University of Zagreb
Faculty of Science
Department of Biology

Gaj Keresteš

Identification and characterization of *BPM2* splice variants in *Arabidopsis thaliana*

Master thesis

Zagreb, 2024.

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Gaj Keresteš

Identifikacija i karakterizacija varijanti alternativnog prekrajanja uročnjakova gena *BPM2*

Diplomski rad

Zagreb, 2024.

This thesis was created in the Plant tissue culture laboratory at the Division of Molecular Biology, Faculty of Science, University of Zagreb, under the supervision of Prof. Dunja Leljak-Levanić, PhD, and co-supervision of assistant lecturer Paula Štancl, mag. biol. mol. This thesis was created within the Croatian Science Foundation scientific research project "Alternative splicing of BPM2 as a mechanism of achieving functional difersity of MATH-BTB proteins in *Arabidopsis thaliana*" (IP-2022-10-7874), led by Prof. Nataša Bauer, PhD. The thesis was submitted for grading to the Department of Biology at the Faculty of Science, University of Zagreb, with the aim of obtaining the Master's degree in molecular biology (mag. biol. mol.).

Acknowledgments

I want to sincerely thank my mentor, Professor Dunja Leljak-Levanić, for the immense amount of support and help she gave me in creating this thesis, and for always being very communicative, open and understanding. I also greatly thank my co-mentor Paula Štancl, for navigating me through the bioinformatic part of this thesis and giving me the much needed encouragement every step of the way.

A big thank you to Mateja Jagić for being my immediate mentor in the lab, showing me how to better perform my experiments and helping me find solutions to the problems I encountered. Thank you to Professor Nataša Bauer for the advice and encouragement when problems with experiments made me doubt myself. Thank you also to Marta Frlin, for all of our conversations about our work and for the mutual reassurance.

Thank you to my colleagues and friends Lana, Sara, Dora, Nikola, Alan, Dona and Kristina for sharing the experience of this study programme and all of the various emotions it put us through, you made the last five years really special.

I am endlessly grateful to my family for all the support I received throughout my whole education and broader life, and for trusting me in creating my own path.

Thank you to my best friend Karlo, for being there for me through thick and thin and for being understanding of the times when we couldn't see each other as much as we would have liked.

Lastly, thank you to Antoni, for seeing me in every state and not looking away.

Thank you all for treating me like a complete person, it meant more than you might know.

BASIC DOCUMENTATION CARD

University of Zagreb
Faculty of Science
Department of Biology   Master thesis

Identification and characterization of *BPM2* splice variants in *Arabidopsis thaliana*

Gaj Keresteš

Horvatovac 102a, 10000 Zagreb, Croatia

MATH-BTB proteins mediate turnover of specific proteins by acting as substrate adaptors of E3 ubiquitin ligases. The *MATH-BTB* gene family has undergone a significant expansion in grasses, with genomes of different grass species having dozens of *MATH-BTB* genes, some of which have roles in fundamental developmental processes. In contrast, the genome of *Arabidopsis thaliana* only contains six *MATH-BTB* genes: *BPM1-6*. In this thesis, alternative splicing was explored as a potential mechanism for *A. thaliana* to achieve the number and diversity of MATH-BTB proteins similar to grasses. A bioinformatic analysis of the newest *A. thaliana* transcriptome, AtRTD3, revealed 56 transcript isoforms encoded by *BPM* genes, 16 of which are encoded by *BPM2*. While proteins encoded by the 16 *BPM2* transcripts differed in domain composition, they lacked the diversity in MATH domains that's present in grass-specific MATH-BTB proteins. Expressions of splice variants *BPM2.3*, *BPM2.8*, *BPM2.9* and *BPM2.15* were analyzed in some vegetative, reproductive and embryonic tissues with standard and quantitative PCR. Splice variants *BPM2.3* and *BPM2.15* had significant differences in expression levels across different tissues, indicating a possible role in developmental transitions. Theoretically unexpected amplicons represent potentially novel splice variants.

Keywords: MATH-BTB, AtRTD3, alternative splicing, plant development
(56 pages, 24 figures, 7 tables, 75 references, original in: English)
Thesis is deposited in Central Biological Library.

Mentor: Prof. Dunja Leljak-Levanić, PhD
Co-mentor: Paula Štancl, mag. biol. mol.

Reviewers:
Prof. Dunja Leljak-Levanić, PhD
Assoc. Prof. Rosa Karlić, PhD
Prof. Nataša Bauer, PhD

Thesis accepted: 04. srpnja 2024.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu
Prirodoslovno-matematički fakultet
Biološki odsjek

Diplomski rad

Identifikacija i karakterizacija varijanti alternativnog prekrajanja uročnjakova gena *BPM2*

Gaj Keresteš

Horvatovac 102a, 10000 Zagreb, Hrvatska

Proteini MATH-BTB sudjeluju u prometu specifičnih proteina djelujući kao adapteri supstrata E3 ubikvitinskih ligaza. Porodica gena *MATH-BTB* je prošla značajnu ekspanziju u travama, pri čemu genomi različitih vrsta trava imaju desetke gena *MATH-BTB*, od kojih neki imaju uloge u temeljnim razvojnim procesima. Nasuprot tome, genom uročnjaka *Arabidopsis thaliana* ima samo šest *MATH-BTB* gena: *BPM1-6*. U ovom radu, alternativno prekrajanje transkripata je istraženo kao potencijalni mehanizam kojim *A. thaliana* postiže broj i raznolikost proteina MATH-BTB sličan travama. Bioinformatička analiza najnovijeg transkriptoma *A. thaliana*, AtRTD3, otkrila je 56 izoformi transkripta kodiranih genima *BPM*, od kojih je 16 kodirano genom *BPM2*. Iako su se proteini kodirani sa 16 transkripata gena *BPM2* razlikovali u sadržaju domena, nedostajala im je raznolikost u MATH domenama koja je prisutna u MATH-BTB proteinima specifičnima za trave. Ekspresije varijanata prekrajanja *BPM2.3, BPM2.8, BPM2.9* i *BPM2.15* analizirane su u nekim vegetativnim, reproduktivnim i embrionalnim tkivima standardnim i kvantitativnim reakcijama PCR-a. Varijante *BPM2.3* i *BPM2.15* imale su značajne razlike u razinama ekspresije u različitim tkivima, što ukazuje na moguću ulogu u razvojnim prijelazima. Standardni PCR je proizveo teoretski neočekivane amplikone koji predstavljaju potencijalno nove varijante prekrajanja.

Ključne riječi: MATH-BTB, AtRTD3, alternativno prekrajanje transkripata, biljni razvitak
(56 stranica, 24 slika, 7 tablica, 75 literaturnih navoda, jezik izvornika: engleski)
Rad je pohranjen u Središnjoj biološkoj knjižnici

Mentor: Prof. dr. sc. Dunja Leljak-Levanić
Komentor: Paula Štancl, mag. biol. mol.

Ocjenitelji:
Prof. dr. sc. Dunja Leljak-Levanić
Izv. prof. dr. sc. Rosa Karlić
Prof. dr. sc. Nataša Bauer

Rad prihvaćen: July 4th, 2024

Contents

Abbrevations

aa – amino-acids

Araport11 – Arabidopsis Information Portal complete reannotation of the *Arabidopsis thaliana* reference genome

AS – alternative splicing

At-Iso – *Arabidopsis thaliana* PacBio Iso-seq transcriptome assembly

AtRTD2 – *Arabidopsis thaliana* Reference Transcript Dataset 2

AtRTD3 – *Arabidopsis thaliana* Reference Transcript Dataset 3

BACK – BTB and C-terminal KELCH protein domain

bHLH – basic helix-loop-helix

bp – base pairs

BPM, AtBPM – *Arabidopsis thaliana* BTB/POZ-MATH proteins

BTB/POZ – Broad-complex, Tramtrack, and Bric-à-brac/Pox virus and Zinc finger protein domain

CD – conserved domain database

cDNA – complementary DNA

CRL3 – CUL3 based RING E3 ligases

Ct – threshold cycles

CUL3 – CULLIN 3

dNTP – deoxyribonucleotide triphosphate

ET – elongation time

gDNA – genomic DNA

HsSPOP – *Homo sapiens* SPOP

Iso-Seq – Isoform sequencing, RNA sequencing of contiguous, full-length transcripts

mapq – mapping quality

MATH – meprin and TRAF homology protein domain

mRNA – messenger RNA

msa – multiple sequence aligment

NLS – nuclear localization signal

NMD – nonsense-mediated decay

NTC – no-template control

ORF – open reading frame

PCR – polymerase chain reaction

Pfam – protein family database

PTC – premature termination codon

PUX7 – plant UBX domain-containing protein 7

qPCR – quantitative polymerase chain reaction

R2R3 MYB – R2 and R3 domain-containing myeloblastosis transcription factors

RING – REALLY INTERESTING NEW GENE

RPKM – reads per kilobase per million mapped reads

RT – reverse transcription

SPOP – Speckle-type POZ protein

Ta – annealing temperature

TAIR – The Arabidopsis Information Resource, a database of genetic and molecular biology data for *Arabidopsis thaliana*

TaMAB – *Triticum aestivum* MATH-BTB proteins

TES – transcription end site

Tm – melting temperature

TPM – transcripts per kilobase million mapped reads

TSS – transcription start site

UTR – untranslated region

ZmMAB – *Zea mays* MATH-BTB proteins

1. Introduction

The MATH-BTB protein family is a large group of proteins widely distributed among eukaryotes. The proteins contains the N-terminal MATH domain and the C-terminal domain of BTB/POZ with a zinc finger motif (Weber et al., 2005) although there are cases where the positions of these two domains are exchanged or they appear as multiple copies (Zapata et al., 2007). These two domains are found in many proteins of the MATH and BTB protein superfamilies, respectively. Each domain is capable of associating with a diverse set of other proteins. The studied mechanism of function of MATH-BTB proteins is their participation in proteasomal degradation where they act as target specificity module of E3 ubiquitin ligase complexes — CULLIN 3 (CUL3)-based REALLY INTERESTING NEW GENE (RING) E3 ligases or CRL3 (Gingerich et al., 2007; Pintard et al., 2003). In this mechanism, the BTB domain binds a CUL3 scaffold protein, while the N-terminal MATH domain targets a highly diverse collection of substrate proteins for proteasomal degradation, mediating a number of diverse developmental processes in plants. In a number of MATH-BTB proteins there is also a (BTB and C-terminal KELCH)-AtBPM-like or BACK domain (Jagić et al., 2022) mainly considered to play a role in orientation of targeted substrates (Stogios et al., 2005). At the very C-terminal end of some Arabidopsis MATH-BTBs, there is a nuclear localization signal sufficient to drive a protein into the nucleus and nucleolus (Leljak Levanić et al.2012).

1.1. Protein family MATH-BTB

Phylogenetic analysis of MATH-BTB families in different plant species points to a significant expansion in the number of MATH-BTB encoding genes in grass species. While Arabidopsis genome contains only six MATH-BTB genes (designated as *BPM1-6*; Weber et al., 2005; Weber and Hellmann, 2009), maize genome contains 31 genes (Juranić et al., 2012), wheat 46 genes (Bauer et al., 2019) and rice MATH-BTB gene family has 74 members (Gingerich et al., 2007). Despite that, only a few plant MATH-BTB proteins have been characterized functionally. These are Arabidopsis BPM proteins (Jagic et al., 2022 and references cited therein), maize ZmMAB1 (Juranić et al., 2012) and wheat TaMAB2 (Bauer et al., 2019). These analysis shows that MATH-BTB proteins participate in diverse physiological mechanisms. ZmMAB1 participates in regulation of cell division by targeting a cytoskeletal protein katanin (Juranić et al., 2012), TaMAB2 potentially targets eukaryotic translation initiation factors eIF3 and eIF4 (Bauer et al., 2019), while Arabidopsis BPM proteins modulate the proteasomal turnover of transcription factors including Apetala2/ethylene responsive, class I homeobox-leucine zipper, R2R3 MYB and bHLH family, as well as protein phosphatases type 2C (cited in Jagić et al., 2022). Recently, the BPM1 protein has been shown to interact with components of RNA-directed DNA methylation machinery (Jagić et al., 2022). Some of these functions are mediated by a CUL3-dependent and some by a CUL3-independent mechanism, which is determined by the roles of individual protein domains in a specific process. Therefore, the function of each domain needs to be studied separately, but also as part of the function of the whole protein.

The MATH domain is a conserved domain positioned closer to the N-terminal end of the BPM protein (Fig. 1). It consists of approximately 180 amino acids. The name MATH comes from the homology with the C-terminal end of the proteins meprin A and B. Meprins are extracellular metalloproteases that have a high homology of the C-terminal end with the TRAF-C domain of TRAF proteins. The TRAF-C domain consists of a high proportion of ß-pleated sheets and it is often called MATH due to its similar composition (Marín, 2015). As part of the MATH-BTB proteins, the MATH domain participates in the recognition and positioning of the substrate for ubiqiutination (Pintard et al., 2003). In addition to the fact that this domain is present in plants, it also occurs in proteins of protozoa and unicellular fungi as well as in iridoviruses, but it has not been detected in any prokaryote (Zapata et al., 2007).

The BTB/POZ domain is a conserved domain located closer to the C-terminal end of the MATH-BTB proteins. It was identified for the first time in *Drosophila melanogaster* as a "bric a brac, tramtrack and broad" motif of transcription factors, as well as in numerous poxvirus proteins from where it gets the name BTB/POZ (eng. Pox virus and Zinc finger) (Zollman et al., 1994). Proteins that have a BTB/POZ domain are divided into several families, the two most important being the BTB/POZ and MATH-BTB/POZ families. The BTB/POZ family in general has a role as a substrate-specific adaptor of the CUL3 E3 complex (Pintard et al., 2003). In addition to a role in the ubiquitin-dependent protein degradation pathway, the BTB/POZ domain also acts on transcriptional repression (Ziegelbauer et al., 2001), cytoskeleton regulation (Ziegelbauer et al., 2001), tetramerization, and closing of ion channels (Minor et al., 2000). Generally, BTB/POZ proteins are classified into two groups, namely those that participate in protein-protein interactions and those that regulate transcription by binding to the DNA molecule. Evolutionarily, the

BTB/POZ domain does not occur in archaebacteria or bacteria except in the Candidatus protochlamidya species, which is considered to be the origin of this domain, which arose after the separation of eukaryotes (Horn et al., 2004).

Based on complementarity with the human SPOP protein, BACK (BTB AND C-TERMINAL KELCH)-AtBPM-like domain lies after BTB domain (Fig. 1). For example, in BPM1, an Arabidopsis MATH-BTB protein, the BTB and BACK domains are separated by one glycine at position 311. In BPM2 there is no space between the BTB and BACK domains, with a cysteine at position 308 even being attributed to both domains (Fig. 1). Although the BACK domain is frequently present in plant MATH-BTB proteins it has not yet been functionally characterized (Jagić et al., 2022). According to the sequence alignment of the human protein SPOP and BPM1, the sequence similarity is only 27%, but crucial amino acids are conserved (Miškec 2019). The role of the BACK domain is coupled with the role of BTB domain and BACK itself enhances the interaction of the MATH-BTB protein with CUL3. Bioinformatics research suggests that all MATH-BTB/POZ proteins that do not have a BACK domain, are not CUL3 adaptors (Zhuang et al., 2009).

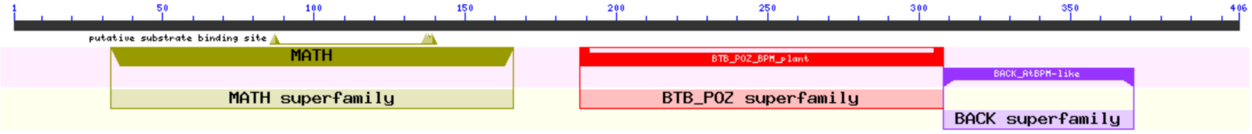


Figure 1. Representation of the conserved domains of the BPM2 protein. The MATH domain is shown in green, the BTB domain in red, and the BACK domain in purple. Source: https://www.ncbi.nlm.nih.gov/.

