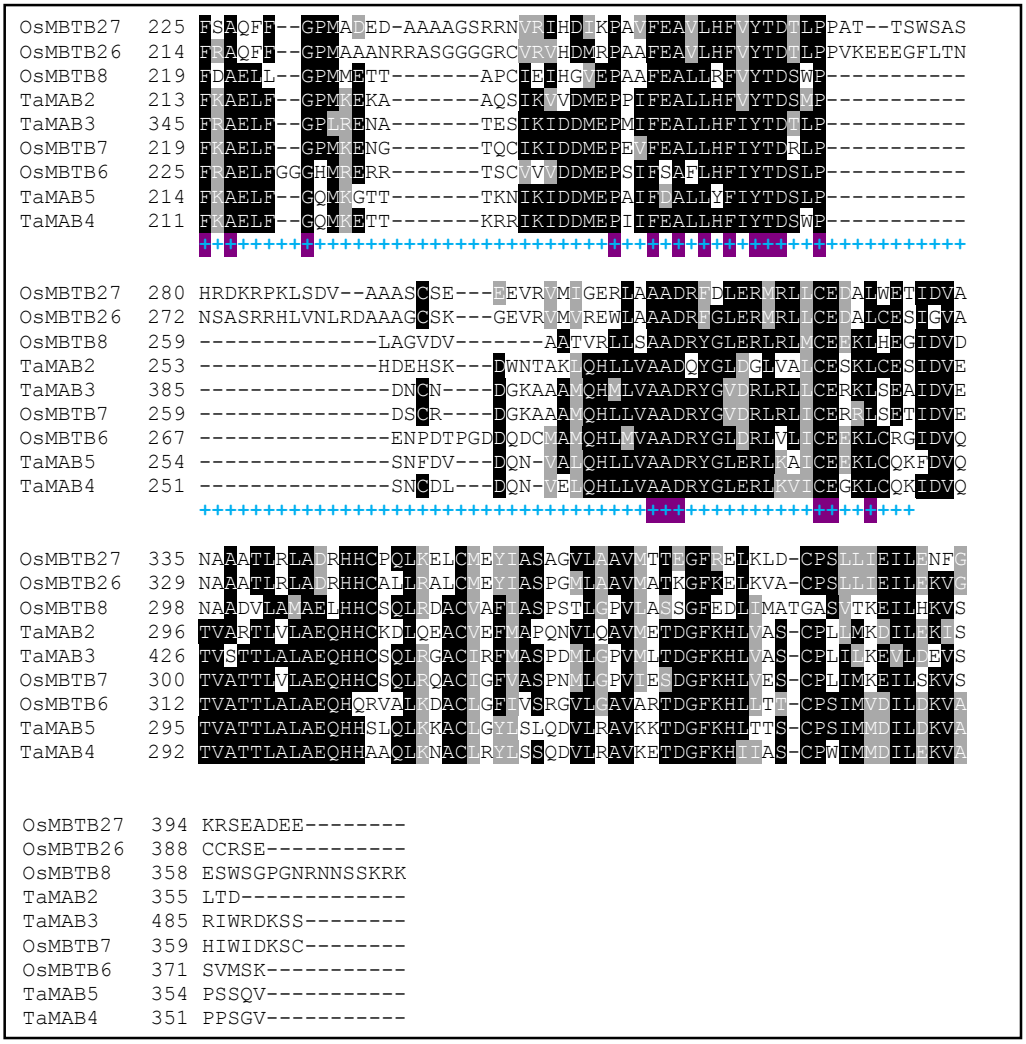


Figure 6. Sequence alignment of 9 wheat MATH-BTB proteins of the E3 subclade. Entire amino acid sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) and Boxshade (http://www.ch.embnet.org/software/BOX_form.html) using a threshold of 50% sequence identity. Identical and similar amino acids are shown in black and grey boxes, respectively. Red plus marks (+) encompass the MATH domain and blue plus marks (+) encompass the BTB domain, as predicted by the NCBI database for TaMAB2 protein. Highlighted in yellow are the amino acids of the MATH domain present in all 9 proteins. Highlighted in purple are the amino acids of the BTB domain present in all 9 proteins.

3.4 Phenotype analysis of *TaMAB2* overexpression and wild type arabidopsis plants

To study the effect of *TaMAB2* overexpression on arabidopsis growth and development, transgenic lines (*35S:TaMAB2-GFP* plants) were used. Both wild type and transgenic lines were grown in the same conditions. Plants were photographed 25, 37 and 47 days after sowing in soil (Figure 7). In the early analysis platform (plates with growth medium), both wild type and transgenic seeds went through characteristic stages of development (as described in Boyes et al., 2001): imbibition, radicle emergence, hypocotyl and cotyledons emergence from seed coat, opening of cotyledons and formation of two, then four rosette leaves (data not shown). In the soil-based platform, wild-type plants (Fig 7, panel 1) continued to exhibit expected characteristics: appearance of ten rosette leaves (Fig 7A), formation of first flower buds (Fig 7B, small square), first flower opening, midflowering stage and completion of flowering with developed siliques (Fig 7C). Remarkably, transgenic plants exhibited an array of phenotypes, which I have classified into three types, following the growth and development of a representative specimen from each type (Fig 7, panels 2-4). Phenotype 1 (Fig 7, panel 2) resembled wild type plants in that it went through all the characteristic stages mentioned above, with no macroscopically observable characteristics of a mutant phenotype. Phenotype 2 (Fig 7, panel 3) exhibited features normally absent in wild type plants. Following the formation of the first 6-7 leaves, which appeared similar to wild type leaves, the younger leaves of the rosette began to curl outwards, resulting in a rugose leaf blade (Fig 7G, H). Each new circle of leaves was smaller and more rugose than the previous one. Moreover, the leafstalks of these inner leaves failed to elongate, positioning younger leaves close to the main stalk (Fig 7H). No flower buds were observed at this stage. 47 days after sowing, some leaf growth could be observed (Fig 7I) but the leaves remained in a rosette formation, there was no stalk elongation and no flowers emerged. Phenotype 3 (Fig 7, panel 4) exhibited all the characteristics observed in phenotype 2 but with greater extent. Thus, 37 days after sowing, a dense bundle of leaves developed close to the plant stalk (Fig 7K), and subsequently failed to grow. This growth regression began after the eight rosette leaf was formed, with the bundle becoming visibly smaller and the rosette leaves wilting (Fig 7L).