On the basis of the above, it is clear that the develpmental function of MATH-BTB proteins is mediated by the structure, presence and completeness of domains, but also by presence/absence of other functional sequences such as NLS within proteins (Leljak Levanić et al., 2012) and these characteristics together determine the specificity of protein functions.

1.2. Phylogenetic analysis of the MATH-BTB family

Phylogenetic analysis of grasses MATH-BTB/POZ proteins shows a branching of clades, showing two dominant groups, a core one common to most of the plant species including grasses and Arabidopsis, and a larger extended and grass-specific group (Fig. 2; Juranić and Dresselhaus, 2014). The functional need for a grass-specific expansion of the MATH-BTB family is not at all understood. *MATH-BTB* genes of the core clade are mainly conserved (60 to 84% amino acid identity with almost no gaps in a sequence alignment) and ubiquitously expressed (Weber and Hellmann, 2009; Lechner et al., 2011) indicating that these genes might be associated with crucial developmental processes. In contrast, the grass-specific and larger clade shows more extended gaps and its members are less homologous to each other (Juranić and Dresselhaus, 2014). Two functionally characterized members of the expanded clade (Juranić et al., 2012, Bauer et al., 2019) show that these genes also control fundamental developmental mechanisms, however their expressions are tightly controlled during development and thus involved in more specific developmental processes. Since expanded clade members have not been identified in Arabidopsis and other dicots, it would be interesting to find out how Arabidopsis and other dicots compensate for the small number of *MATH-BTB* genes and increase the variety of their functions. One possible answer is alternative splicing, as it was shown that a smaller number of MATH-BTB genes correlates with a greater number of splice variants. For instance, the human MATH-BTB gene SPOP was reported to generate 23 functional splice variants likely resulting in proteins with different activities and targets (Juranić and Dresselhaus, 2014).

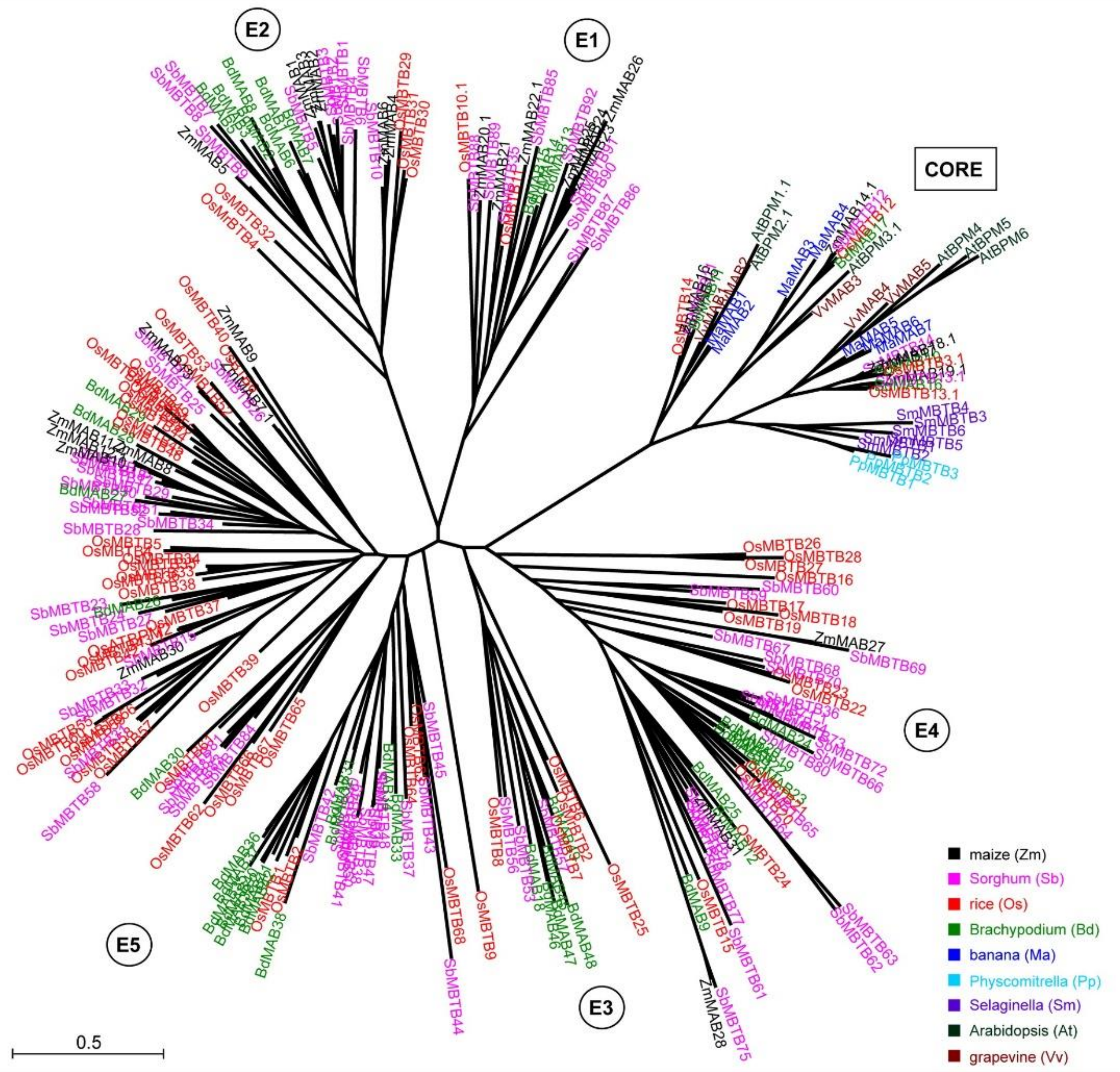


Figure 2. Phylogentic tree of genes encoding MATH-BTB proteins from 9 representative plant species. Analysis is done with Seaview v.4.3.4. Genes are color-coded by species (legend). Core clade MATH-BTB genes form a separate clade. Grass-specific expanded clade forms five subclades (E1 to E5). The number after decimal point for designated gene presents a splicing variant used for phylogenetic analysis. The bar of 0.5 is a branch length, which represents nucleotide substitutions per site. Source: Juranić and Dresselhaus (2014).

1.3. Alternative splicing

In addition to the transcriptional control of gene expression, other less studied regulatory mechanisms including alternative splicing (AS) are essential. Alternative splicing is a mechanism that increases the complexity of gene expression by impacting production of mature mRNA from the primary mRNA transcript (precursor mRNA). Alternative splicing is a key mechanism for expanding proteome diversity, providing a greater number of proteins with more diverse functions being transcribed from a limited number of genes. The four major types of AS include exon skipping, intron retention, and alternative splice donor and acceptor choices, all of which have been observed in all eukaryots (Grau-Bové et al., 2018). Additionally, by disrupting the main open reading frame (ORF) of a gene, AS can effectively lead to downregulation of its expression, either by creating a truncated protein (Lewis et al., 2003) or by producing a mature mRNA that is degraded in the nonsense-mediated decay (NMD) pathway. NMD is a cytoplasmic pathway that is triggered by certain transcript features, most commonly unusually long 3'-UTRs and premature termination codons (PTC) with downstream splice junctions. There are also a number of transcripts with NMD features that aren't degraded by NMD. Namely, in plants, transcripts with NMD features caused by intron retention tend to avoid degradation by remaining within the nucleus (Kalyna et al., 2012; Göhring et al., 2014).

AS in general is a mechanisms of differential processing of introns and exons in pre-mRNAs to produce multiple mature transcript isoforms from one gene and is the most important contributor to transcriptome diversification in both plants and animals (Irimia and Blencowe, 2012; Reddy et al., 2013; Staiger and Brown, 2013). Regulation of alternative splicing is a complex process which demands highly effective synchronisation of transcription and splicing (Wang et al., 2015). AS plays an important role in cellular differentiation and organism development. In mammals, the biological importance of AS is relatively well understood. A large number of human diseases are caused by a dysregulation of AS and production of wrong gene splice variants (Scotti et al., 2016). Significant activation of AS is connected with reproductive development in mice. Zhang et al. (2024) showed that the accurate expression of appropriate isoforms is essential for transition from totipotency to pluripotency, for proper embryonic development and for transition from pattern control of development to embryogenic control via zygotic genome activation. The authors show that embryos exposed to splicing-disrupting drugs were arrested at the 2-cell stage.

During the early human embryonic development, AS activity is also shown as vigorous even in the direction of disruption of open reading frames in some transcripts, resulting in their expression termination suggesting that some genes are specifically silenced during the zygotic genome activation process in mice embryos by an AS-dependent mechanism. (Torre et al., 2023).

Understanding the functional consequences of alternative splicing events has not been sufficiently explored in plants. Most studies in plants have concentrated on the splicing events themselves while the impact of alternative splicing on plant development and physiology has often been neglected.

Comparative analysis of AS in three major animal models, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens* with *Arabidopsis thaliana* shows the highest fraction of alternatively spliced genes in humans, followed by *A. thaliana* and *D. melanogaster* with comparable AS levels, which are significantly higher than AS in *C. elegans*. In Arabidopsis, AS is found as a mechanism dominantly involved in stress responses and to a lesser extend in development or tissue determination (Martín et al., 2021).

To date, specific splicing events in *BPM* genes, or other plant *MATH-BTB* genes, have not been investigated in relation to development. However, the importance of alternative splicing is evident in Arabidopsis, where 6 *BPM* genes encode at least 16 BPM proteins (Škiljaica, 2022). According to TAIR (Berardini et al., 2015), the stress-related *BPM2* gene encodes five splice variants, which code for proteins that significantly differ in amino acid content, sequence length and domain content. The five known *BPM2* splice variants are highly similar in their untranslated regions and the 1st and 2nd exon. The significant differences appear in the 3rd and 4th exon (Škiljaica, 2022). According to sequences retrieved from the Ensembl Plants database, the *BPM2.1–2.5* splice variants encode 406, 295, 301, 298 and 355 aa-long proteins, respectively. All splice variants encode an identical MATH domain (134 aa) at the N-terminal end of the protein. However, only protein isoforms BPM2.1 and BPM2.5 contain a putative BTB domain (121 aa) and a BACK domain (64 and 35 aa, respectively). Isoforms BPM2.2, BPM2.3 and BPM2.4 contain a truncated BTB domain (59, 70 and 70 aa, respectively), recognized as part of a BTB domain in both the Pfam and CD database. In all three protein variants, the truncated BTB sequence is followed by a 49, 44 and 41 aa-long stretch, respectively, which is not recognized as part of either BTB or BACK domain in the Pfam or CD database. Moreover, preliminary results indicate the existence of additional *BPM2* gene splice variants (Pali, 2020).

1.4. Arabidopsis transcriptome dataset AtRTD3

Arabidopsis thaliana Reference Transcript Dataset 3 (AtRTD3) is currently the most comprehensive Arabidopsis transcriptome, containing twice as many transcripts as the previously best Arabidopsis transcriptome, *Arabidopsis thaliana* Reference Transcript Dataset 2 (AtRTD2; Zhang et al., 2017). The increased number of transcripts mainly comes from novel isoforms produced by AS events and transcription start site (TSS) and transcription end site (TES) variation. Using Pacific Biosciences single-molecule long-read sequencing (PacBio Iso-seq) and novel sequencing analysis methods enabled the identification of these novel isoforms with more accurately defined splice junctions, TSS and TES. AtRTD3 consists of transcripts obtained by PacBio Iso-seq (At-Iso, 77.9%) and transcripts from previous short-read assemblies, AtRTD2 (14.7%) and Araport 11 (Cheng et al., 2017; 7.4%). Transcripts were taken from previous assemblies if they had splice junctions or loci not present in At-Iso (Zhang et al., 2022).

At-Iso includes transcripts from different Arabidopsis Col-0 tissues, developmental stages, plants exposed to abiotic stress conditions, infected with different pathogens, as well as RNA degradation mutants. However, a lot of the different samples were pooled before library construction, preventing the reads from being assigned to a specific developmental stage. There was also no sample of leaves, neither from the rosette nor from the stem, of plants that were grown in control conditions (Zhang et al., 2022).

2. Research aims

The first aim of the thesis is to characterize alternative splicing variants of *BPM2* gene transcripts based on data obtained by sequencing complete mRNAs in *Arabidopsis thaliana* (AtRTD3, Zhang et al., 2022), and to identify novel isoforms that are not described in the publicly available TAIR database. By mapping long reads obtained by Zhang et al. (2022) to the AtRTD3 transcriptome, it will be determined which of the *BPM2* transcripts appear in plants exposed to temperature stress (heat or cold).

After the identification of novel splicing variants, the representation of several novel isoforms will be analyzed in different tissues of *Arabidopsis thaliana*, with a special focus on tissues that appear during developmental transitions and global reprogramming of the transcriptome with the aim to identify the presence of specific variants in particular tissues, *in vivo*.

The third aim is the evaluation whether the *BPM2* gene encodes additional novel splice variants whose presence was not observed up to now in any available dataset including the most comprehensive AtRTD3. This will be based on the results of the splicing variant expression analysis and detection of theoretically unexpected amplicons.

3. Materials and methods

3.1. Bioinformatics

3.1.1. Transcript data

I used the R statistical programming language (version 4.3.3; R Core Team, 2023) for data analysis. I imported the AtRTD3 transcriptome (Zhang et al., 2022) GTF and BED files using the package rtracklayer (version 1.64.0; Lawrence et al., 2009), while the FASTA file was imported using the readDNAStringSet function from the Biostrings package (version 2.70.1; Pagès et al., 2017). I filtered these data objects to contain only the transcripts of the BPM2 gene, specifically those labeled with the identifier "AT3G06190" from the TAIR database. Also, I imported Table S9 from Zhang et al. (2022), which contains information on transcript coding potential, characteristics and translations.

I identified transcript isoforms for all *BPM* genes in the AtRTD3 dataset by searching for their corresponding TAIR identifiers: AT5G19000 for *BPM1*, AT3G06190 for *BPM2*, AT2G39760 for *BPM3*, AT3G03740 for *BPM4*, AT5G21010 for *BPM5* and AT3G43700 for *BPM6*.

3.1.2. Characterization of *BPM2* splice variants

In the text of this thesis, each *BPM2* transcript isoform is referred to as "*BPM2*" followed by a period and the variant number from AtRTD3. From the transcript sequences, I determined the length, GC content and individual nucleotide content of each of the 16 transcripts of *BPM2* using the package Biostrings (version 2.70.1; Pagès et al., 2017). I calculated the number of exons in each transcript from the GTF file. Using the Wilcoxon rank sum test, I examined these characteristics differ significantly with transcript coding potential (coding versus non-coding). Using the R package msa (version 1.36.0; Bodenhofer et al., 2015), I performed multiple alignment with the ClustalW algorithm on the *BPM2* transcript isoforms and the gDNA sequence of *BPM2*. I manually fixed certain alignment errors based on exon coordinates. I created an identity distance matrix of these alignments using the package seqinr (version 4.2-36; Charif and Lobry, 2007). The identity distance matrix contained squared roots of the pairwise distances. Using neighbor-joining (NJ) method within the package ape (version 5.8; Paradis and Schliep, 2019), I clustered the transcripts and the gDNA based on the identity distance matrix. To determine which *BPM2* transcripts correspond to those in the TAIR database, I repeated the described multiple alignment and clustering procedure after adding *BPM2* transcript sequences from TAIR to the multiple alignment. I also added the known part of the sequence of the new transcript isoform identified by Pali (2020) to this analysis.

3.1.3. Characterization of *BPM2* splice variant translation products

Using the amino-acid sequences from Table S9, I determined the length of translation products (BPM2 protein isoforms) and compared them between coding and unproductive transcripts using the Wilcoxon rank sum test. The names used for the proteins in the rest of the text denote their lengths in amino acids. Using the R package msa (version 1.36.0; Bodenhofer et al., 2015), I performed multiple alignment with the Muscle algorithm on the BPM2 protein isoforms. I manually adjusted the alignment for the bpm2_295 protein (encoded by *BPM2.8*) because of poor alignments achieved by available algorithms. I created an identity distance matrix and a similarity distance matrix based on the Fitch matrix of mutational distance (Fitch, 1966) with squared roots of the pairwise distances using the package seqinr (version 4.2-36; Charif and Lobry, 2007). I clustered the proteins based on the identity distance matrix using the NJ method within the package ape (version 5.8; Paradis and Schliep, 2019). Based on the identity distance matrix, I determined which transcripts code for identical proteins. I searched for conserved domains in BPM2 protein isoforms using the online tool Batch CD-Search (Wang et al., 2023; Marchler-Bauer et al., 2011; Marchler-Bauer and Bryant, 2004).

3.1.4. Transcriptome mapping

I downloaded reads from the RNA-seq libraries created by Zhang et al. (2022; study accession PRJNA755474) that were related to temperature stress (sample accessions: SRR23291381, SRR23291382, SRR23291390, SRR23291398 and SRR23291399). Characteristics of the libraries related to temperature stress are listed in Table 1. I performed a quality check using the tool FastQC (Galaxy Version 0.74+galaxy0) within the web platform Galaxy (The Galaxy Community, 2024) and, based on the results, trimmed the first 100 bases from the 5’-end with FASTQ Trimmer (Galaxy Version 1.1.5).

Table 1. Characteristics of RNA-seq libraries related to temperature stress (cold and heat).

library	condition	Age	tissue	library type
SRR23291381	Cold	5-week-old plants	Rosettes from different exposures - pooled	Telo2
SRR23291382	Cold	5-week-old plants	Rosettes from different exposures - pooled	Clontech
SRR23291390	Cold	5-week-old plants	Rosettes from different exposures - pooled	Telo
SRR23291398	Heat	5-week-old plants; seedlings	Different temperatures and exposures - pooled	Telo2
SRR23291399	Heat	5-week-old plants; seedlings	Different temperatures and exposures - pooled	Telo

I mapped the trimmed reads to the AtRTD3 transcriptome (Zhang et al., 2022) using Minimap2 (Li, 2018; Galaxy Version 2.26+galaxy0) with the preset *PacBio/Oxford Nanopore read to reference mapping (-Hk19) (map-pb)*, K-mer size = 16 and disabled spliced alignment. This mapping is referred to as k16 in the rest of the thesis. To obtain a more specific mapping, I performed another mapping with reads that only had 75 bases trimmed from the 5'-end, with increased K-mer size (28) and increased minimizer window size (19). This mapping is referred to as k28 in the rest of the thesis. All other parameters were kept the same. I examined the mapping qualities using the tool Samtools flagstat (Galaxy Version 2.0.5).

I imported the Pairwise mApping Format (PAF) mapping files into R to analyze the reads mapped to *BPM2* transcript isoforms. I set the criteria for unique mapping to be mapq > 0. I compared the number and mapq values of uniquely mapped reads in each library using the Kruskal-Wallis test. Using packages DESeq2 (version 1.44.0; Love et al., 2014) and pheatmap (version 1.0.12; Kolde 2019), I created count matrices of uniquely mapped reads for all libraries and used them to group the reads in the different libraries. For grouping, I used a complete linkage method on a matrix of Euclidean distances. I examined the reads that map to *BPM2* transcript isoforms - their total number, mapq values, and number of reads from each library uniquely mapped to each *BPM2* transcript isoform. Using R packages DGEobj.utils (version 1.0.6; Thompson et al., 2022) and edgeR (version 4.2.0; Chen et al., 2024), I performed TPM (Transcripts Per Kilobase Million) and RPKM (Reads Per Kilobase Million) normalization based on uniquely mapped reads. Recommended control genes (Škiljaica et al., 2022) had very little reads uniquely mapped to them

(< 10 for the vast majority) and had significant fluctuations in different libraries, so they were not used for normalization.

3.2. Plant Material and Growth Conditions

Arabidopsis thaliana (L.) Heynh. ecotype Col-0 plants were used in this work. Plant cultivation, somatic embryogenesis and tissue harvesting were performed by Dunja Leljak Levanić and Mateja Jagić, as described in Ivanić (2022). Seven types of tissue were harvested – oval leaves of the rosette, flower buds, open flowers, ovules, zygotic embryos in the cotyledonary phase, somatic embryos in the induction phase and somatic embryos in the maturation phase. For oval leaves of the rosette, flower buds and open flowers (Fig. 3), 20-51 mg of tissue per sample was collected. For ovules, zygotic (Fig. 4) and somatic embryos (Fig. 5), less than 10 mg of tissue per sample was collected.

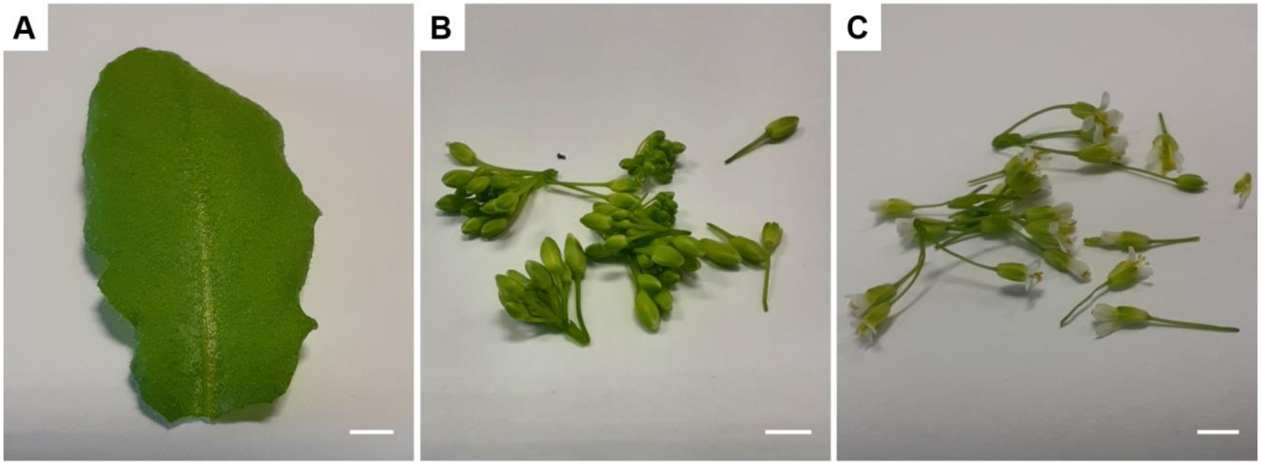


Figure 3. Tissues harvested from *Arabidopsis thaliana* (L.) Heynh. ecotype Col-0 plants. A) oval leaf of the rosette, B) flower buds and C) open flowers. Scale indicates 2 mm.

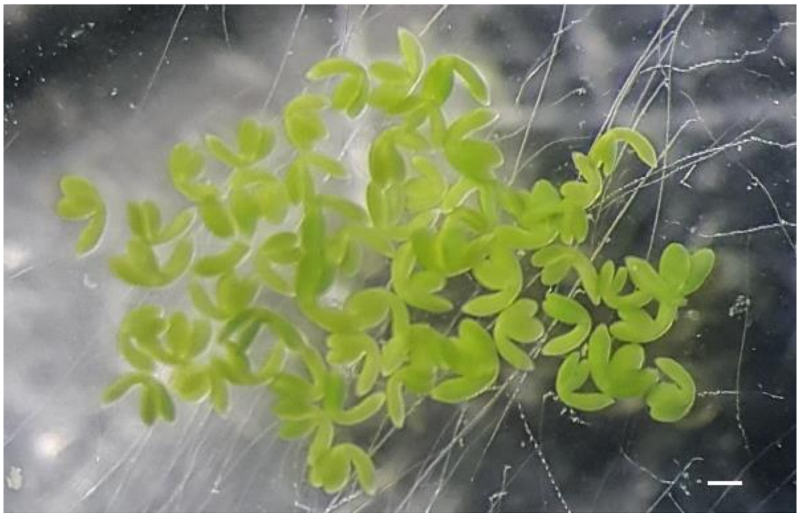


Figure 4. Cotyledonary zygotic embryos of *Arabidopsis thaliana* (L.) Heynh. ecotype Col-0. Scale indicates 200 µm. Source: Demir (2021).

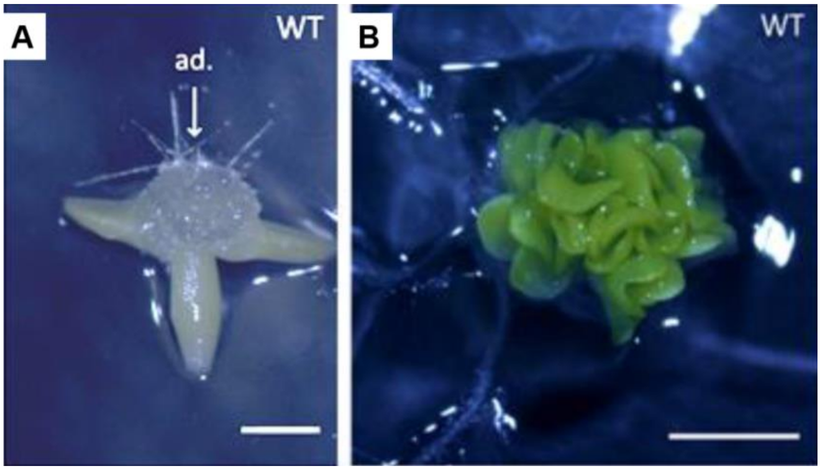


Figure 5. Somatic embryos wild type (WT) *Arabidopsis thaliana* (L.) Heynh. ecotype Col-0 in the A) induction and B) maturation phase. Scale indicates 1 mm. Arrow pointed shows induction of embryogenic calus on the adaxial side of the cotyledon. Source: Ivanić (2022).

3.3. RNA Extraction and Reverse Transcription

I extracted total RNA from tissue samples with biomass > 10 mg (oval rosette leaves, flower buds and open flowers) using the MagMAX™ Plant RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific) according to the manufacturer's instructions. I measured RNA concentrations and purities (260/280 and 260/230 scores) using NanoDropTM 1000 Spectrophotometer (Thermo Fisher Scientific). I prepared three biological replicates for each sample. For the reverse transcription (RT) reaction, I mixed 635.95 ng total RNA, 5 µM

Oligo(dT)$_{18}$ primer (Thermo Fisher Scientific), 200 units of RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific), 20 units of RiboLock RNase inhibitor (Thermo Fisher Scientific), 1 mM dNTP mix (Sigma-Aldrich; 1 mM of each dNTP) and 1× Reaction Buffer (Thermo Fisher Scientific) in a total volume of 20 µL. Mixtures of RNA and Oligo(dT)$_{18}$ primers were incubated for 5 min at 65 °C, after which I added the rest of the reagents (enzymes, buffer and dNTP mix). The RT reaction mixtures were incubated for 60 min at 42 °C and 10 min at 70 °C. I diluted the cDNA solutions 4× in nuclease-free water for downstream applications and stored them at 4 °C.

I extracted messenger RNA from tissue samples with biomass < 10 mg (ovules, zygotic embryos, somatic embryos in the induction phase and somatic embryos in the maturation phase) using the Dynabeads™ mRNA DIRECT™ Micro Purification Kit (Invitrogen, Thermo Fisher Scientific) according to the manufacturer's instructions, without the last Tris-HCl wash. I eluted the mRNA by incubating the magnetic beads in 7.5 µL of 10 mM Tris-HCl at 80 °C for 2 min and transferring the supernatant to a clean 200 µL PCR tube. I repeated the incubation step in a new 7.5 µL of 10 mM Tris-HCl and added the supernatant to the first eluate. This double elution was performed to increase mRNA yield. I prepared at least two independent extractions (biological replicates) for each sample. For the reverse transcription (RT) reaction, I mixed 30.50 ng of extracted mRNA, 2.86 µM Oligo(dT)$_{18}$ primer (Thermo Fisher Scientific), 40 units of RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific), 20 units of RiboLock RNase inhibitor (Thermo Fisher Scientific), 0.286 mM dNTP mix (Sigma-Aldrich; 0.286 mM of each dNTP) and 1× Reaction Buffer (Thermo Fisher Scientific) in a total volume of 14 µL. Mixtures of RNA and Oligo(dT)$_{18}$ primers were incubated for 5 min at 65 °C, after which I added the rest of the reagents (enzymes, buffer and dNTP mix). The RT reaction mixtures were incubated for 60 min at 42 °C and 10 min at 70 °C. I diluted the cDNA solutions 2× in nuclease-free water for downstream applications and stored them at 4 °C.

I checked for presence of cDNA in RT mixtures and possible contamination with gDNA by performing PCR with ACT3 primers, described in section Primer Design and Specificity Determination. I performed the PCR as described in section 3.5. Polymerase Chain Reactions (Fig. A4, Fig. A5).

3.4. Genomic DNA extraction

I extracted gDNA from an oval rosette leaf of an *Arabidopsis thaliana* (L.) Heynh. ecotype Col-0 plant using a "quick and dirty" method. I harvested the tissue in a 1.5 mL tube containing 10 µL of glass homogenization beads (SiLibeads Type S, Sigmund Linder) and immediately flash-froze it in liquid nitrogen. I homogenized the tissue using a silamat (Silver Mix, C.M.F. Srl.) in two 8 s intervals, freezing the tissue in liquid nitrogen in between intervals. I added 100 µL of extraction buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA pH 8.0, 0.5% SDS), vortexed the mixture and centrifuged it at 14,000 g for 5 min at 25 °C using a Brinkmann Eppendorf 5415 C Centrifuge. I transferred 75 µL of the supernatant to a clean 1.5 mL tube, added 150 µL of ethanol (96% v/v) and vortexed the mixture. I centrifuged the tubes at 14,000 g for 10 min at 25 °C, removed the supernatant and left the tubes open for 20 min to dry the precipitate. I resuspended the percipitate in 50 µL of TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0) by vortexing, centrifuged the suspension at 14,000 g for 5 min at 25 °C and transferred the supernatant to a clean 1.5 mL tube. I prepared a 10× diluted aliquot of the gDNA for downstream use and stored both the dilution and the non-diluted solution at 4 °C.

3.5. Polymerase Chain Reactions

For all PCR reactions, I prepared mixtures containing 1× EmeraldAmp® GT PCR Master Mix (Takara Bio Inc.), 200 nM forward and reverse primer and 1 µL of template solution in a total volume of 25 µL. PCR was performed in a GeneAmp® PCR System 2700 (Applied Biosystems) with the initial denaturation at 98 °C for 3 min, followed by 40 cycles of denaturation at 98 °C for 10 s, annealing at 58, 60 or 61 °C for 30 s, extension at 72°C for 1 min/kb, and a final extension at 72°C for 5 minutes. After amplification, the reaction mixtures were stored at 4 °C.

3.6. Agarose-gel electrophoresis

I prepared 1.5% agarose gels in TAE buffer (40 mM Tris base, 20 mM glacial acetic acid, pH 8.0, 1 mM EDTA) and loaded 5 µL of PCR samples and 3 µL of molecular markers into wells. I used GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific) and GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific) (Fig. 6) as molecular markers depending on the expected size of the amplicon. I separated the DNA fragments by electrophoresis for 30 min at 100 V (RunOne™ System, Embi Tec). I stained the gels in a 10 ng/L ethidium bromide solution for 15-20 min and photographed them under UV light using a Kodak EDAS 290 hood, with 3.5 s exposure time and 100% UV strength.