25 days after sowing	37 days after sowing	47 days after sowing

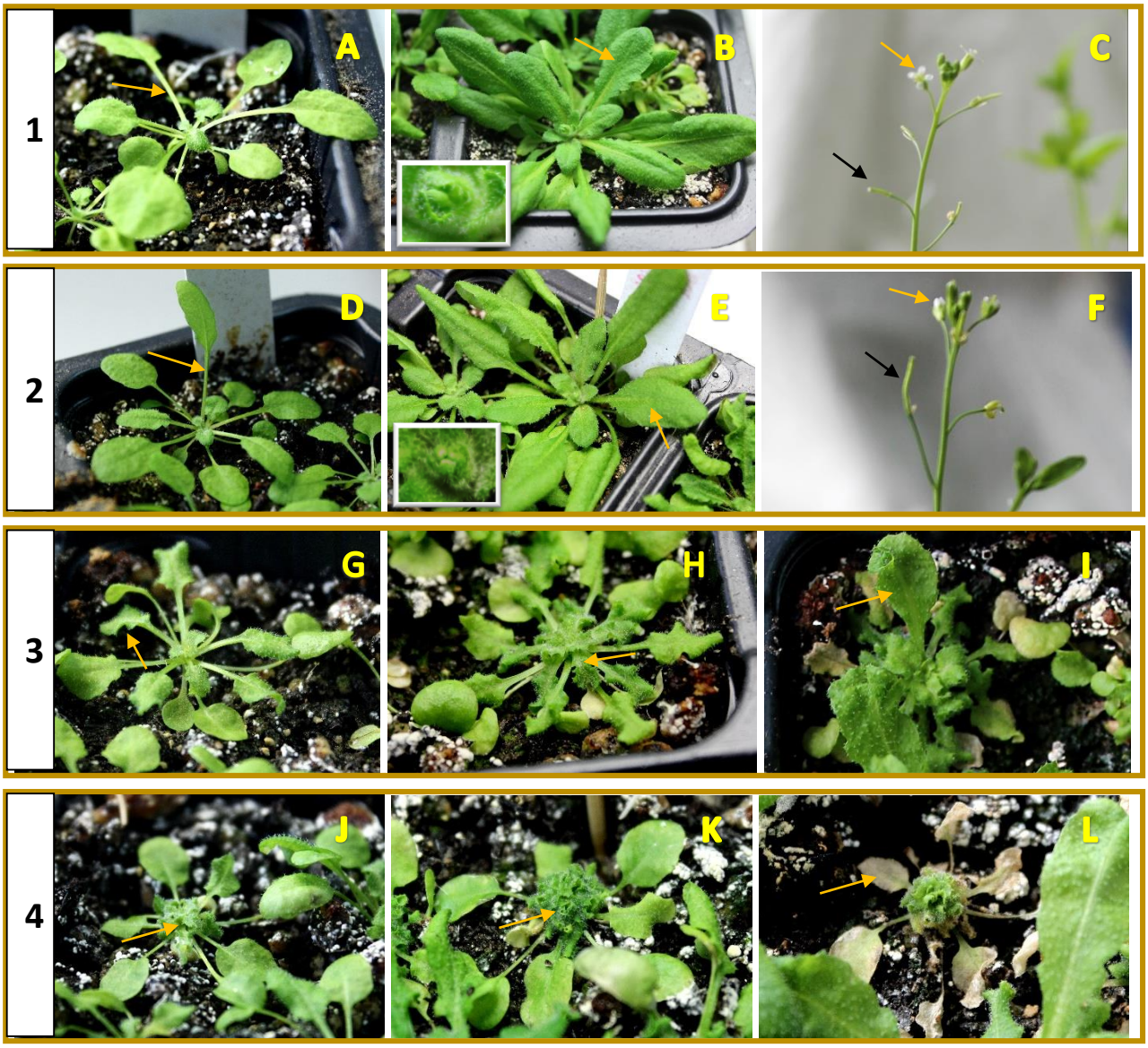


Figure 7. Wild type and *TaMAB2* overexpression lines of *Arabidopsis thaliana* in a soil-based platform. **(Panel 1**) Wild type plants with elongated rosette leaves (**A**, arrow), leaf elongation (**B**, arrow), developed flower buds (**B**, inset) flowers (**C**, yellow arrow) and siliques (**C**, black arrow). **Panels 2-4** show the three *TaMAB2* overexpression phenotypes. **(Panel 2)** Phenotype 1, with elongated rosette leaves (**D**, arrow), leaf elongation (**E**, arrow), developed flower buds (**E**, inset) flowers (**F**, white arrow) and siliques (**F**, black arrow). **(Panel 3)** Phenotype 2, with rugose leaf blades (**G**, arrow), failed leaf stalk elongation (**H**, arrow) and minor leaf growth in later stages (**I**, arrow). **(Panel 4)** Phenotype 3, with a bundle of small and rugose leaves near the main stem (**J**, arrow), failed leaf stalk elongation (**K**, arrow) and wilting of rosette leaves (**L**, arrow). Plants were photographed 25 (column 1), 37 (column 2) and 47 (column 3) days after sowing.

3.5 Isolation of zygotic embryos from *TaMAB2* overexpression and wild type arabidopsis plants

Zygotic embryos were isolated from Phenotype 1 *TaMAB2* overexpression lines and wild type line of arabidopsis. Siliques were chosen based on position on the plant stem, specifically the fourth and the seventh silique from top of plant for both the transgenic line and the wild type. No major morphological differences were observed between wild type and transgenic line embryos. Silique 4 of wild type contained both immature ovules possibly containing early globular stages of embryos and mature ovules containing later stages (torpedo or cotyledonary). Silique 4 of transgenic line mostly contained immature ovules (Fig 8B). In one silique, a heart-stage embryo was found (Fig 8C). Silique 7 of wild type plants contained mostly mature ovules. The majority of embryos found there were heart-stage embryos (Fig 8A). Silique 7 of transgenic plants contained a wide array of developmental stages. The ovules were either immature (Fig 8D) or completely mature with globular-stage embryos (Fig 8E), heart-stage embryos (Fig 8F), torpedo-stage embryos (Fig 8G) and cotyledonary-stage embryos (Fig 8H). Due to technical difficulties with the isolation process and thus a small sample size, no statistically relevant results could be assembled.

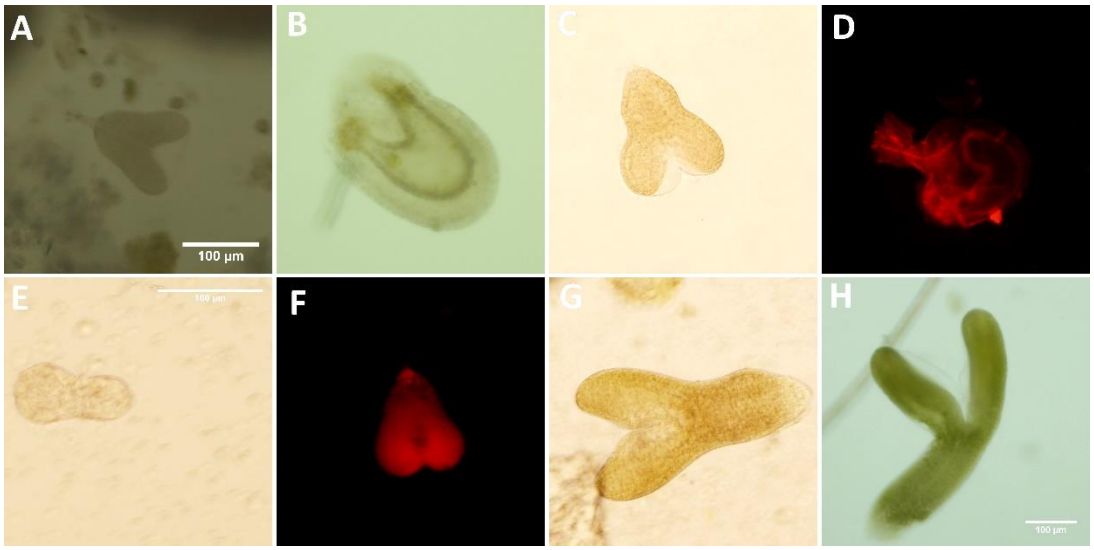


Figure 8. Zygotic embryos isolated from **(A)** wild type and **(B-H)** TaMAB2 overexpression lines of arabidopsis plants. **(A)** Heart-stage embryo isolated from wild type silique **(**7th from top of plant). **(B)** Immature ovule and **(C)** heart-stage embryo isolated from transgenic line siliques (4th from top of plant). **(D)** Immature ovule, **(E)** early globular embryo, **(F)** heart-stage embryo, **(G)** torpedo-stage embryo and **(H)** cotyledonary-stage embryo isolated from transgenic line siliques (7th from top of plant). Images **(D)** and **(F)** show samples stained with propidium iodide (PI, excited at 493 nm). Scale bar = 100 µm. Images without a scale bar correspond to (A) in scale size.

3.6 Somatic embryogenesis in *TaMAB2* overexpression and wild type arabidopsis plants

To induce somatic embryogenesis (SE), immature zygotic embryos (IZE) of wild type and transgenic plants were plated onto E5 induction medium supplemented with 2,4-dichlorophenoxyacetic acid (2,4-D) and then transferred onto E5 medium without 2,4-D. The development of somatic embryos was monitored during the early induction phase and the late maturation phase. Both wild type and transgenic line explants displayed a successful induction phase. The previously bent cotyledons of IZE started to straighten and expand and the cotyledonary node started to swell. By the end of second week of culture, most explants developed immature somatic embryos directly, on the adaxial side of cotyledons, in the area proximal to the cotyledonary node (Fig 9A and B). Formation of callus tissue could be observed on the abaxial side of cotyledons (shown in Fig 9B). Callus tissue usually forms somatic embryos by the end of week 3 of culture, through a process of indirect embryogenesis (not shown). An early attempt of setting up a culture for somatic embryogenesis resulted with mostly degenerated zygotic embryos displaying intensive callus growth (shown in Fig 9C). This was most likely due to errors in the isolation process, such as causing extensive damage to the explant tissue. During the late phase of somatic embryogenesis, wild type and transgenic line explants showed a distinct difference in somatic embryo development. The majority of wild type explants formed clusters of trumpet-shaped embryos (Fig 10A) with a 53% frequency of SE. Transgenic line explants, however, showed only 4% frequency of SE. The formation of somatic embryos could be observed in some explants (Fig 10B) but most explants were either highly degenerated (Fig 10C and D) or developed only callus tissue on all sides of cotyledons (Fig 10E and F). In the early attempt of somatic embryo culture, the frequency of SE was 42% for wild type and 33% for transgenic line. Taken together, the frequency of SE averages to 47.5% for wild type and 18.5% for transgenic line.

3.7 Protein interaction analysis

3.7.1 Cloning of selected genes in Y2H system

For protein interaction analysis, a Gal4-based yeast two hybrid assay was performed. Four proteins were tested for interaction with TaMAB2: the cytoskeletal protein actin 11 expressed

predominantly during reproductive development (from *A. thaliana*), the ubiquitously expressed histone variant H3.3 (from *A. thaliana*), the RNA binding protein with nuclease activity Tudor-SN1 (from *A. thaliana*) and the microtubule-severing protein Katanin (from wheat, *T. aestivum*). Therefore, all five genes had to be separately cloned in a Y2H system. The cloning proved successful: each individual yeast clone exhibited colony growth on a selective medium for growth of single transformants (not shown). Actin 11, H3.3 and Tudor-SN1 were tested for interaction with TaMAB2 in both orientations, meaning that each potential interactor was fused with either the AD or BD of Gal4 and then brought in close proximity with TaMAB2 fused with either the BD or AD of Gal4, respectively. Neither of these three proteins showed positive results in the His prototrophy assay or the β-galactosidase assay using X-gal as substrate (data not shown). To assess its suspected homodimerization ability, TaMAB2 was additionally tested for interaction with itself and a positive result was obtained in the His prototrophy assay. Although the β-galactosidase assay failed to show signs of interaction, a weak positive result was obtained in the liquid β-galactosidase assay using ONPG as substrate (Fig 11).

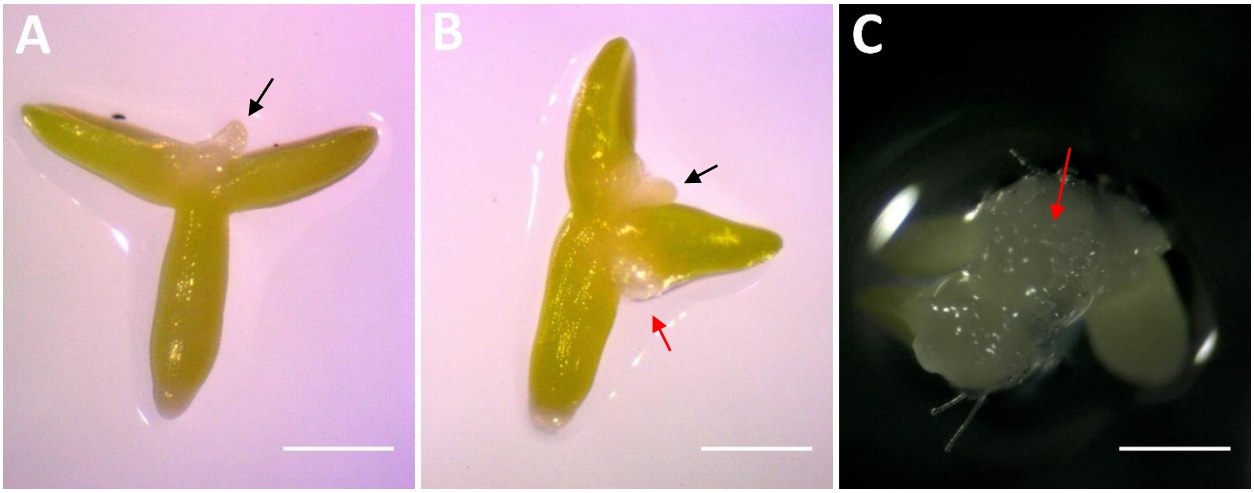


Figure 9. Induction phase of somatic embryogenesis in **(A)** wild type and **(B, C)** *TaMAB2* overexpression line of *A. thaliana*. Early somatic embryos are visible on the adaxial side of cotyledons of immature zygotic embryos (**A, B**, black arrows). Panels **(B)** and **(C)** show callus tissue growing from the explant (red arrows). Scale bar = 1 mm.

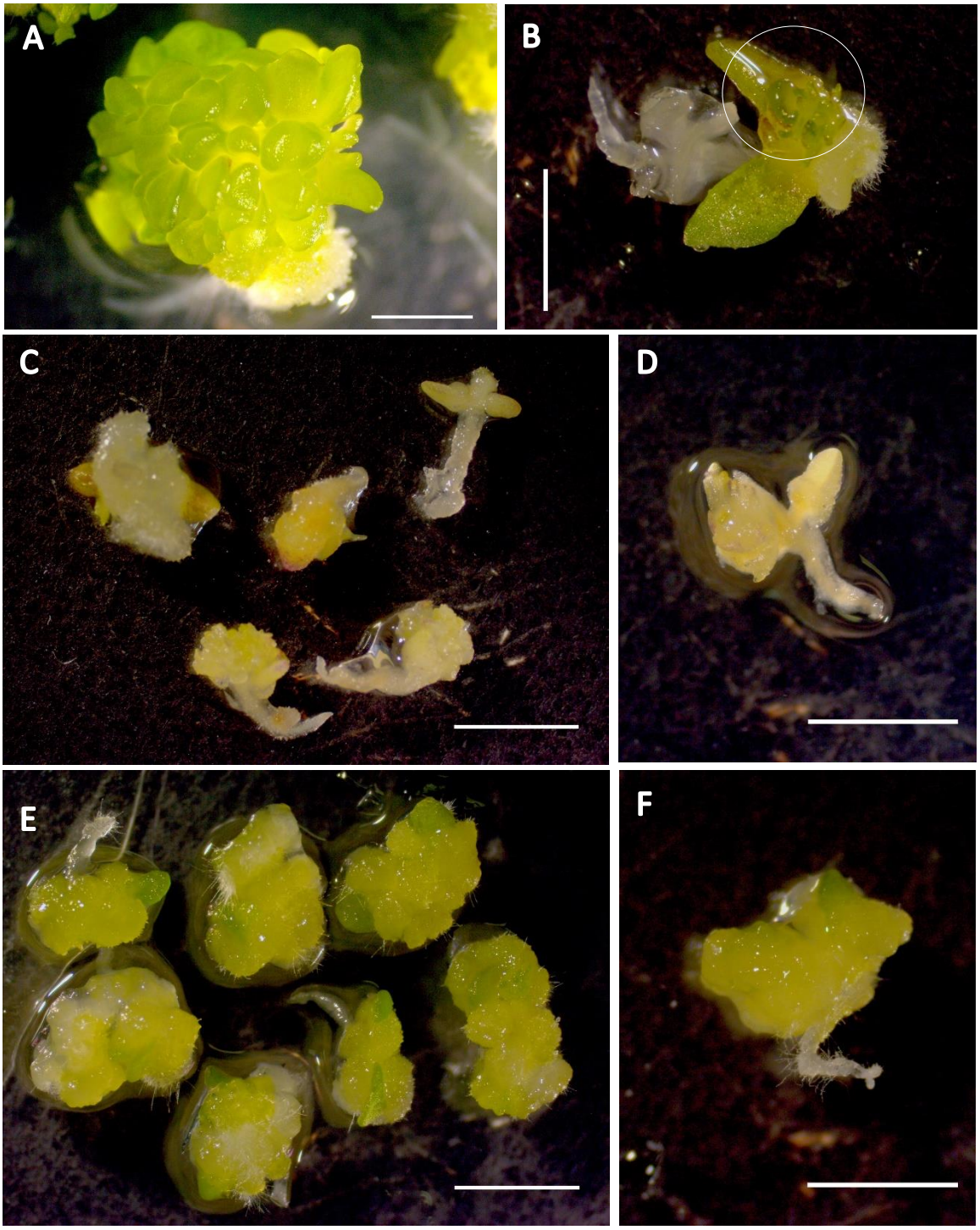


Figure 10. Late phase of somatic embryogenesis in **(A)** wild type and **(B-F)** *TaMAB2* overexpression line of *A. thaliana*. **(A)** Clusters of trumpet-shaped embryos are visible on the surface of wild type and **(B)** transgenic line explants (circled). **(C, D)** Highly degenerated transgenic line explants with no observable SE. **(E, F)** Transgenic line explants with intensive callus growth on all sides of IZE cotyledons but no SE. Scale bar = 2 mm.