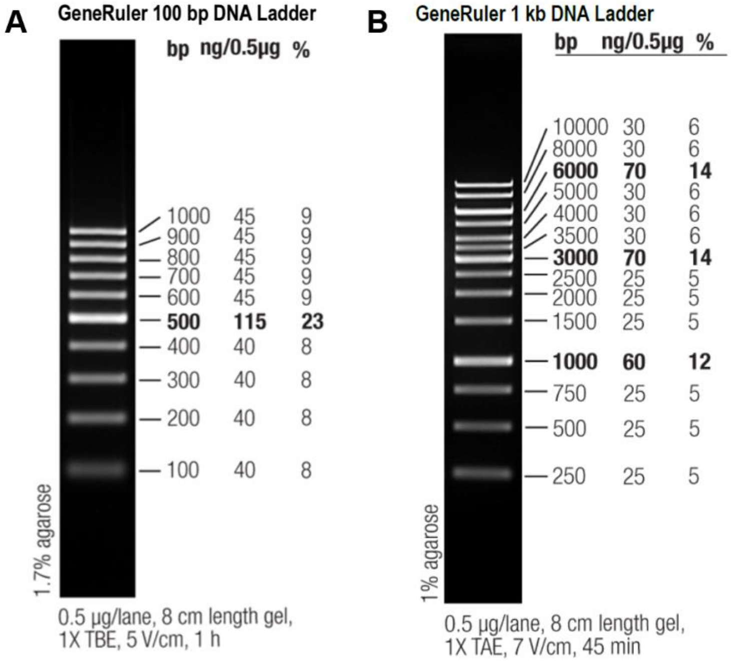


Figure 6. Molecular markers for DNA fragment sizes. A) GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific), B) Gene Ruler 1 kb DNA Ladder (Thermo Fisher Scientific).

3.7. Quantitative Polymerase Chain Reactions

Because embryonic tissues are hardly available and have very small biomass, zygotic embryos weren't used in qPCR experiments, and only one biological replicate of somatic embryos in the maturation phase was used. Two biological replicates were used for all other tissues. For all qPCR reactions, I prepared mixtures containing 1× GoTaq® qPCR Master Mix reagent, (Promega), 200 nM forward and reverse primer and 1 µL of template solution (equivalent to 7.95 ng total RNA extracted using the MagMAX™ Plant RNA Isolation Kit or 1.09 ng mRNA extracted using the Dynabeads™ mRNA DIRECT™ Micro Purification Kit) in a total volume of 10 µL. I performed three technical replicates for all qPCR reactions. The reactions were performed in a Mic qPCR Cycler (Bio Molecular Systems). The thermal profile for amplicons *PUX7*, *BPM2*.9 and *BPM2.3-15* was the following: initial denaturation at 95 °C for 5 min, followed by 8 pre-cycles of touchdown PCR with denaturation at 95 °C for 5 s and combined annealing and extension step at 62 → 58 °C for 20 s (lowering the annealing/extension temperature by 0.5 °C in each of the 8 pre-cycles), followed by 40 cycles of 95 °C for 5 s and 60 °C for 20 s. The thermal profile for the *BPM2.8* amplicon was the same as for the other amplicons, except the annealing/extension step was 60 s instead of 20 s due to the length of the amplicon. I determined the number of different amplicons in a single reaction by examining the number of peaks in the melting curve obtained after amplification using the following parameters: 72 °C to 95 °C with ramp speed of 0.3 °C per second.

For each sample, I discarded the technical replicate with the biggest deviation from the mean. I analyzed relative expression of transcripts *BPM2.8*, *BPM2.9* and *BPM2.3-15* as described in Škiljaica (2022), using *PUX7* as a reference gene. I calibrated the expression in flower buds and open flowers to the expression in oval rosette leaves; and the expression in somatic embryos to the expression in ovules. I tested the difference in expression of each variant in different tissues using the Kruskal-Wallis test with the Pairwise Wilcoxon test with Bonferroni correction as a post-hoc test. Differences with a P value of < 0.05 were regarded as significant.

3.8. Primer Design and Specificity Determination

Sequences of primers used for standard PCR, along corresponding with annealing temperatures, elongation times, expected targets and fragment sizes, are listed in Table 2. Primers BPM2.3-15_fw and BPM2_univ_rev amplify both *BPM2.3* and *BPM2.15* variants which can to be distinguished based on fragment size. Sequences of primers used for qPCR are listed in Table 3. I checked primer specificity by filtering the AtRTD3 transcriptome (Zhang et al., 2022) for those transcripts which contained the sequence complementary to the 10 most 3'-terminal bases of the primer in either strand of cDNA, and then using NCBI Primer-BLAST (Ye et al., 2012) on those cDNA sequences as a custom database. I used two databases (filtered by forward and reverse primer complementarity) for each primer pair Primer-BLAST.

Table 2. Primers for standard PCR reactions. Primers ACT3_fw and ACT3_rev were taken from Tokić (2024). Primers BPM2.8_fw, BPM2.9_fw, BPM2.3-15_fw and BPM2_univ_rev were designed by Mateja Jagić. Target genes are based on NCBI Primer-BLAST (Ye et al., 2012) results. Ta is annealing temperature used in PCR reactions, and ET is elongation time used in PCR reactions.

Primer	Sequence	target	product size / bp	Ta / °C	ET / min:s
ACT3_fw	CTGGCATCATACTTTCTACAATG	*ACT3* cDNA (AT3G53750)	650	58	1:00
		gDNA:			
		ACT3 (AT3G53750)	733		
		ACT1 (AT2G37620)	750		
		ACT2 (AT3G18780)	728		
		ACT4 (AT5G59370)	798		
ACT3_rev	CACCACTGAGCACAATGTTAC	*ACT7* (AT5G09810)	736		
		ACT8 (AT1G49240)	746		
		ACT9 (AT2G42090)	902		
		ACT11 (AT3G12110)	805		
		ACT12 (AT3G46520)	749		
		actin-like atpase superfamily protein (AT2G42100)	792		
BPM2.8_fw	CTTTAGAAGTTGAGGCTGAAAGCTG	*BPM2.8* (AT3G06190.8)	310	60	0:30
BPM2_univ_rev	GCTAGCTGAACAACACAGATCAAC				
BPM2.9_fw	GTACAAGCCCCTATTTTCAAGACTTG	*BPM2.9* (AT3G06190.9)	565	60	0:30
BPM2_univ_rev	GCTAGCTGAACAACACAGATCAAC				
BPM2.3-15_fw	TTCTCCCTTTAACTCTCTTTCTGGAC	*BPM2.3* (AT3G06190.3)	1221	61	1:30
		BPM2.15 (AT3G06190.15)	876		
BPM2_univ_rev	GCTAGCTGAACAACACAGATCAAC	*BPM2* gDNA (AT3G06190)	1221		

Table 3. Primers for qPCR reactions. Primers qA-PUX7-Fw and qA-PUX7-Rev were taken from Tokić (2024). Primers qBPM2.6_fw and qBPM2.6_rev were taken from Pali (2020). Primers qBPM2.2_fw and qBPM2.2_rev were designed by Andreja Škiljaica. Primers BPM2.8_fw, BPM2_univ_rev, BPM2_univ_fw and qBPM2.9_rev were designed by Mateja Jagić.

primer	Sequence	target	product size / bp
qA-PUX7-Fw	GTTTCTCAGACTATCAAAGCCA	*PUX7* (AT1G14570.1)	120
qA-PUX7-Rev	ATCAATTACAAGCACCACGG		
qBPM2.2_fw	AGTTGATGGAGAAACATTTCCTG	*BPM2.8* (AT3G06190.8)	121
qBPM2.2_rev	AGCCTCAACTTCTAAAGAATTGG		
BPM2.8_fw	CTTTAGAAGTTGAGGCTGAAAGCTG	*BPM2.8* (AT3G06190.8)	310
BPM2_univ_rev	GCTAGCTGAACAACACAGATCAAC		
BPM2_univ_fw	CTGCAGTTTTCAGGGCACAGC	*BPM2.9* (AT3G06190.9)	115
qBPM2.9_rev	AGAGTTGATGCCCATTTCAAGTCTTG		
qBPM2.6_fw	CCCTATTTTCAAGGTTCTCCCT	*BPM2.3* (AT3G06190.3) *BPM2.15* (AT3G06190.15)	107
qBPM2.6_rev	CAGCCTCAACTTCTAAAGCTAC		

4. Results

4.1. Splice variants of all *BPM* genes

Searching AtRTD3 (Zhang et al., 2022) by TAIR gene ID, I identified 15 splice variants of *BPM1*, 16 splice variants of *BPM2,* 11 splice variants of *BPM3*, 6 splice variants of *BPM4*, 2 splice variants of *BPM5* and 6 splice variants of *BPM6*. In total, there were 56 splice variants of *BPM* genes. In this thesis, splice variants of *BPM2* are referred to as *BPM2* followed by the accession number from AtRTD3.

4.2. *BPM2* splice variants

Of the 16 splice variants of BPM2 identified, 15 were sourced from the new Iso-seq-based AtRTD3 transcriptome, while *BPM2.8* was obtained from AtRTD2, as indicated by their annotations in the AtRTD3 BED file. There were 8 transcripts described as coding, and 8 described as unproductive, suspected to be substrates of the NMD pathway based on the position of the stop codon in the first ORF. Alignment of *BPM2* transcripts to chromosome 3 is shown in Figure 7.

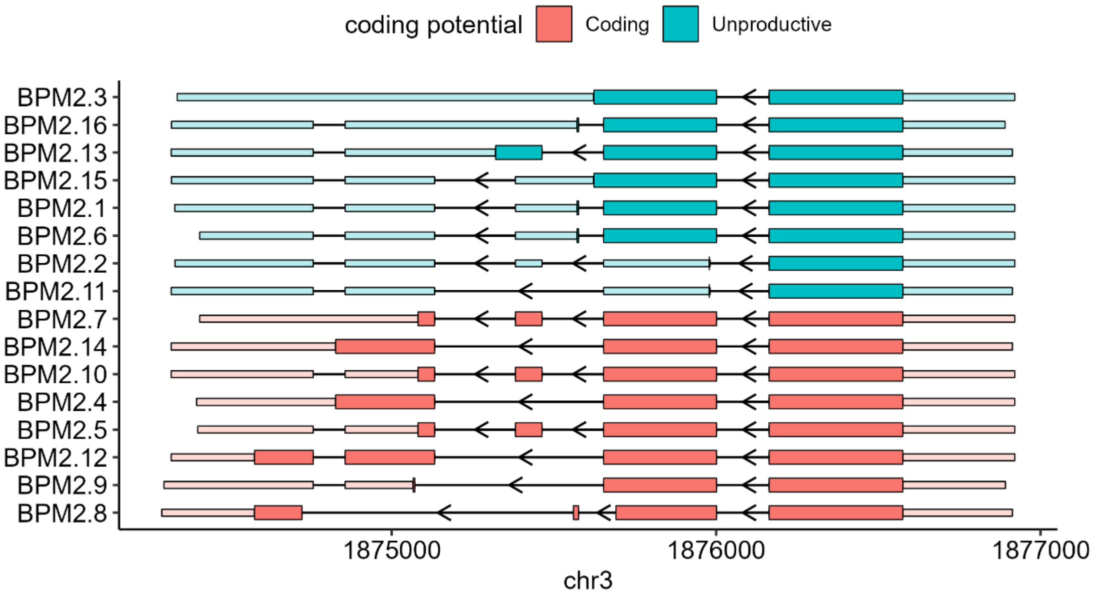


Figure 7. Transcripts of gene *BPM2* aligned by their coordinates on chromosome 3. Transcripts are sorted by coding potential and length. Thick colored lines represent the CDS, thinner lightly colored lines represent 5'-UTRs and 3'-UTRs, while thin blask lines represent introns.

Since the names of the *BPM2* splice variants in TAIR don't correspond to those in AtRTD3, I inquired the corresponding transcripts in the two databases by clustering all available sequences

(Fig. 8) and examining which transcripts have the same splice sites. I also did this for the variant identified by Pali (2020). Transcripts obtained by Iso-seq have different and more accurately determined TSS and TES than those described previously, so TSS and TES weren't taken into consideration while matching transcripts from different databases. Transcripts *BPM2.4, BPM2.5* from TAIR and the variant *BPM2.6* identified by Pali (2020) couldn't be unambiguously assigned to a single AtRTD3 transcript, because there are variants in AtRTD3 that differ only in TSS and TES location, but have all the same internal splice sites.

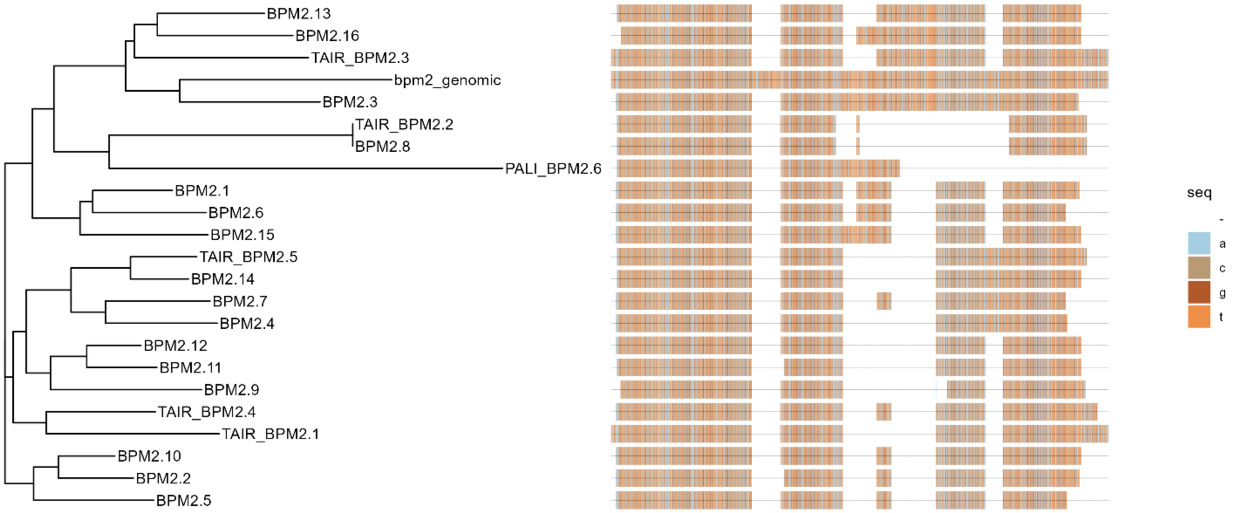


Figure 8. Clustering of sequences of *BPM2* transcripts from AtRTD3, *BPM2* transcripts from TAIR (Berardini et al., 2015) and *BPM2* gDNA. Clustering was performed based on an identity distance matrix of aligned sequences, by the neighbour-joining algorithm. Sequence alignment was done with *ClustalW* and fixed manually based on exon coordinates.

The corresponding transcript pairs from AtRTD3 and TAIR, as well as the new variant identified by Pali (2020), are listed in Table 4. Splice variants *BPM2.1*, *BPM2.2*, *BPM2.3, BPM2.6, BPM2.7*, *BPM2.9*, *BPM2.11*, *BPM2.15* and *BPM2.16* were not described in the publicly available TAIR database. The partial sequence identified by Pali (2020) was present in variants *BPM2.3* and *BPM2.15*.

Table 4. Names of corresponding transcript variants of gene BPM2 in databases TAIR (Berardini et al., 2015) and AtRTD3. Partial sequence of variant *BPM2.6* was identified by Pali (2020) and isn't present in the TAIR database.

TAIR	AtRTD3
BPM2.1	*BPM2.12*
BPM2.2	*BPM2.8*
BPM2.3	*BPM2.13*
BPM2.4	*BPM2.5 / BPM2.10*
BPM2.5	*BPM2.4 / BPM2.14*
*BPM2.6**	*BPM2.3 / BPM2.15*

* identified by Pali (2020), not present in the TAIR database.

In downstream analysis I examined certain characteristics of transcripts isoforms from AtRTD3 database. Lengths of *BPM2* transcripts ranged from 1512 bp (*BPM2.8*) to 2422 bp (*BPM2.3*), with unproductive transcripts being significantly longer than coding transcripts (Fig. 9A, B). All transcripts had between 2 and 5 exons, with no significant difference based on predicted coding potential (Fig. 9C, D).

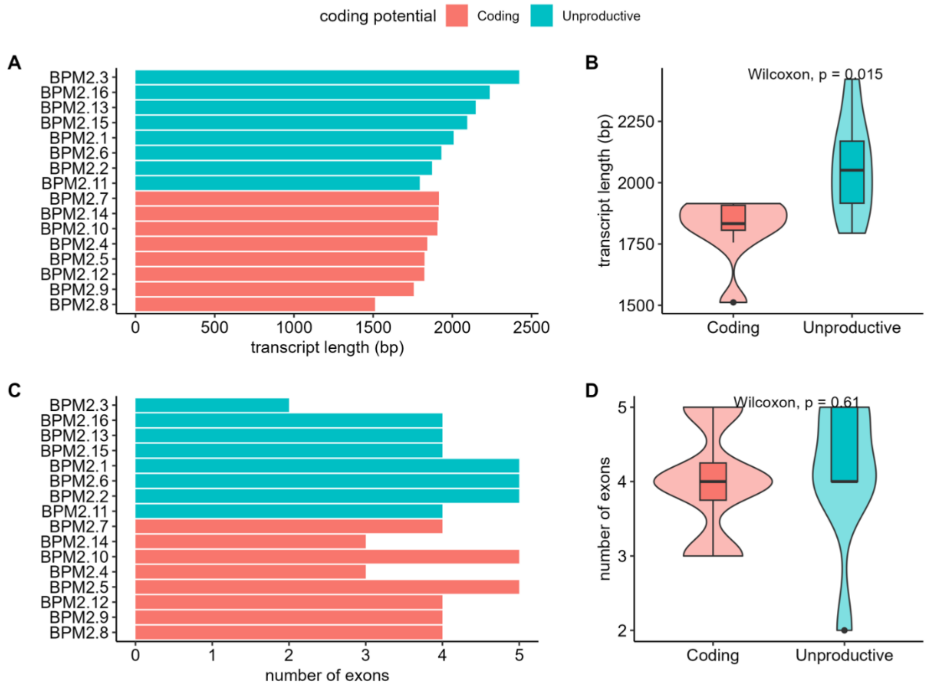


Figure 9. Characteristics of *BPM2* transcript isoforms. **A)** transcript lengths (bp), sorted by coding potential and value. **B)** Distribution of lengths of coding and unproductive transcripts. Unproductive transcripts are significantly longer than coding transcripts (p = 0.015, Wilcoxon rank sum test). **C)** Number of exons in transcripts, sorted in the same way as in 2.A). **D)** Distribution of the number of exons in coding and unproductive transcripts. There is no significant difference in the number of exons between coding and unproductive transcripts (p = 0.61, Wilcoxon rank sum test).

GC content in transcripts was 0.447 ± 0.006 with no significant difference between coding and unproductive transcripts (Fig. 10A). Looking at individual nucleotide content (Fig. 10B), cytosine was somewhat underrepresented while thymine is somewhat overrepresented in all transcripts. There is significantly more adenine in coding than in unproductive transcripts (p = 0.003, Wilcoxon rank sum test), and significantly more thymine in unproductive than in coding transcripts (p = 0.038, Wilcoxon rank sum test).

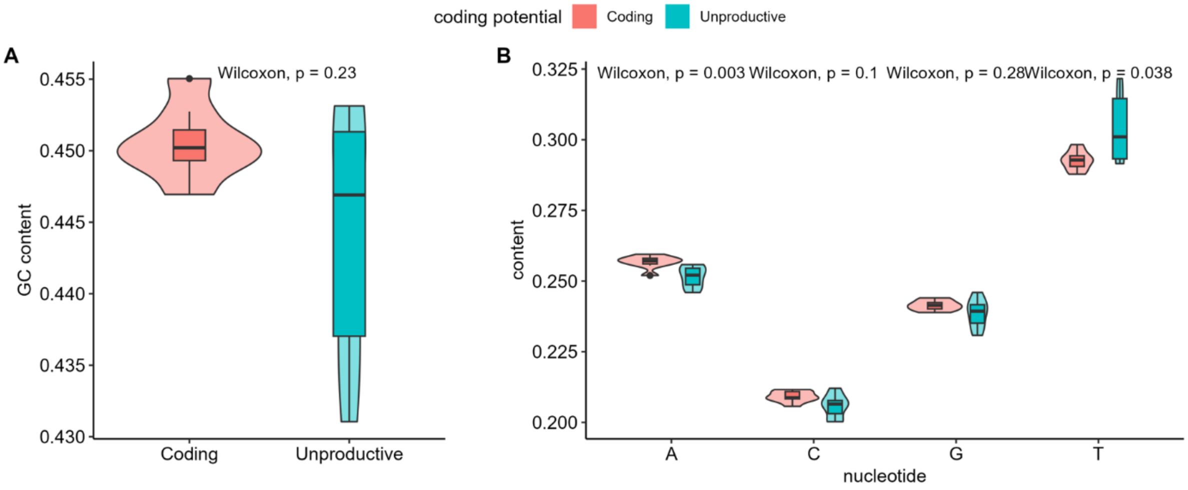


Figure 10. Sequence content of *BPM2* transcripts. **A)** GC content and **B)** individual nucleotide content in coding and unproductive transcripts. The only statistically significant differences based on predicted coding potential are in adenine (p = 0.003, Wilcoxon rank sum test) and thymine content (p = 0.038, Wilcoxon rank sum test).

4.3. *BPM2* splice variant translation products

After characterizing the transcript isoforms of *BPM2*, I investigated the proteins encoded by these variants from AtRTD3 database. The lengths of proteins encoded by *BPM2* splice variants ranged from 137 to 406 amino acids (aa). Proteins derived from coding transcripts were significantly longer than those produced from unproductive transcripts, as shown in Figure 11.

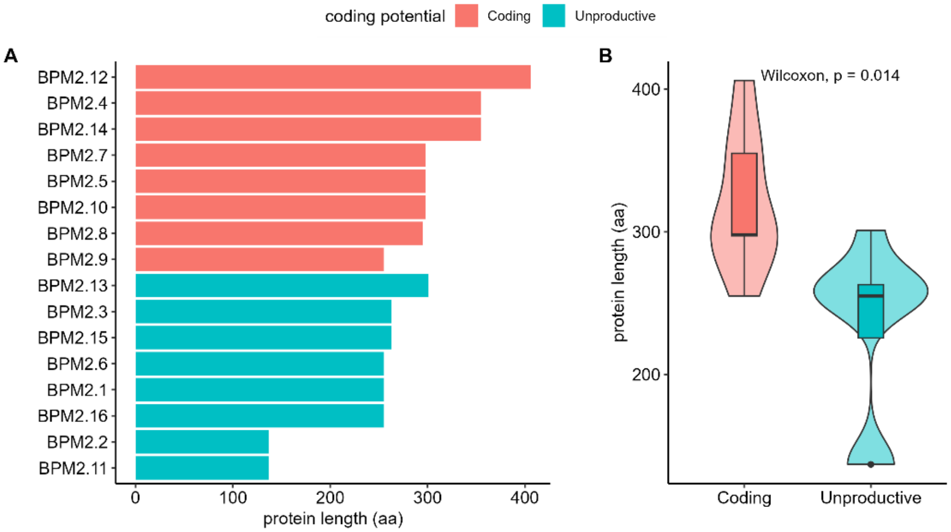


Figure 11. Lengths of *BPM2* transcript variant translation products. **A)** protein lengths in amino acids (aa), sorted by coding potential and value. **B)** Distribution of lengths of coding and unproductive transcripts. Proteins encoded by coding transcripts are significantly longer than those encoded by unproductive transcripts (p = 0.014, Wilcoxon rank sum test).

Certain splice variants coded for identical proteins, as is shown in the dendrogram based on the identity distance matrix of aligned sequences (Fig. 12, Fig. A1). I found that 16 splice variants of *BPM2* encode 9 unique proteins. *BPM2.2* and *BPM2.11* encode for the shortest protein, while the protein encoded by *BPM2.12* is the longest, followed by the protein encoded by *BPM2.4* and *BPM2.14*.

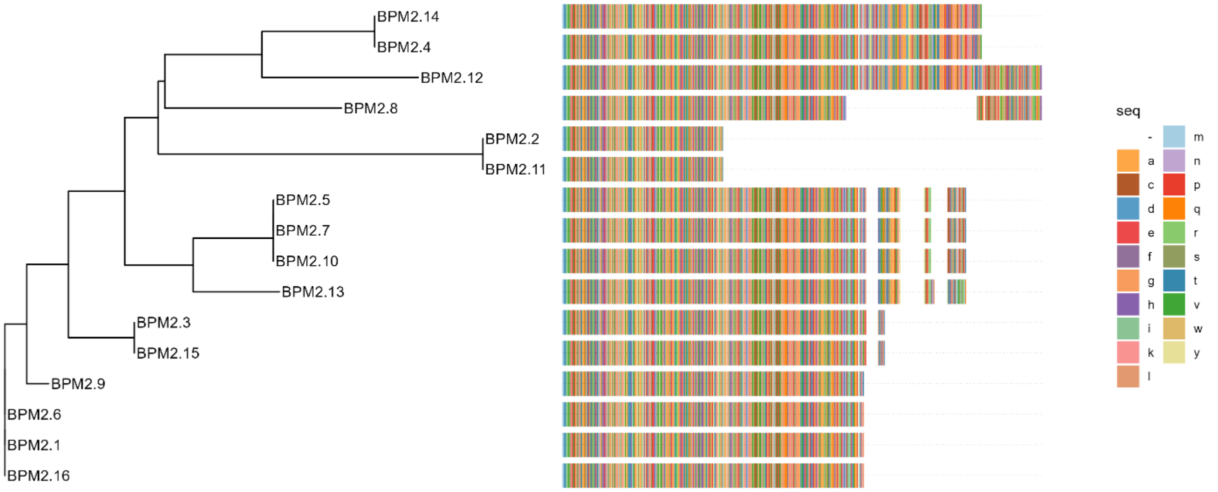


Figure 12. Clustering of sequences of *BPM2* transcript variant translation products. Clustering was performed based on an identity distance matrix of aligned sequences, by the neighbor-joining (NJ) algorithm. Sequence alignment was done using the *Muscle* algorithm and adjusted manually for BPM2.8. Proteins in the same terminal nodes have identical sequences.

Furthermore, the proteins and transcripts that encode them, as well as their conserved domains determined by Batch CD-Search, are listed in Table 5. The names used for the proteins in the rest of the text denote their lengths in amino acids. I identified a complete conserved N-terminal MATH domain in eight proteins. However, the MATH domain in the shortest protein, bpm2_137, lacked its C-terminal region and was the only conserved domain present in this variant. Only the two longest proteins (bpm2_406 and bpm2_355) had a complete BTB domain. Only the N-terminal half of the BTB domain was present in the remaining 6 proteins that had it (bpm2_301, bpm2_298, bpm2_295, bpm2_263, bpm2_255a and bpm2_255b). The two longest proteins (bpm2_406 and bpm2_355) also had a BACK domain, although it was missing the C-terminus in bpm2_355. Detailed schematic representation of these domains analyzed with Batch CD-Search are in Figure A2.

Table 5. Proteins of all *BPM2* splice variants, sorted by predicted coding potential of the transcript and by protein size (aa). The names used for the proteins denote their lengths in aa. Transcripts are labeled with their AtRTD3 and TAIR ID. Putative conserved domains were predicted using Batch CD-Search (Wang et al., 2023; Marchler-Bauer et al., 2011; Marchler-Bauer and Bryant, 2004).

Coding potential	**index**	**protein**	**AtRTD3**	**TAIR**	**Conserved domains**	
					complete	**missing C-terminus**
Coding	1	bpm2_406	*BPM2.12*	*BPM2.1*	MATH, BTB, BACK	-
	2	bpm2_355	*BPM2.4*	*BPM2.5*	MATH, BTB	BACK
			BPM2.14	*BPM2.5*		
	3	bpm2_298	*BPM2.5*	*BPM2.4*	MATH	BTB
			BPM2.7	-		
			BPM2.10	*BPM2.4*		
	4	bpm2_295	*BPM2.8*	*BPM2.2*	MATH	BTB
	5	bpm2_255a	*BPM2.9*	-	MATH	BTB
Unproductive	6	bpm2_301	*BPM2.13*	*BPM2.3*	MATH	BTB
	7	bpm2_263	*BPM2.3*	-	MATH	BTB
			BPM2.15	-		
	8	bpm2_255b	*BPM2.1*	-	MATH	BTB
			BPM2.6	-		
			BPM2.16	-		
	9	bpm2_137	*BPM2.2*	-	-	MATH
			BPM2.11	-		

4.4. Mapping reads from heat and cold treated libraries to AtRTD3

Reads from libraries related to cold or heat stress were mapped to the AtRTD3 transcriptome using Minimap2, the gold standard tool for long read mapping (Liyanage et al., 2023; Liu et al., 2022; Li et al., 2018). Two mappings were done with different read pre-processing and different K-mer sizes, as described in Materials and methods 3.1.4. Transcriptome mapping. According to Samtools flagstat, at least 99.69% of reads from each library were successfully mapped to the AtRTD3 transcriptome in each mapping. Primary mappings were achieved for 98.47% to 99.29% of reads. Numbers and percentages of total and primary read mappings in each library for each mapping are listed in Table A1. In the original k16 mapping, 57.28% of reads had a unique mapping (mapq > 0), while the rest of the reads mapped equally well to at least two places in the transcriptome. In the more specific k28 mapping, 63.84% had a unique mapping (mapq > 0).

Most transcripts in AtRTD3 had under 10 reads that uniquely mapped to them in each library, and most of those reads had a mapq < 10 (Fig. 13). Values of mapq were significaltly higher in the k28 mapping than in the k16 mapping ($p < 2.2e-16$, Wilcoxon rank sum test), with mean values 14.40 and 13.28, respectively.