3.7.2 Katanin is a potential substrate of TaMAB2

Despite several attempts, no bacteria carrying the *Katanin* gene cloned into the pGAD424 plasmid could be obtained. Therefore, only Katanin fused with BD of Gal 4 was cloned into the Y2H system and tested for interaction with AD-TaMAB2 fusion product. Positive result was detected in all three assays performed: histidine prototrophy assay, β-galactosidase assay using X-gal as substrate and liquid β-galactosidase assay using ONPG as substrate (Fig 11).

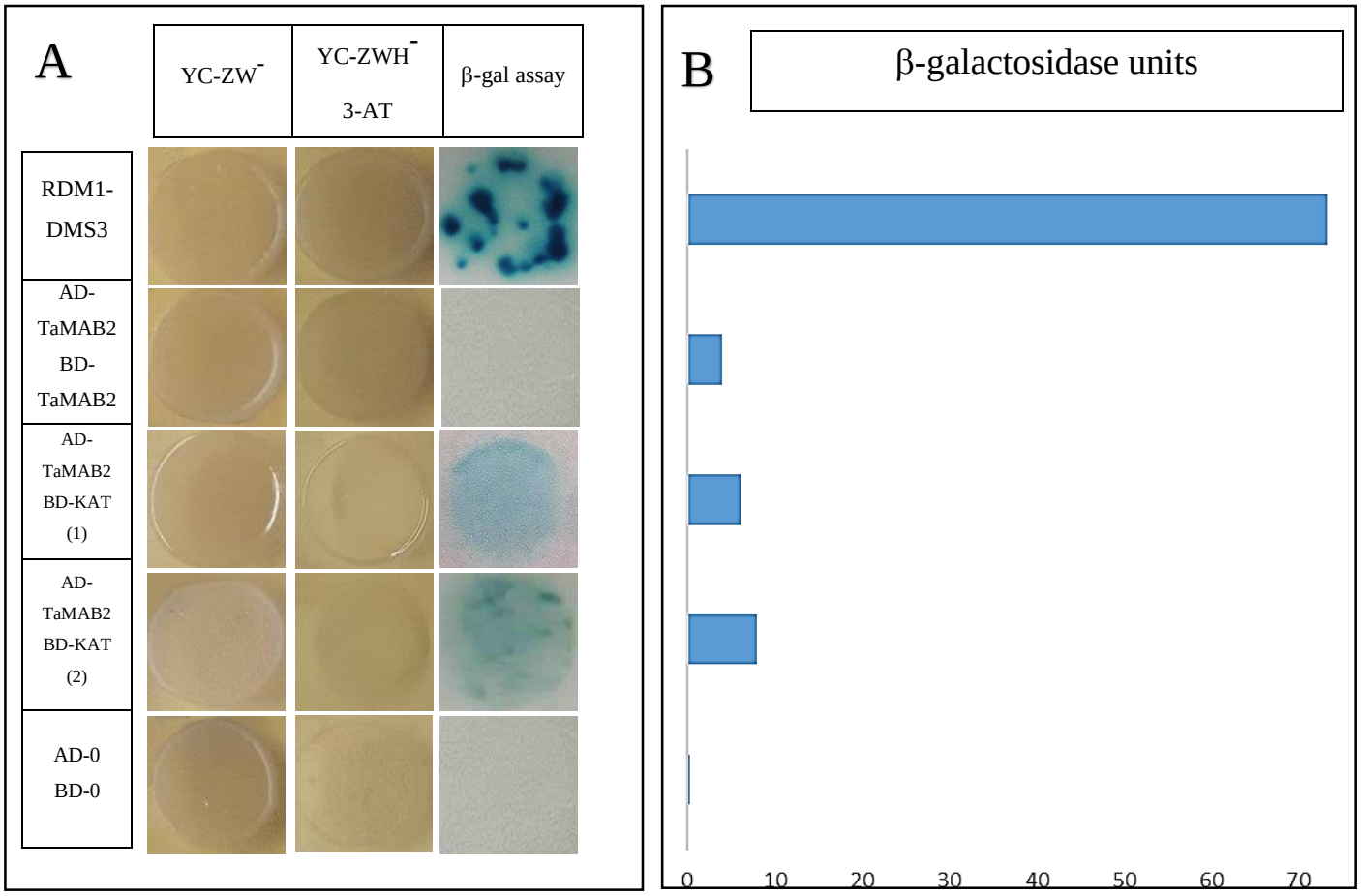


Figure 11. Yeast two-hybrid assay. **(A)** His prototrophy assay and β-galactosidase (β-gal) assay. Interaction of AD-TaMAB2 and BD-KAT is shown as result of two individual assays: (1) and (2). YC-ZW$^-$ medium lacks leucine and tryptophan, for growth of all co-transformants (left column). YC-ZWH$^-$ 3-AT medium lacks leucine, tryptophan and histidine, for growth of co-transformants with positive protein-protein interaction in the His-prototrophy assay (middle column). 3-AT inhibits unspecific interactions. β-gal assay (right column) where blue colour signals positive protein-protein interaction. In the β-gal assay, no interaction was shown for AD-TaMAB2 and BD-TaMAB2. **(B)** Liquid β-galactosidase assay using ONPG as substrate. Blue bars represent β-galactosidase units (one equals the amount of β−galactosidase which hydrolyses 1 μmol of ONPG to *o*-nitrophenol and D-galactose per minute per cell). RDM1-DMS3 is positive control and empty prey (AD-0) and bait (BD-0) vectors are negative control.

4 DISCUSSION

4.1 Phylogeny of wheat MATH-BTB proteins and selected homologs

The results of a phylogenetic analysis conducted within this research concur with several previously published works on the phylogeny of plant *MATH-BTB* genes. These analyses started with arabidopsis and rice *MATH-BTB* genes (Gingerich et al., 2005) but were later expanded to include other monocots (sorghum and wheat), dicots (*Medicago truncatula* and poplar), a gymnosperm (pine) a moss and a bryophyte (Gingerich et al., 2007), and more recently maize (Juranić et al., 2012), *Brachypodium distachium*, banana, grapevine and a lycophyte *Selaginella moellendorffi* (Juranić & Dresselhaus, 2014). The result reported unanimously in those works and reiterated here is the existence of two distinct clades of plant *MATH-BTB* genes: the smaller core clade and the expanded clade, which further clusters into five subclades. As of yet, wheat *MATH-BTB* genes have not been phylogenetically analysed. However, none of the previously analysed dicots clustered into the expanded clade, making it a monocot-specific clade as proposed by Gingerich et al. (2007). Furthermore, according to a hypothesis proposed and confirmed by Juranić & Dresselhaus (2014), the expanded clade is also grass-specific, with none of the analysed non-grass monocots (banana), dicots (*A. thaliana*) or eudicots (grapevine) clustering into it. Therefore, it was also expected of *TaMAB* genes to cluster primarily into the expanded clade. Another interesting observation regarding *TaMAB* genes is the obvious lack of splicing variants. This could be explained by a theory proposed by Sasidharan and Gerstein (2008) and revisited by Juranić & Dresselhaus (2014), according to which the number of splicing variants reflects the number of genes in a gene family. For instance, arabidopsis has 6 *MATH-BTB* genes, of which 3 genes have 2 splicing variants. Furthermore, of the 2 known *MATH-BTB* genes in humans, one is speculated to have 23 splicing variants. In comparison, the *MATH-BTB* gene family appeared to have expanded in the three grasses; sorghum (47 genes), *Brachypodium* (46 genes) and wheat (49 genes); but was accompanied by a decrease in splicing variants, with sorghum and *Brachypodium* each having only 1 gene with 2 splicing variants and wheat having no splicing variants. Further corroborating this theory, the gene loss process which possibly occurred in the maize genome, resulting in 31 *MATH-BTB* genes, was accompanied by an increase in the number of splicing variants (10 genes with 2, 3 or even 5 splicing variants). Juranić & Dresselhaus (2014) proposed

that the splicing variants present in the maize genome possibly exert a regulatory role (via short interfering RNAs) on full-length MATH-BTB proteins, thus alleviating the effects of a major genetic loss. It would be interesting to study the genetic events possibly occurring within the wheat genome to better understand the evolutionary pathways shaping the *MATH-BTB* family in plants.