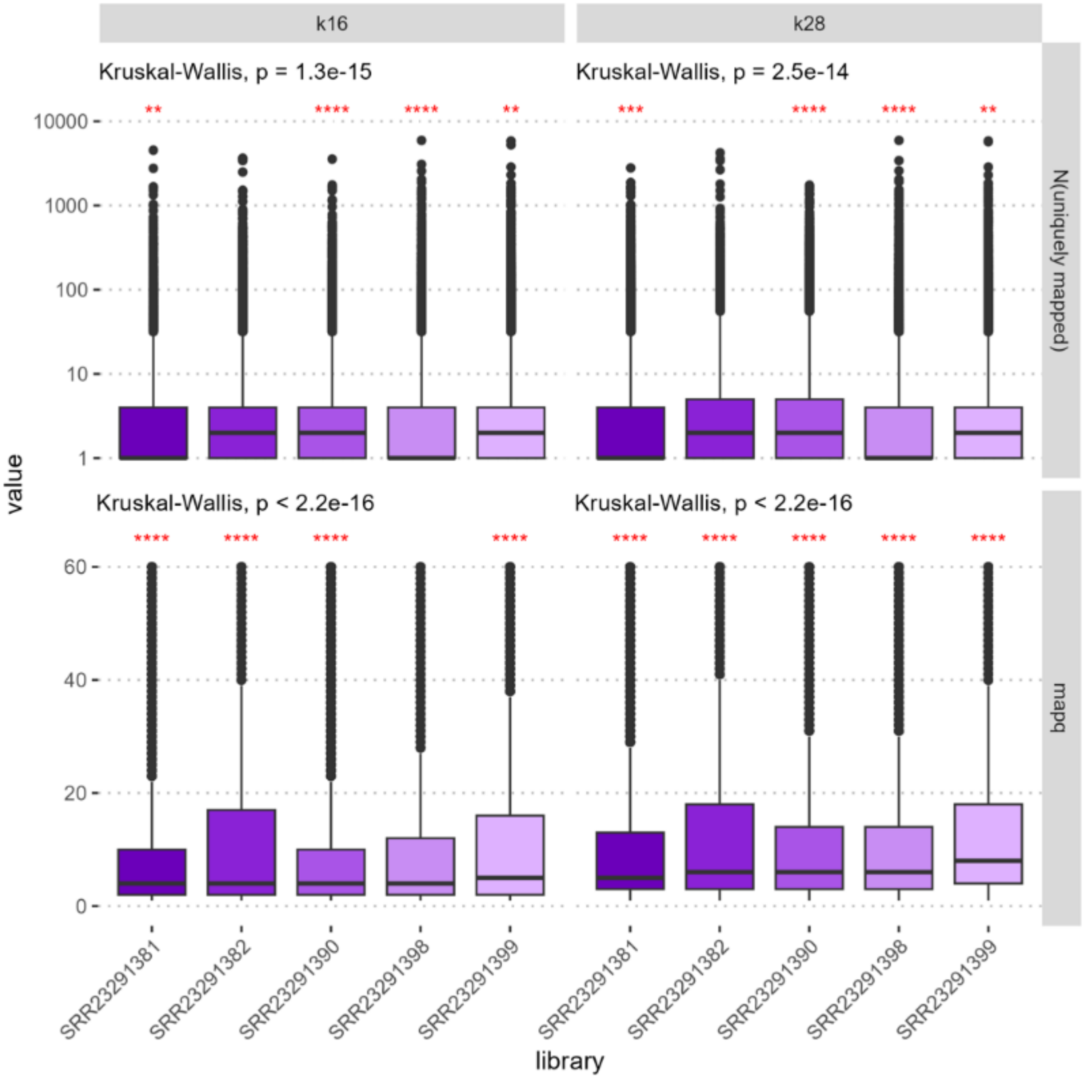


Figure 13. Number of reads per transcript and mapping quality (mapq) in the k16 mapping (100 bp trimmed from 5'-end of reads, k-mer size = 16) and the k28 mapping (100 bp trimmed from 5'-end of reads, k-mer size = 16, minimizer window size = 19). Only uniquely mapping reads (mapq > 0) were counted. (Kruskall-Wallis test, p-value < 0.05). Wilcoxon rank sum test, ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001.

Reads from the k16 mapping showed the strongest clustering corresponding to the treatment conditions and tissue (rosettes from cold treated 5-week-old plants; pooled heat treated seedlings and rosettes from heat treated 5-week-old plants), with libraries from heat treated samples further clustering based on library type – Telo and Telo2 libraries showed less distance to each other than to the Clontech library (Fig. 14).

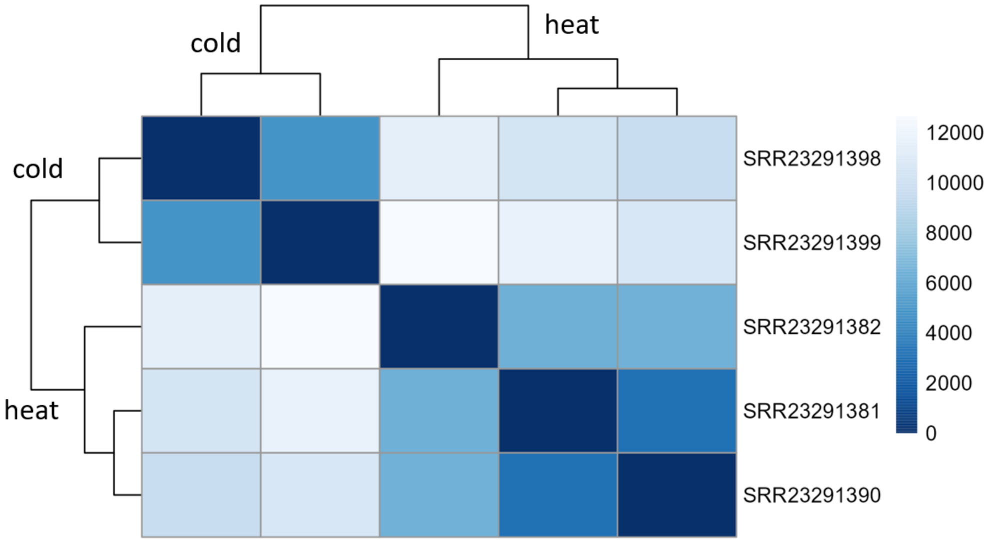


Figure 14. Clustering of read mappings in different libraries. Clustering is done with complete linkage based on the Euclidean distance matrix of count matrices of uniquely mapped reads. Libraries differ in biological sample and library type, as listed in Table 1.

4.4.1. Reads uniquely mapped to *BPM2* splice variants

The variant *BPM2.12* had the greatest number of reads uniquely mapped to it out of all *BPM2* splice variants. It was also the only one present in all 5 libraries. Other represented variants were *BPM2.1*, *BPM2.14* and *BPM2.15*, present in cold-treated sample libraries, and *BPM2.9*, present in one heat-treated sample library. The raw number, TPM and RPKM values of reads uniquely mapped to specific *BPM2* splice variants are shown in Figure 15A-C. There were 68 reads mapped to *BPM2* splice variants in the k16 mapping. Of those, 60 were uniquely mapped to a *BPM2* splice variant, 7 had no unique mapping, and 1 was uniquely mapped to a non-*BPM2* transcript. In the k28 mapping, 67 total reads were mapped to *BPM2* splice variants, of which 59 were uniquely mapped to a *BPM2* splice variant and 8 had no unique mapping. Most of the reads that uniquely mapped to *BPM2* splice variants were mapped to *BPM2.12*, with other variants only being represented by one or two reads. There was also a significant increase in mapq for reads uniquely mapped to *BPM2* splice variants in k28 compared to k16 (Fig. 15D).

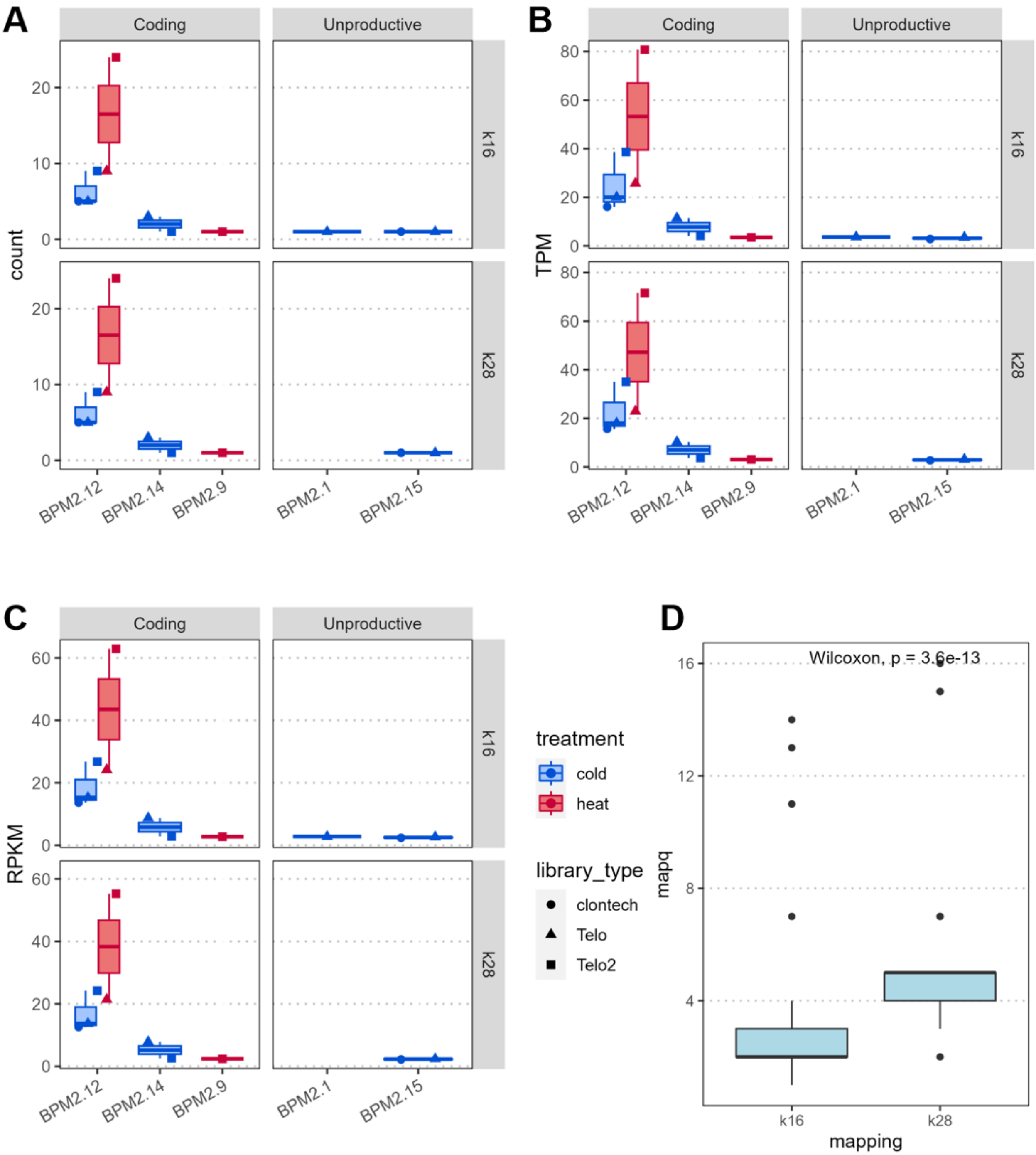


Figure 15. A) Raw number, B) TPM, C) RPKM and D) mapping quality (mapq) of reads uniquely mapped to *BPM2* splice variants in the k16 mapping (100 bp trimmed from 5'-end of reads, k-mer size = 16) and the k28 mapping (100 bp trimmed from 5'-end of reads, k-mer size = 16, minimizer window size = 19). Values of mapq are significantly higher in the k28 mapping compared to the k16 mapping (p = 3.6e-13, Wilcoxon rank sum test).

4.5. Presence of splice variants *BPM2.3*, *BPM2.8*, *BPM2.9* and *BPM2.15* in various vegetative and reproductive tissues

Electrophoresis after amplification with specific *BPM2.8* primers showed a band around 350 bp in all seven tissues (Fig. 16). This band was slightly larger than the expected 310 bp. When using a slightly higher annealing temperature (61 °C instead of 60 °C) and a longer elongation time (90 s instead of 30 s), another fainter band was present in the gel at around 500 bp in all samples that had the 350 bp band (Fig. 17). Primer dimers were also visible in gels as diffuse bands around 100 bp, in all samples (including gDNA and no-template controls), with all primer pairs specific for *BPM2* splice variants. As expected, in the genomic DNA sample, no fragments amplified with specific primers for *BPM2.8* nor with specific primers for *BPM2.9* (Fig. 16, 17 and 18) because those primers span exon-exon junctions specific to those splice variants, and are therefore not complementary to the genomic sequence of *BPM2*.

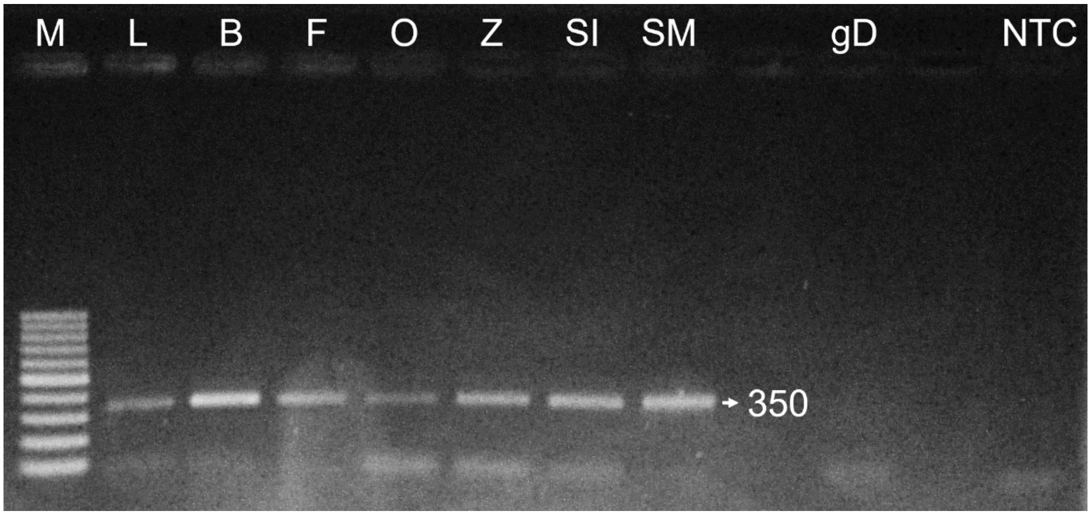


Figure 16. Agarose gel showing PCR products (**Ta = 60 °C, ET = 30 s)** using primers **BPM2.8**_fw and BPM2_univ_rev on cDNA from oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), ovules (**O**), cotyledonary zygotic embryos (**Z**), somatic embryos in the induction phase (**SI**), somatic embryos in the maturation phase (**SM**), a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Marker used (**M**) is GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific). Numbers denote band sizes in bp. Abbrevation: Ta-anneling tempelature, ET-elongation time.

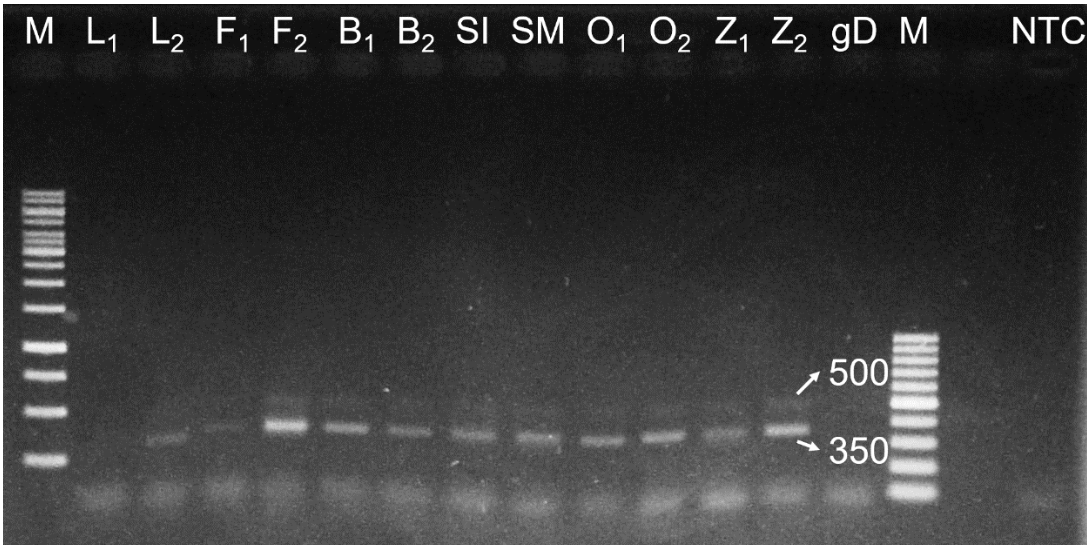


Figure 17. Agarose gel showing PCR products (**Ta = 61 °C, ET = 90 s**). using primers **BPM2.8**_fw and BPM2_univ_rev on cDNA from oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), ovules (**O**), cotyledonary zygotic embryos (**Z**), somatic embryos in the induction phase (**SI**), somatic embryos in the maturation phase (**SM**), a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Index numbers mark different biological samples. Markers used (**M**) are GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific) on the left and GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific) on the right. Numbers denote band sizes in bp. Abbrevation: Ta-anneling tempelature, ET-elongation time.

Expected size of the *BPM2.9* fragment amplified using specific primers was 565 bp. Bands of that size were only present in cotyledonary zygotic embryos and somatic embryos in the maturation phase, but its appearance was hardly visible due to the dominant band present in all samples around 610 bp. Open flowers, ovules and both somatic embryo samples also has a band around 750 bp. Zygotic embryos and somatic embryos in the maturation phase each had an additional unique band, at approximately 900 bp and 300 bp, respectively. In total, there were 5 different bands from amplification with specific *BPM2.9* primers, as shown in Figure 18.

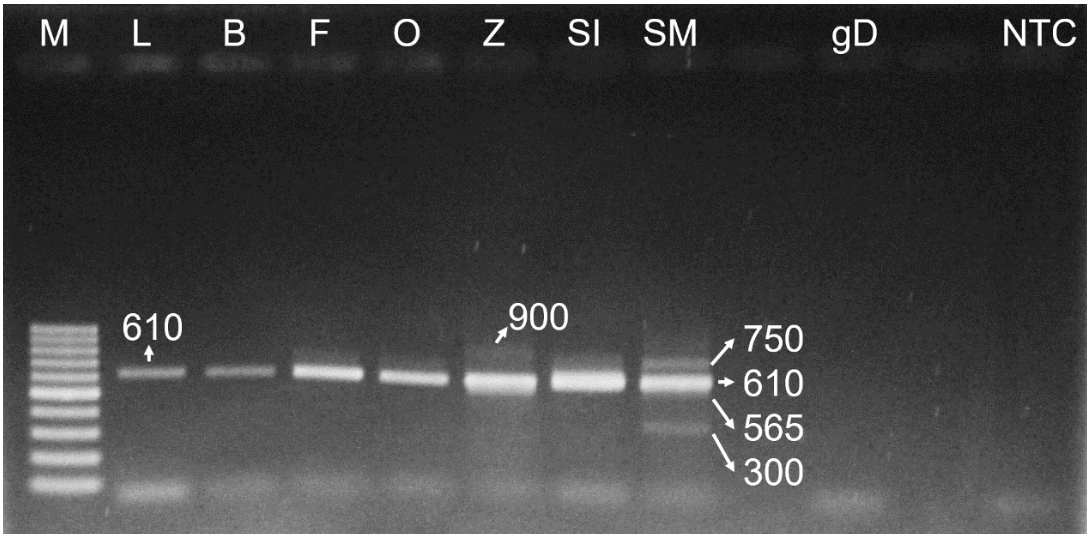


Figure 18. Agarose gel showing PCR products (**Ta = 60 °C, ET = 30 s)** using primers **BPM2.9**_fw and BPM2_univ_rev on cDNA from oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), ovules (**O**), cotyledonary zygotic embryos (**Z**), somatic embryos in the induction phase (**SI**), somatic embryos in the maturation phase (**SM**), a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Markers used (**M**) are GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific) on the left and GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific) on the right. Numbers denote band sizes in bp. Abbrevation: Ta-anneling tempelature, ET-elongation time.

Primers BPM2.3-15_fw and BPM2_univ_rev were expected to amplify cDNA of splice variants *BPM2.3* and *BPM2.15*, resulting in amplicons of 1221 bp and 876 bp, respectively. The primer pair also amplified a 1221 bp region of *BPM2* in gDNA, as expected. The band corresponding to *BPM2.15* was present in all seven tissues, although not in all biological replicates of oval rosette leaves and ovules (Fig. A3). The band corresponding to *BPM2.3* was present in all seven tissues except in ovules. There were two additional bands present in all tissues, one very bright around 1100 bp and one very faint around 2000 bp. Ovules, zygotic and somatic embryos had an additional faint band around 1500 bp. Two additional bands were only present in one type of tissue – one around 990 bp in oval rosette leaves, and a faint one around 260 bp in cotyledonary zygotic embryos. In total, there were 7 different bands from amplification with primers specific to *BPM2.3* and *BPM2.15*, as shown in Figure 19.

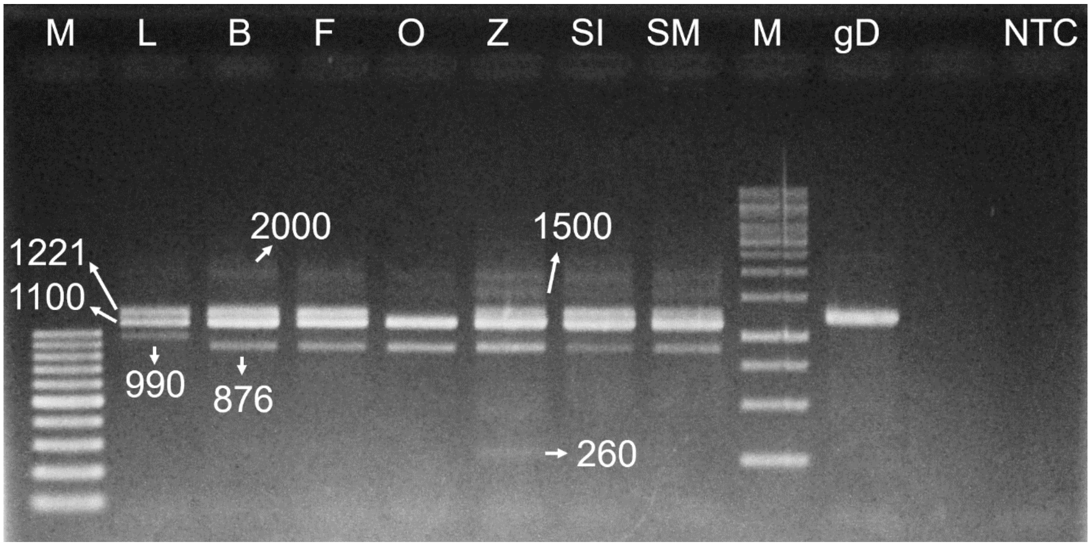


Figure 19. Agarose gel showing PCR products (**Ta = 61 °C, ET = 90 s)** using primers **BPM2.3-15**_fw and BPM2_univ_rev on cDNA from oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), ovules (**O**), cotyledonary zygotic embryos (**Z**), somatic embryos in the induction phase (**SI**), somatic embryos in the maturation phase (**SM**), a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Markers used (**M**) are GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific) on the left and GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific) on the right. Numbers denote band sizes in bp. Abbrevation: Ta-anneling tempelature, ET-elongation time.

4.6. Relative quantification of splice variants *BPM2.3*, *BPM2.8*, *BPM2.9* and *BPM2.15* in various vegetative and reproductive tissues

Different tissues in which RNA was isolated by the same method were compared regarding expressions of each splice variant – *BPM2.8*, *BPM2.9*, and combined *BPM2.3* and *BPM2.15*. Oval rosette leaves, flower buds and open flowers were compared with Kruskal-Wallis tests, and ovules, somatic embryos in the induction phase (SI) and somatic embryos in the maturation phase (SM) were compared with independent Kruskal-Wallis tests. There was no amplification in any sample with qPCR primers specific for *BPM2.8* (qBPM2.2_fw and qBPM2.2_rev), so standard PCR primers for *BPM2.8* (BPM2.8_fw and BPM2_univ_rev) were used.

Expression of *BPM2.3* and *BPM2.15* was significantly higher (p = 0.029, Pairwise Wilcoxon test with Bonferroni correction) in oval rosette leaves than in flower buds and open flowers (Fig 20A, Table 6). No significant difference was observed in the expression of *BPM2.8* between oval rosette leaves, flower buds and open flowers, as it varied more between biological replicates than between the different tissues (Table 6). The differences in expression of *BPM2.9* between these tissues wasn't significant either (Fig 20B, Table 6).

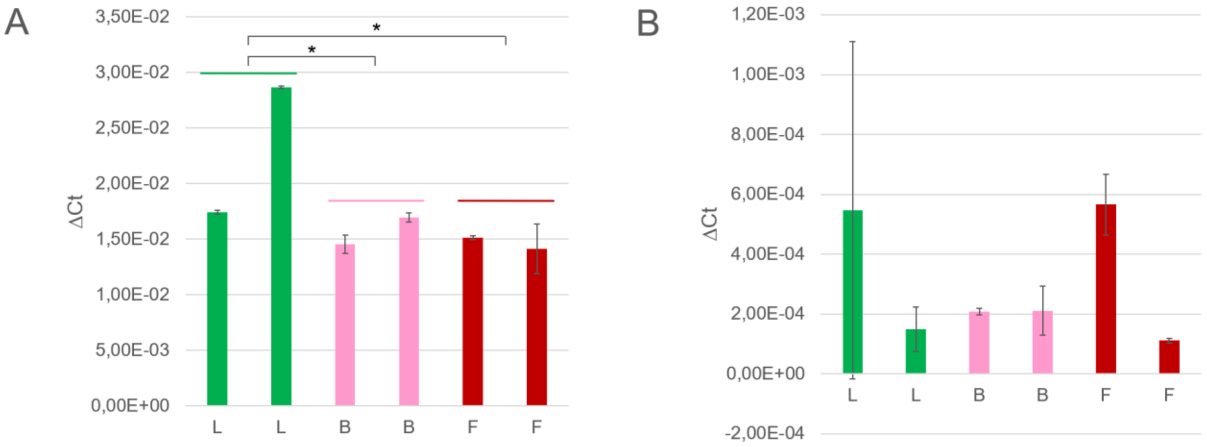


Figure 20. Relative expression values (ΔCt) of **A)** *BPM2.3-15*, **B)** *BPM2.9* in oval rosette leaves (L), flower buds (B) and open flowers (F). Different bars represent independent biological samples. There are two technical replicates for each biological sample. Difference in expression in different tissues (were tested using the Kruskal-Wallis test ($p < 0.05$) with the Pairwise Wilcoxon test with Bonferroni correction as a post-hoc test (ns: $p > 0.05$, *: $p <= 0.05$).

Table 6. Values and standard deviations (SD) of ΔΔCt for oval rosette leaves (L), flower buds (B) and open flowers (F), calibrated to the expression in oval rosette leaves.

transcript	tissue	ΔΔCt	SD
BPM2.8	L	1.00	2.47
	B	1.34	1.88
	F	1.59	2.43
BPM2.9	L	1.00	1.56
	B	0.73	0.17
	F	0.88	0.95
BPM2.3-15	L	1.00	0.29
	B	0.70	0.07
	F	0.65	0.06

Expression of *BPM2.3* and *BPM2.15* was significantly lower ($p = 0.029$, Pairwise Wilcoxon test with Bonferroni correction) in ovules than in SI. Although the expression of *BPM2.3* and *BPM2.15* in SM showed no statistically significant difference to either ovules or SI, it was much more similar to SI, with ΔΔCt being over 12 for both SI and SM when calibrated against ovules (Fig 21A, Table 7). The differences in expression of *BPM2.8* between ovules, SI and SM weren't significant, but expression in the SM sample seemed to be lower than in the other samples

(Fig. 21B, Table 7). That reduced expression in the SM sample wasn't noticed for *BPM2.9*, which also had no significant difference in expression between ovules, SI and SM (Fig. 21C, Table 7).

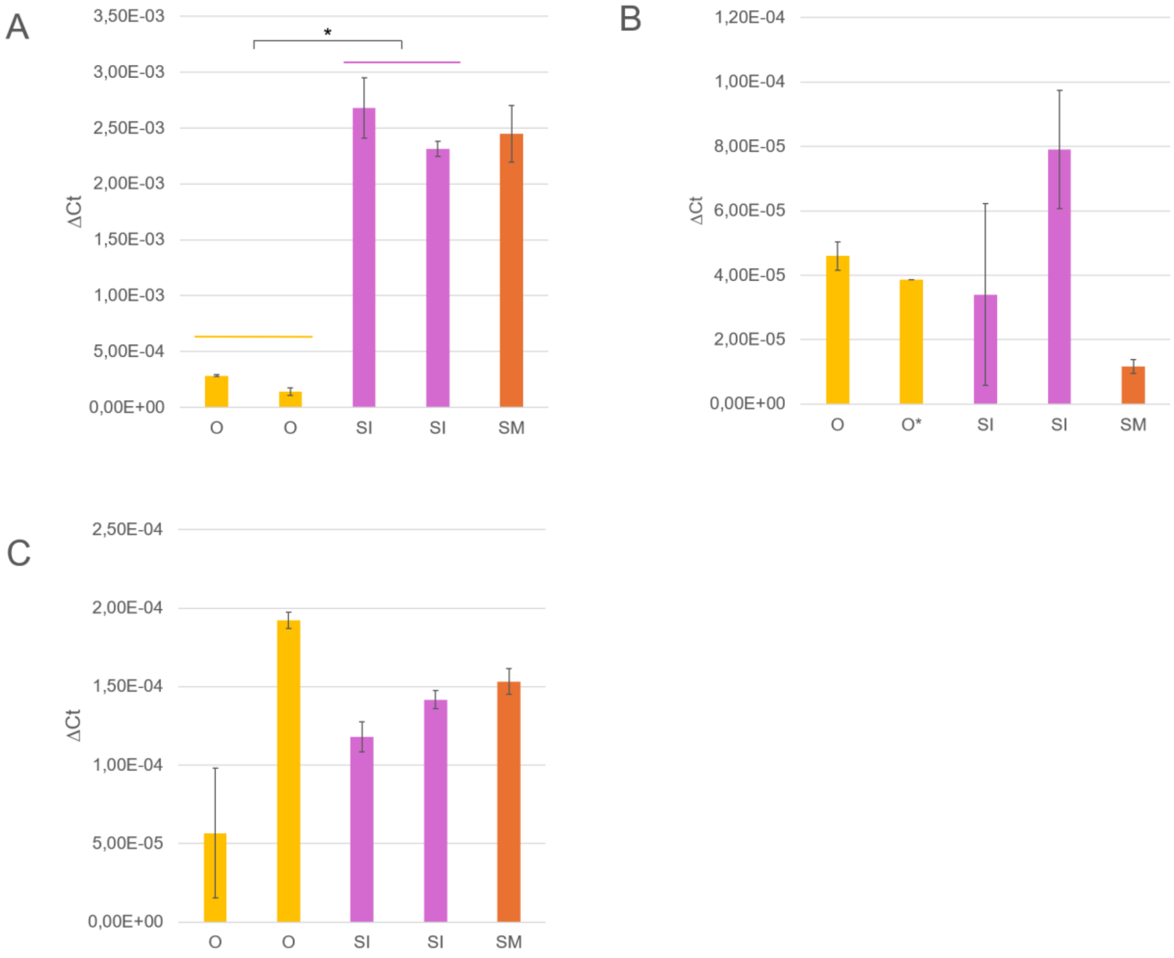


Figure 21. Relative expression values (ΔCt) of **A)** *BPM2.3-15*, **B)** *BPM2.8* and **C)** *BPM2.9* in ovules (O), somatic embryos in the induction phase (SI) and somatic embryos in the maturation phase (SM). Different bars represent independent biological samples. There are two technical replicates for each biological sample except O*, for which there is only one technical replicate. Difference in expression in different tissues were tested using the Kruskal-Wallis test ($p < 0.05$) with the Pairwise Wilcoxon test with Bonferroni correction as a post-hoc test (ns: $p > 0.05$, *: $p <= 0.05$).