Of the 49 analysed *TaMAB* genes, *TaMAB21-24* clustered in the core group. It was previously reported that the core group genes show the highest level of evolutionary conservation (Gingerich et al., 2007; Juranić & Dresselhaus, 2014) and that they possibly share the same function. Of the core group genes analysed here, only arabidopsis genes have been functionally studied and brought in relation with abscisic acid response (Lechner et al., 2011) ethylene response (Weber & Hellman, 2009) and regulation of flowering (Chen et al., 2014). It would be interesting to study the function of the wheat *TaMAB* core group genes to address this hypothesis. Moreover, a possibility arises that genes in each subclade of the expanded clade also share some functional similarity (Juranić & Dresselhaus, 2014). The only TaMAB protein studied in more detail is the TaMAB2, which presumably regulates the zygotic division and the establishment of polarity or cell fate in the developing embryo (Leljak-Levanić et al., 2013). It would therefore be interesting to see whether *TaMAB2* shares the same functions with *TaMAB3*, *4* and *5*, which cluster along with *TaMAB2* in the E3 subclade. Presuming that functional similarity is preceded by similarity in amino acid sequence, the results of multiple sequence alignments presented here further corroborate the hypothesis that proteins within a specific subclade share similar functions. In MATH-BTB proteins which function as part of a Cul3-based E3 ligase, MATH domain is responsible for target recognition and binding. Higher overall sequence similarity, but more specifically, higher sequence similarity in the MATH domain observed among members of the E3 subclade could mean that proteins of the E3 subclade (four wheat and five rice proteins) bind similar substrates. TaMAB2 protein is presumed to interact with Katanin (discussed in section 4.2). Protein interaction analyses could therefore be conducted in which other members of the E3 subclades are tested for interaction with Katanin. Furthermore, of the 11 amino acids fully conserved among all proteins of the E3 subclade, some might be accountable for specificity of substrate recognition and/or binding. Thus, future multiple sequence alignments could include other plant MATH-BTB proteins which cluster into the E3 subclade to narrow down the number of highly conserved amino acids between them and subsequent biochemical analyses could reveal

whether these amino acids are genuinely responsible for recognition and/or binding of their potential substrates.

4.2 TaMAB2 interacts with Katanin and itself

Very little is known about the function of wheat MATH-BTB proteins. So far, one study has reported that *TaMAB2* gene expression was detected in the zygote and the two-celled embryo, with subsequent downregulation in later stages (Leljak-Levanić et al., 2013) but no functional analyses have been made on TaMAB2 or any of its wheat homologs, to date. The result presented in this work suggests a possible interaction of TaMAB2 and the microtubule-severing protein Katanin, shedding some light on its physiological function during wheat embryogenesis. Several studies reported an interaction between a BTB domain-containing protein and Katanin. A MATH-BTB protein of *C. elegans*, MEL-26, recruits the katanin protein MEI-1 required for meiotic spindle formation. This interaction targets MEI-1 for ubiquitin-dependent degradation prior to mitosis (Pintard et al., 2004). A similar mechanism for mitotic regulation was observed in mammalian cells, where the catalytic subunit of Katanin is targeted for degradation via the Kelch repeat-containing BTB adaptor protein KLHDC5 (Cummings et al., 2009). According to Genschik et al. (2013), this mechanism seems to be conserved among metazoans. However, a plant MATH-BTB protein from maize, ZmMAB1, also appears to interact with the catalytic subunit of Katanin. Like MEL-26, ZmMAB1 likely acts as a substrate-specific adaptor of a Cul3-based E3 ligase. Mutations in ZmMAB1 also lead to chromosome segregation defects and short spindle formation during meiosis. Therefore, it has been suggested that ZmMAB1 and MEL-26 have a similar role in organising microtubular spindles during meiosis-to-mitosis transition (Juranić et al., 2012). The amino acid sequences of ZmMAB1 and TaMAB2 display significant similarity (Leljak-Levanić et al., 2013) and both proteins appear to localize in the nucleus and cytoplasm (Juranić et al., 2012; Leljak-Levanić et al., 2013), although the cytoplasmic presence of TaMAB2 was possibly an anomaly caused by ectopic over-expression of the fusion protein (Leljak-Levanić et al., 2013). Nevertheless, the novel finding that TaMAB2 interacts with Katanin corroborates an early assumption of Leljak-Levanić et al. (2013), that TaMAB2 protein functions as part of a Cul3-based E3 ligase in the zygote and/or proembryo where it possibly, like MEL-26 and ZmMAB1, regulates

the mitotic spindle assembly and consequently the zygotic division and progression of early embryogenesis.

4.3 *TaMAB2* overexpression has an ambiguous effect on arabidopsis development

Because *TaMAB2* knockout mutants of wheat could not be obtained (Leljak-Levanić, unpublished), *Arabidopsis thaliana* plants were transformed to obtain transgenic lines with overly expressed *TaMAB2* gene. The analysis of *TaMAB2* overexpression lines revealed an interesting phenomenon. Transgenic lines displayed an array of phenotypes which could be ranged by intensity of mutation effect. Some plants displayed near wild-type characteristics, while others exhibited extremely retarded growth and failed to elongate plant stems and produce flowers. This could be due to mutagenicity of the plant transformation process itself. Transgene insertion, a central feature of plant transformation, is not often a precise event, resulting in deletions, rearrangements of host chromosomal DNA and introduction of superfluous DNA at transgene insertion sites. These unintended mutations have the potential to result in inadvertent loss, acquisition, or misexpression of important traits, partly because transgenes insert into or near functional genes. Furthermore, when transgene insertion causes rearrangements or insertion of superfluous DNA then juxtaposition of promoter sequences and coding fragments may produce sense or antisense transcripts which, like siRNAs and miRNAs, might interfere with expression of genes containing homologous or similar sequences (Latham et al., 2006). Therefore, further genetic analysis should be carried out to establish whether the phenotypes observed in this work are a consequence of intended *TaMAB2* overexpression or unintended mutations caused by technical shortcomings of the transformation procedure. Nevertheless, one of the observed traits in nearly all specimens of the *TaMAB2* overexpression line was failed elongation of the plant stem, leaf stalks and leaf blades. As discussed in section 4.2., the interaction of TaMAB2 and Katanin would presumably lead to degradation of Katanin by the 26S proteasome. If this assumption is correct, Katanin levels might be depleted in TaMAB2 overexpression lines. However, an analysis of *Arabidopsis thaliana* lines with overexpressed Katanin (AtKTN1) revealed a similar effect on arabidopsis phenotype as was observed in *TaMAB2* overexpression lines: reduced expansion of cells and organs, and a decrease in cell wall strength (Burk et al., 2007). In order for any conclusion

to be made regarding the *TaMAB2* overexpression plants, further gene expression analyses should be conducted.

Analysis of zygotic embryogenesis in *TaMAB2* overexpression lines of arabidopsis did not reveal any major phenotypic discrepancies when compared to wild type embryos. However, adjustments in the isolation process, specifically a greater sample size and optimization of the isolation protocol to achieve maximum embryo yield could be of great benefit for future research.

Regarding the effect of *TaMAB2* overexpression on somatic embryogenesis of arabidopsis, a more straightforward observation can be made. Specifically, *TaMAB2* overexpression seems to have a strong inhibitory effect on somatic embryo development in culture, which becomes more evident in later stages of SE. Although the transgenic line showed a relatively high frequency of SE (33%) in the early attempt of setting up an SE culture, due to high level of tissue damage inflicted upon zygotic embryos during isolation, those results are not given as much gravity as results obtained in the second attempt (4% frequency of SE). Since exact function of TaMAB2 is not known, the effect of TaMAB2 on SE on a molecular level can only be speculated. TaMAB2 presumably acts as part of Cul3-based E3 ubiquitin ligase (Leljak-Levanić et al., 2013). Therefore, its overexpression might lead to excessive degradation of its target substrates (or, specifically, their homologues in arabidopsis) which might account for the observed inability of transgenic explants to produce somatic embryos.

5 CONCLUSION

Searching through the NCBI and EnsemblPlants databases with TaMAB2 sequence as query resulted in 46 novel wheat MATH-BTB proteins (TaMAB proteins). Phylogenetic analysis of these proteins, along with known and previously unreported MATH-BTB proteins of maize, rice and arabidopsis revealed major clustering of TaMAB proteins in the expanded, grass-specific clade. The four proteins clustering in the core clade had the highest number of exons. TaMAB2 protein clustered into the E3 subclade of the expanded clade. Multiple sequence alignment of selected wheat MATH-BTB proteins from all subclades, as well as MATH-BTB proteins from the E3 clade only, revealed higher sequence similarity between proteins of the E3 subclade. The similarity was especially prominent within the MATH domain. This finding corroborates the hypothesis of higher functional similarity between proteins of a specific subclade. Yeast two-hybrid assay revealed an interaction of TaMAB2 protein with microtubule-severing protein Katanin, making Katanin a possible physiological substrate of TaMAB2 during zygote and early embryo development. Additionally, TaMAB2 showed a weak interaction with itself, suggesting its homodimerization ability. *TaMAB2* overexpression had an ambiguous effect on growth and development of adult arabidopsis plants, resulting in a wide array of mutant phenotypes, most of which were characterized by failed elongation of all plant tissues and, to a partial extent, inability to produce flowers. Additionally, *TaMAB2* overexpression appeared to have an inhibitory effect on somatic embryogenesis in arabidopsis zygotic embryo culture.

6 REFERENCES

Abrash, E. B., & Bergmann, D. C. (2009). Asymmetric Cell Divisions: A View from Plant Development. *Developmental Cell*, *16*, 783–796.

Agatep, R., Kirkpatrick, R. D., Parchaliuk, D. L., Robin, A., & Gietz, R. D. (1998). Transformation of Saccharomyces cerevisiae by the lithium acetate/single-stranded carrier DNA/ polyethylene glycol protocol. *Technical Tips Online*, *3*, 133–137.

Boscá, S., Knauer, S., & Laux, T. (2011). Embryonic development in Arabidopsis thaliana: from the zygote division to the shoot meristem. *Frontiers in Plant Science*, *2*, 1–6.

Boyes, D. C., Zayed, A. M., Ascenzi, R., Mccaskill, A. J., Hoffman, N. E., Davis, K. R., & Görlach, J. (2001). Growth Stage – Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants. *Plant Cell*, *13*, 1499–1510.

Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., & Laux, T. (2008). Differential Expression of WOX Genes Mediates Apical-Basal Axis Formation in the Arabidopsis Embryo. *Developmental Cell*, *6*, 867–876.

Burk, D. H., Burk, D. H., Zhong, R., & Ye, Z. (2007). The Katanin Microtubule Severing Protein in Plants. *Journal of Integrative Plant Biology*, *49*, http://doi.org/10.1111/j.1672-9072.2007.00544.x

Chen, L., Bernhardt, A., Lee, J., & Hellmann, H. (2014). Identification of Arabidopsis MYB56 as a Novel Substrate for CRL3 BPM E3 Ligases. *Molecular Plant*, *8*, 242–250.

Chen, L., Lee, J. H., Weber, H., Tohge, T., Witt, S., Roje, S., Fernie, A. R. & Hellmann, H. (2013). Arabidopsis BPM proteins function as substrate adaptors to a cullin3-based E3 ligase to affect fatty acid metabolism in plants. *Plant Cell*, *25*, 2253-2264.

Cheng, D., Qian, W., Meng, M., Wang, Y., Peng, J., & Xia, Q. (2014). Identification and Expression Profiling of the BTB Domain-Containing Protein Gene Family in the Silkworm, Bombyx mori. International Journal of Genomics, http://dx.doi.org/10.1155/2014/865065

Dow, M. R., & Mains, P. E. (1998). Genetic and molecular characterization of the Caenorhabditis elegans gene, mel-26, a postmeiotic negative regulator of mei-1, a meiotic-specific spindle component. *Genetics*, *150*, 119–128.

Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*, 1792-1797.

Fehér, A. (2008). The initiation phase of somatic embryogenesis: what we know and what we don't. *Acta Biologica Szegediensis, 52*, 53–56.

Fehér, A. Pasternak, T. P. & Dudits, D. (2003). Transition of somatic plant cells to an embryogenic state. *Plant Cell, Tissue and Organ Culture*, *74*, 201-228.

Feilotter, H. E., Hannon, G. J., Ruddell, C. J., & Beach, D. (1994). Construction of an improved host strain for two hybrid screening. *Nucleic Acids Research*, *22*, 1502–1503.

Figueroa, P., Gusmaroli, G., Serino, G., Habashi, J., Lingeng, M., Shen, Y., Feng, S., Bostick, M., Callis, J., Hellman, H. & Deng, X. W. (2005). Arabidopsis Has Two Redundant Cullin3 Proteins That Are Essential for Embryo Development and That Interact with RBX1 and BTB Proteins to Form Multisubunit E3 Ubiquitin Ligase Complexes in Vivo. *The Plant Cell*, *17*, 1180–1195.

Gaj, M. (2011). Somatic Embryogenesis and Plant Regeneration in the Culture of Arabidopsis thaliana (L.) Heynh. Immature Zygotic Embryos. *Methods in Molecular Biology*, *710*, 257-265.

Genschik, P., Sumara, I., & Lechner, E. (2013). The emerging family of CULLIN3-RING ubiquitin ligases (CRL3s): cellular functions and disease implications. *The EMBO Journal*, *32*, 2307–2320.

Gingerich, D. J., Gagne, J. M., Salter, D. W., Hellmann, H., Estelle, M., Ma, L. & Vierstra, R. D. (2005). Cullins 3a and 3b assemble with members of the broad complex/tramtrack/bric-a-brac (BTB) protein family to form essential ubiquitin–protein ligases (E3s) in Arabidopsis. *Journal of Biological Chemistry*, *280*, 18810–18821.

Gingerich, D. J., Hanada, K., Shiu, S., & Vierstra, R. D. (2007). Large-Scale, Lineage-Specific Expansion of a Bric-a-Brac / Tramtrack / Broad Complex Ubiquitin-Ligase Gene Family in Rice. *The Plant Cell*, *1*, 1–21.

Graaff, E. Van Der, Laux, T., & Rensing, S. A. (2009). The WUS homeobox-containing (WOX) protein family. *Genome Biology*, *10*, 1–9.

Gutiérrez-Mora, A., González-Gutiérrez, A. G., Rodríguez-Garay, B., Ascencio-Cabral, A. & Li-Wei, L. (2012). Plant somatic embryogenesis: some useful considerations. In: *Embryogenesis*. Sato, K. I. (ed.) Rijeka: InTech, pp. 229–241.

Hove, C. A., Lu, K., & Weijers, D. (2015). Building a plant: cell fate specification in the early Arabidopsis embryo. *Development*, *142*, 420–430.

Jiménez, V. M. (2001). Regulation of in vitro somatic embryogenesis with emphasis on the role of endogenous hormones. *Revista Brasileira de Fisiologia Vegetal*, *13*, 196-223.

Juranić, M., & Dresselhaus, T. (2014). Phylogenetic analysis of the expansion of the MATH-BTB gene family in the grasses. *Plant Signaling & Behavior*, *9*, doi:10.4161/psb.28242

Juranić, M., Srilunchang, K., Graciele Krohn, N., Leljak-Levanić, D., Sprunck, S. & Dresselhaus., T. (2012). Germline-Specific MATH-BTB Substrate Adaptor MAB1 Regulates Spindle Length and Nuclei Identity in Maize. *The Plant Cell*, *24*, 4974–4991.

Koonin, E. V., Senkevich, T. G. & Chernos, V. I. (1992). A family of DNA virus genes that consists of fused portions of unrelated cellular genes. *Trends in Biochemical Sciences*, *17*, 213-214.

Latham, J., Wilson, A. K. & Steinbrecher, A. K. (2006). The Mutational Consequences of Plant Transformation. *Journal of Biomedicine and Biotechnology*, doi:10.1155/JBB/2006/25376

Lechner, E., Leonhardt, N., Eisler, H., Parmentier, Y., Alioua, M., Jacquet, H., Leung, J. & Genschik, P. (2011). MATH / BTB CRL3 Receptors Target the Homeodomain-Leucine Zipper ATHB6 to Modulate Abscisic Acid Signaling. *Developmental Cell*, *21*, 1116–1128.

Leljak-Levanić, D., Juranić, M., & Sprunck, S. (2013). De novo zygotic transcription in wheat (Triticum aestivum L.) includes genes encoding small putative secreted peptides and a protein involved in proteasomal degradation. *Plant Reproduction*, *26*, 267-285.

Lotan, T., Ohto, M., Yee, K. M., West, M. A. L., Lo, R., Kwong, R. W., Yamagishi, K., Fischer, R. L., Goldberg, R. B. & Harada, J. J. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*, *93*, 1195–1205.

Luke-Glaser, S., Pintard, L., Lu, C., Mains, P. E., Peter, M., & Tn, A. (2005). The BTB Protein MEL-26 Promotes Cytokinesis in C. elegans by a CUL-3-Independent Mechanism. *Current Biology*, *15*, 1605–1615.

Mani, R-S. (2014). The emerging role of speckle-type POZ protein (SPOP) in cancer development. *Drug Discovery Today*, *19*, 1498–1502.

Mansfield, S. G. & Briarty, L. G. (1991). Early embryogenesis in Arabidopsis thaliana. II. The developing embryo. *Canadian Journal of Botany*, *69*, 461-476.

Marín, I. (2015). Origin and Diversification of Meprin Proteases, *PLoS One*, *10*, doi:10.1371/journal.pone.0135924

Miller, J. H. (1972). Experiments in molecular genetics. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory. isbn:0879691069.

Murashige, T. & Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, *15*, 473-479.

Pagnussat, G. C., Alandete-Saez, M. & Bowman, J. (2009). Auxin-Dependent Patterning and Gamete Specification in the Arabidopsis Female Gametophyte. *Science*, *324*, 1684-1689.

Patton, E. E., Willems, A. R. & Tyers, M. (1998). Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. *Trends in Genetics*, *14*, 236–243.

Peters, J-M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Reviews Molecular Cell Biology*, *7*, 644-656.

Pintard, L., Willems, A., & Peter, M. (2004). Cullin-based ubiquitin ligases: CUL3-BTB complexes join the family. *The EMBO Journal*, *23*, 1681–1687.

Pintard, L., Willis, J. H., Willems, A., Johnson, J. L., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers, M., Bowerman, B. & Peter, M. (2003). The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature*, *425*, 311–316.

Portillo, L., Santacruz-Ruvalcaba, F., Gutiérrez-Mora, A. & Rodríguez-Garay, B. (2007). Somatic embryogenesis in Agave tequilana Weber cultivar azul. *In Vitro Cellular and Developmental Biology – Plant*, *43*, 569-575.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406-425.

Sasaki, T., Lorković, Z. J., Liang, S.- C., Matzke, A. J. M., & Matzke, M. (2014). The ability to form homodimers is essential for RDM1 to function in RNA-directed DNA methylation. *PloS One*, *9*, http://doi.org/10.1371/journal.pone.0088190

Sasidharan, R. & Gerstein, M. (2008). Genomics: protein fossils live on as RNA. *Nature*, *453*, 729-731.

Smet, I. De, & Beeckman, T. (2011). Asymmetric cell division in land plants and algae: the driving force for differentiation. *Nature Publishing Group*, *12*, 177–188.

Stogios, P. J., Downs, G. S., Jauhal, J. J. S., Nandra, S. K., & Privé, G. G. (2005). Sequence and structural analysis of BTB domain proteins. *Genome Biology*, *6*, 1–18.

Vondráková, Z, Krajňáková, J., Fischerová, L., Vágner, M. & Eliášová, M. (2016). Physiology and role of plant growth regulators in somatic embryogenesis. In: Vegetative Propagation of Forest Trees. Park, Y-S., Bonga, J. M., Moon, H-K. (eds.) National Institute of Forest Science (NIFoS). Seoul, pp. 123-169.

Weber H., Bernhardt A., Dieterle, M., Hano, P., Mutlu, A., Estelle, M., Genschik, P. & Hellmann, H. (2005). Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZ-MATH protein family. *Plant Physiology*, *137*, 83–93.

Weber, H. & Hellmann, H. (2009). Arabidopsis thaliana BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *The FEBS Journal*, *276*, 6624–6635.

Willems, A. R., Schwab, M. & Tyers, M. (2004). A hitchhiker' s guide to the cullin ubiquitin ligases: SCF and its kin. *Biochimica et Biophysica acta*, *1695*, 133–170.

Willemsen, V. & Scheres, B. (2004). Mechanisms of pattern formation in plant embryogenesis. *Annual Review of Genetics, 38*, 587-614.

Xu, L., Wei, Y., Reboul, J., Vaglio, P., Shin, T.H., Vidal, M., Elledge, S.J. & Harper, J.W. (2003). BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3. *Nature, 425*, 316–321.

Yoshida, S., Barbier de Reuille, P., Lane, B., Bassel, G. W., Prusinkiewicz, P., Smith, R. S. & Weijers, D. (2014). Genetic control of plant development by overriding a geometric division rule. *Developmental Cell, 29*, 75-87.

Zapata, J. M., Martínez-garcía, V. & Lefebvre, S. (2007). Phylogeny of the TRAF/MATH domain. In: TNF receptor associated factors (TRAFs). Wu, H. (ed.), New York, Springer, pp. 1-24.

Zavatierri, M. A. (2010). Induction of somatic embryogenesis as an example of stress-related plant reactions. *Electronic Journal of Biotechnology*, *13*, DOI: 10.2225/vol13-issue1-fulltext-4

Zollman, S., Godt, D., Priv´e, G. G., Couderc, J. & Laski, F. A. (1994). The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 10717–10721.

Zhuang, M., Calabrese, M. F., Liu, J., Waddell, M. B., Nourse, A., Hammel, M., Miller, D. J., Walden, H., Duda, D. M., Seyedin, S. N., Hoggard, T., Harper, W., White, K. P. & Schumann, B. A. (2010). Structures of SPOP-Substrate Complexes: Insights into Molecular Architectures of BTB-Cul3 Ubiquitin Ligases. *Molecular Cell*, *36*, 39–50.

7 APPENDIX

Supplemental Table 1. Updated list of *MATH-BTB* genes identified in the maize genome. Thirty-one gene has been previously described (Juranić et al., 2012) and denominated *ZmMAB1-31* (*Z. mays* MATH-BTB). Two additional *MATH-BTB* genes, *ZmMAB15like* and *ZmMAB17like*, have been included in the list. Listed in the table are each gene's sequence identifier from www.maizesequence.org or NCBI database, number of splicing variants, length of protein product as the number of amino acids (aa) and number of coding exons. For genes with more than one splicing variant, a number was added after decimal point to denote the variant used for phylogenetic analysis. Numbers in parenthesis designate the number of amino acids and exons in all splicing variants of a given gene. Highlighted in red are *ZmMAB* genes of the core group.

gene product name	sequence identifier (Locus ID)	number of splicing variants per gene	translation length of transcript (aa)	number of exons
ZmMAB1	AC195147.3_FG001	1	347	1
ZmMAB2	GRMZM2G404188	1	352	1
ZmMAB3	GRMZM2G337139	1	328	1
ZmMAB4	GRMZM2G081441	1	345	1
ZmMAB5	GRMZM2G372171	1	344	1
ZmMAB6	GRMZM2G125162	1	399	3
ZmMAB7.1	GRMZM2G110531_P01	2	371 (371)	1 (1)
ZmMAB8	GRMZM2G418031	1	372	1
ZmMAB9	GRMZM2G574887	1	358	1
ZmMAB10.1	GRMZM2G154437_P01	2	369 (306)	1 (2)
ZmMAB11.2	GRMZM2G077428_P02	2	371 (205)	1 (1)
ZmMAB12.1	GRMZM2G181276_P01	2	453 (140)	1 (2)
ZmMAB13	GRMZM2G027688	1	373	1
ZmMAB14.1	GRMZM2G052985_P01	3	399 (143, 60)	4 (2, 1)
ZmMAB15	GRMZM2G148213	1	422	4
ZmMAB16	GRMZM2G172210	1	426	4
ZmMAB17	GRMZM2G166049	1	428	4
ZmMAB18.1	GRMZM2G060765_P01	5	427 (428, 375, 320, 243)	5 (4, 4, 2, 3)
ZmMAB19.1	GRMZM2G074323_P01	3	432 (292, 59)	4 (2, 1)
ZmMAB20.1	GRMZM2G009724_P01	2	364 (364)	1 (1)
ZmMAB21	GRMZM2G109738	1	350	1

Table continued on page ii.

ZmMAB22.1	GRMZM2G046238_P01	2	368 (155)	1 (1)
ZmMAB23	GRMZM2G143782	1	379	2
ZmMAB24	GRMZM2G103251	1	362	2
ZmMAB25	GRMZM2G088086	1	356	2
ZmMAB26	GRMZM2G161610	1	274	1
ZmMAB27	GRMZM2G161569	1	359	1
ZmMAB28	GRMZM2G041963	1	353	1
ZmMAB29	GRMZM2G077951	1	786	1
ZmMAB30	GRMZM2G319215	1	357	1
ZmMAB31	/	1	367	1
ZmMAB15like	ACG38997.1 (NCBI)	/	422	/
ZmMAB17like	ACG37150.1 (NCBI)	/	428	/

Supplemental Table 2. List of *MATH-BTB* genes identified in the rice genome. Sixty-seven genes have been previously described (Gingerich et al., 2007) and denominated *OsMBTB* (*O. sativa* MATH-BTB). Listed in the table are each gene's sequence identifier from the TIGR database, number of splicing variants, length of protein product as the number of amino acids (aa) and number of coding exons. For genes with more than one splicing variant, a number was added after decimal point to denote the variant used for phylogenetic analysis. Numbers in parenthesis designate the number of amino acids and exons in all splicing variants of a given gene. Highlighted in red are the *OsMBTB* genes of the core group.

gene product name	sequence identifier (Locus ID)	number of splicing variants per gene	translation length of transcript (aa)	number of exons
OsMBTB1	LOC_Os02g20590	1	353	1
OsMBTB2	LOC_Os02g20720	1	390	1
OsMBTB4	LOC_Os04g35310	1	368	2
OsMBTB5	LOC_Os04g35370	1	373	2
OsMBTB6	LOC_Os04g53400	1	375	1
OsMBTB7	LOC_Os04g53410	1	366	1
OsMBTB8	LOC_Os04g53430	1	373	1
OsMBTB9	LOC_Os04g56460	1	353	1
OsMBTB10.1	LOC_Os06g14060.1	2	352 (352)	2 (2)
OsMBTB11	LOC_Os06g45730	1	364	2
OsMBTB15	LOC_Os08g03470	1	370	2
OsMBTB16	LOC_Os08g03480	1	300	1
OsMBTB17	LOC_Os08g03500	1	370	1
OsMBTB18	LOC_Os08g03510	1	377	1

Table continued on page iii.

OsMBTB19	LOC_Os08g03530	1	382	1
OsMBTB20	LOC_Os08g13000	1	365	1
OsMBTB21	LOC_Os08g13030	1	363	1
OsMBTB22	LOC_Os08g13060	1	386	1
OsMBTB23	LOC_Os08g13070	1	355	2
OsMBTB24	LOC_Os08g13180	1	384	1
OsMBTB25	LOC_Os08g25240	1	419	1
OsMBTB26	LOC_Os08g31420	1	392	1
OsMBTB27	LOC_Os08g31430	1	401	1
OsATBPM2	LOC_Os10g29020	1	312	2
OsMBTB29	LOC_Os08g41120	1	341	1
OsMBTB30	LOC_Os08g41150	1	359	1
OsMBTB31	LOC_Os08g41180	1	321	1
OsMBTB32	LOC_Os08g41190	1	361	1
OsMBTB33	LOC_Os10g28760	1	382	1
OsMBTB34	LOC_Os10g28770	1	372	1
OsMBTB36	LOC_Os10g28780	1	383	2
OsMBTB36	LOC_Os10g28790	1	373	1
OsMBTB37	LOC_Os10g28810	1	395	1
OsMBTB38	LOC_Os10g28820	1	374	1
OsMBTB39	LOC_Os10g28860	1	372	1
OsMBTB40	LOC_Os10g28870	1	385	1
OsMBTB42	LOC_Os10g28990	1	386	1
OsMBTB43.1	LOC_Os10g29050.1	2	363 (329)	2 (3)
OsMBTB44	LOC_Os10g29100	1	369	1
OsMBTB45	LOC_Os10g29110	1	409	1
OsMBTB46	LOC_Os10g29150	1	390	1
OsMBTB47.1	LOC_Os10g29180.1	2	375 (369)	2 (1)
OsMBTB48	LOC_Os10g29220	1	356	1
OsMBTB49	LOC_Os10g29230	1	369	1
OsMBTB50	LOC_Os10g29290	1	363	1
OsMBTB51	LOC_Os10g29310	1	363	1
OsMBTB52	LOC_Os10g29380	1	370	1
OsMBTB53	LOC_Os10g29410	1	401	1
OsMBTB55	LOC_Os10g29740	1	370	1
OsMBTB56	LOC_Os10g29750	1	366	1

Table continued on page iv.

OsMBTB57	LOC_Os10g29790	1	395	1
OsMBTB58	LOC_Os10g29810	1	397	1
OsMBTB59	LOC_Os10g29840	1	359	1
OsMBTB60	LOC_Os10g29850	1	355	4
OsMBTB61	LOC_Os11g24550	1	374	1
OsMBTB62	LOC_Os11g40220	1	342	1
OsMBTB63	LOC_Os11g40490	1	371	2
OsMBTB64	LOC_Os11g40680	1	370	1
OsMBTB65	LOC_Os11g41300	1	358	1
OsMBTB66	LOC_Os11g41310	1	380	1
OsMBTB67	LOC_Os11g41350	1	392	1
OsMBTB68	LOC_Os11g45560	1	370	3
LOC_Os10g29501	LOC_Os10g29495.1	1	718	12
OsMBTB3.1	LOC_Os03g57854.1	2	431 (378)	4 (4)
OsMBTB13.1	LOC_Os07g07270.1	2	424 (371)	4 (4)
OsMBTB12	LOC_Os07g01140	1	395	4
OsMBTB14	LOC_Os07g46160	1	434	4

Supplemental Table 3. List of *MATH-BTB* genes identified in the *Arabidopsis thaliana* genome, denominated *AtBPM* (*A. thaliana* BTB/POZ-MATH). Listed in the table are each gene's sequence identifier from the TAIR database, number of splicing variants, length of protein product as the number of amino acids (aa) and number of coding exons. For genes with more than one splicing variant, a number was added after decimal point to denote the variant used for phylogenetic analysis. Numbers in parenthesis designate the number of amino acids and exons in all splicing variants of a given gene. Highlighted in red are the *AtBPM* genes of the core group.

gene product name	sequence identifier (Locus ID)	number of splicing variants per gene	translation length of transcript (aa)	number of exons
AtBPM1.1	AT5G19000.1	2	407 (402)	4 (5)
AtBPM2.1	AT3G06190.1	2	406 (295)	4 (4)
AtBPM3.1	AT2G39760.1	2	408 (343)	4 (3)
AtBPM4	AT3G03740.1	1	465	4
AtBPM5	AT5G21010.1	1	410	4
AtBPM6	AT3G43700.1	1	415	4

CURRICULUM VITAE

Andreja Škiljaica

Personal information

Address: Fortunata Pintarića 5, 48 000, Koprivnica
Email: andreja.skiljaica@gmail.com
Date of birth: October, 24th 1992.

Education

2014 – 2016 Graduate program of Molecular Biology
University of Zagreb, Faculty of Science, Division of Biology
2011 – 2014 Undergraduate program of Molecular Biology
University of Zagreb, Faculty of Science, Division of Biology
Title**: Bachelor in Molecular Biology,** univ. bac. mol. biol.

Training

Laboratory practice in the Laboratory for plant tissue culture, at University of Zagreb, Faculty of Science, Division of Biology

Lab Demonstrator for the following courses: Mechanisms of Plant Development, Plant Tissue Culture and Introduction to Genetic Engineering

Skills

Languages: Croatian (mother tongue), English
Computer: MS Windows, MS Office, Photoshop
Hobbies: Creative writing, photography