Table 7. Values and standard deviations (SD) of ΔΔCt for ovules (O) and somatic embryos in the induction (SI) and maturation (SM) phase, calibrated to the expression in ovules.

transcript	tissue	ΔΔCt	SD
BPM2.8	O	1.00	0.12
	SI	1.19	0.70
	SM	0.27	0.05
BPM2.9	O	1.00	0.75
	SI	1.24	0.14
	SM	1.47	0.08
BPM2.3-15	O	1.00	0.41
	SI	12.36	1.34
	SM	12.16	1.26

4.6.1. Melting curves of splice variants *BPM2.8* and *BPM2.9*

Melting curves obtained after amplification with primers BPM2.8_fw and BPM2_univ_rev had either one peak around 77 °C (melting curve type 8.A) or two peaks, a dominant one around 83 °C and a lower one around 79 °C (melting curve type 8.B), as shown in Figure 22 and Figure A6. Melting curves type 8.A appeared sporadically in all qPCR experiments, not consistently within the three technical replicates, and in negative controls of two out of three experiments. Replicates with type 8.A curves were disregarded in qPCR analyses. Melting curve type 8.B indicates existence of two distinct amplicons that might represent different splice variants. Melting curve type 8.A indicates the formation of primer dimers.

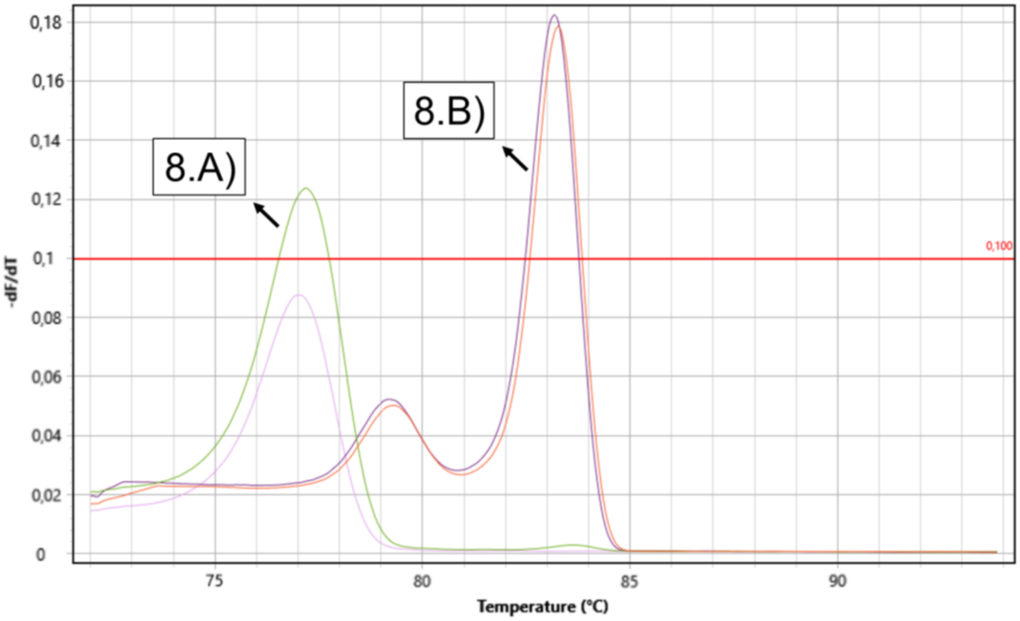


Figure 22. Melting curves of amplicons obtained using primers BPM2.8_fw and BPM2_univ_rev in the no-template control (light purple), and three technical replicates of an ovules sample (green, orange, dark purple). Created with Mic qPCR Cycler (Bio Molecular Systems) device and software.

Each melting curve obtained after amplification with primers BPM2_univ_fw and qBPM2.9_fw had a single peak. For samples in which RNA was isolated using the Dynabeads™ mRNA DIRECT™ Micro Purification Kit (Invitrogen, Thermo Fisher Scientific), there were two significantly different (p = 3.451e-05, Welch Two Sample t-test) peaks – one with Tm = 80.21 °C ± 0.05 °C, and one with Tm = 80.89 °C ± 0.11 °C (Fig. 23). For samples in which RNA was isolated using the MagMAX™ Plant RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific), the peaks Tm values ranged from 80.15 °C to 81.49 °C with a mean of 81.15 °C, but weren't clearly separated into groups (Fig. 24). The Tm values of peaks were not consistent within technical replicates.

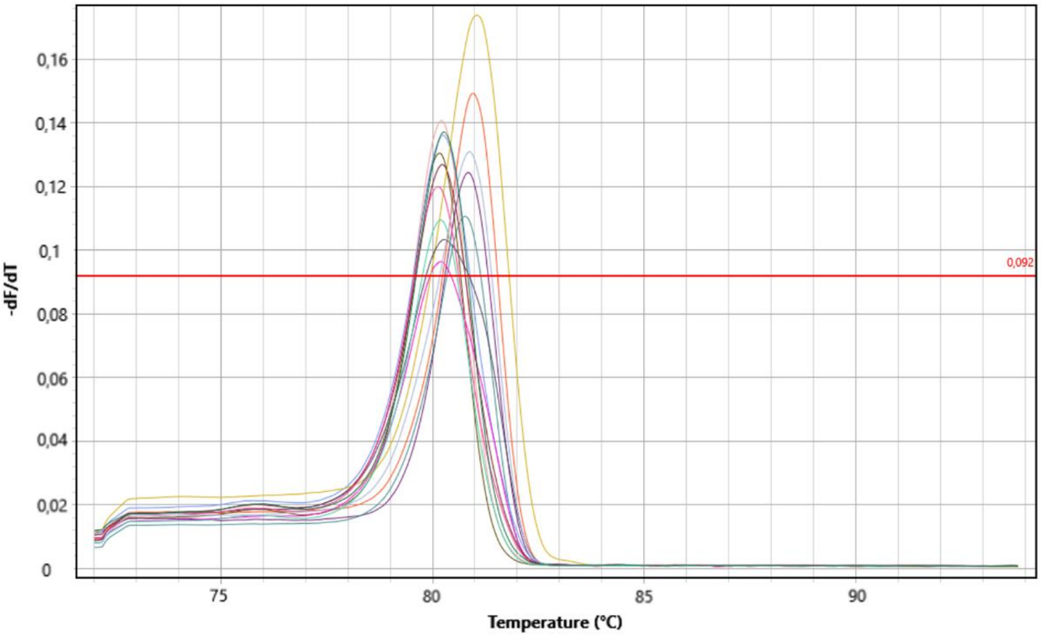


Figure 23. Melting curves of amplicons obtained using primers BPM2_univ_fw and qBPM2.9_fw for samples in which RNA was isolated using the Dynabeads™ mRNA DIRECT™ Micro Purification Kit (Invitrogen, Thermo Fisher Scientific). Created with Mic qPCR Cycler (Bio Molecular Systems) device and software.

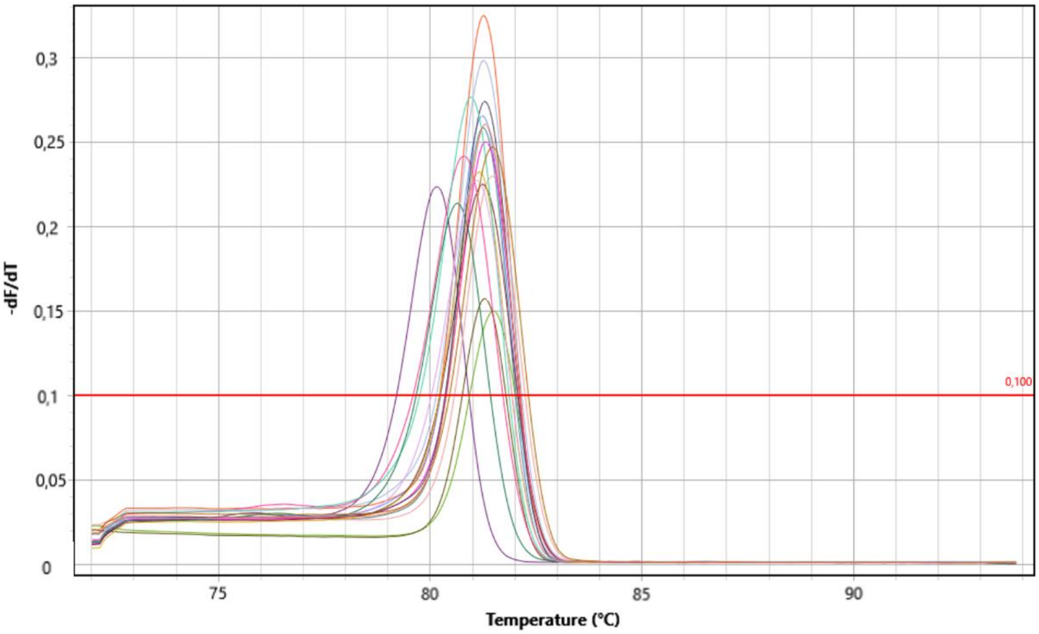


Figure 24. Melting curves of amplicons obtained using primers BPM2_univ_fw and qBPM2.9_fw for samples in which RNA was isolated using the MagMAX™ Plant RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific). Created with Mic qPCR Cycler (Bio Molecular Systems) device and software.

5. Discussion

5.1. Novel splice variants of Arabidopsis *BPM* genes

The results of this thesis indicate that the Arabidopsis *BPM2* gene encodes at least 16 different splice variants that code for nine different proteins, instead of the five splice variants (*BPM2.1*, *BPM2.2*, *BPM2.3*, *BPM2.4*, *BPM2.5*) annotated in TAIR and NCBI databases. The 16 *BPM2* splice variants were identified with a bioinformatic analysis of AtRTD3, the most comprehensive Arabidopsis transcriptome to date (Zhang et al., 2022). AtRTD3 contains seven *BPM2* splice variants with internal splice sites corresponding to the five variants characterized in TAIR (Table 4), and nine variants with novel splice sites. Transcripts *BPM2.4*, *BPM2.5* from TAIR couldn't be unambiguously assigned to a single AtRTD3 transcript because they only differed in the positions of TSS and TES, which are less accurately determined in the TAIR database. Accession numbers of splice variants in AtRTD3 do not correspond to those in TAIR and NCBI. In this thesis, splice variants were referred to as *BPM2* followed by the accession number from AtRTD3.

Novel transcript isoforms for all six known *BPM* genes were also identified in AtRTD3. According to AtRTD3, *BPM1-6* encode 56 splice variants instead of the 17 described in the TAIR database. Genes *BPM1*, *BPM2*, *BPM3*, *BPM4*, *BPM5* and *BPM6* encode 15, 16, 11, 6, 2 and 6 different transcripts, respectively, instead of the 3, 5, 4, 1, 1, and 3 isoforms described in TAIR. These results are significant considering that the *MATH-BTB* gene family is characterized by a significant expansion in genomes of grass species, such as rice, sorghum, Brachypodium, maize (Gingerich et al., 2007; Juranić and Dresselhaus, 2014) and wheat (Škiljaica, 2022) as well as in the nematode *C. elegans* (Stogios et al., 2005). An example of this expansion which is present in a small number of species is part of the lineage-specific expansion and is most commonly associated with expansion of gene families involved in pathogen and stress response, transcription regulation, controlled protein degradation mediated by the ubiquitin system, protein modification, signal transduction, chemoreception, and small molecule metabolism (Lespinet et al., 2002).

Interestingly, the *MATH-BTB* family is not expanded in the *A. thaliana* genome (Gingerich et al., 2005, 2007), nor in the monocot *Musa* lineage (banana). According to Juranić and Dresselhaus (2014), this indicates that expansion of the *MATH-BTB* family in the above listed species is characteristic of grasses and that the proteins of the expanded family possibly possess grass-specific functions. However, functional characterization of certain MATH-BTB proteins in grasses shows that their functions are essential, whether they are involved in cytoskeleton regulation during the reproductive development of maize (Juranić et al., 2012) and wheat (Bauer et al., 2019) or in methylation (Jagić et al., 2022). For example, maize protein ZmMAB1 is involved in the organization of microtubules of the spindle apparatus and in nuclei positioning and identity determination during meiosis II and the first mitotic division in both female and male germlines (Juranić et al., 2012). ZmMAB1 interacts with CUL3, which indicates a function mediated by the CRL3 complex. A similar function is described for the MEL26 protein in nematode *C. elegans* (Pintard et al., 2003). Wheat protein TaMAB2 is expressed only in early embryogenesis, namely in the zygote and the proembryo (Leljak-Levanić et al., 2013). Cytoskeleton proteins actin and tubulin, as well as eukaryotic translation initiation factors eIF3 and eIF4 have been identified as potential interactors and substrates for degradation. These interactions suggest a role of TaMAB2 in translation initiation in early embryogenesis. However, TaMAB2 hasn’t been shown to interact with CUL3, suggesting it might preferentially act via a CUL3-independent mechanism (Bauer et al., 2019).

Since both of the described functions of grass MATH-BTB proteins couldn’t be considered as grass-specific functions, it opens the question of how other, non-grass species compensate for the small number of *MATH-BTB* genes. This question is partially answered in this thesis, since a search of AtRTD3 revealed 56 splice variants of *BPM* genes, comparable to the number of *MATH-BTB* genes in grass species. This indicates that alternative splicing has the potential to achieve the same evolutionary contribution to developmental features as lineage-specific expansion (Lespinet et al., 2002). However, all nine proteins encoded by the 16 splice variants of *BPM2* were found to have an identical MATH domain sequence (or a truncated MATH domain with the partial sequence identical to proteins encoded by other variants, in case of the shortest BPM2 protein), while MATH-BTB proteins of the grass-specific expanded clade have more diverse MATH domains (Gingerich et al., 2007; Juranić and Dresselhaus, 2014; Škiljaica, 2022) which potentially target a greater number of substrates for proteosomal degradation. Furthermore, a phylogenetic analysis conducted by Leljak-Levanić (personal communication) revealed that all 16 *BPM2* splice variants either cluster within the core clade or form a new clade unrelated to the expanded clade. Taking this into consideration, the *BPM2* variants characterized in this thesis probably can’t achieve the multiplicity of potential functions of MATH-BTB proteins of the grass-specific expanded clade. This leaves the potential functions of the novel *BPM2* variants, whose existence was confirmed in the experimental part of this thesis, open to discovery in further research.

The nine proteins encoded by the 16 splice variants of *BPM2* differed in regions downstream of the MATH domain, mostly in the C-terminal part of the BTB domain and the BACK domain (Fig. 12, Fig. A2, Table 5). In the protein encoded by variant *BPM2.12* (TAIR *BPM2.1*), all three domains were complete and the BACK domain was recognized as BACK-AtBPM-like by NCBI’s Batch CD-Search. The protein encoded by *BPM2.4* and *BPM2.14* also had a BTB and a BACK domain, but BACK domain was truncated and was not recognized as BACK-AtBPM-like. The shortest BPM2 protein, encoded by *BPM2.2* and *BPM2.11,* had only a truncated MATH domain, which would limit its function to the functions of MATH proteins (Marín et al., 2015, Oelmüller et al., 2005). In the remaining six proteins that had the BTB domain, it was truncated. (Fig. A2). The differences in structure, presence and completeness of domains, but also in presence/absence of other functional sequences such as NLS within proteins mediate the develpmental function of MATH-BTB variants and proteins (Leljak Levanić et al., 2012).

5.1.1. Predicted coding potential of *BPM2* transcripts

NMD is a pathway that prevents the production of truncated proteins, functioning as a mRNA quality control pathway and an additional layer of gene expression regulation (Kalyna et al., 2012; Göhring et al., 2014). Zhang et al. (2022) classified transcripts from protein-coding genes as either coding or unproductive based on the presence of typical NMD target features. In case of *BPM2* transcripts, these were long 3’-UTRs, premature termination codons (PTC) and, for most transcripts labeled as unproductive, splice junctions downstream of the PTC. These characteristics were assigned to transcripts by the program TranSuite (Entizne et al., 2020), more specifically the TransFeat module. TransFeat identifies termination codons as premature if their position is below a certain threshold relative to the longest ORF in the gene. This doesn’t take into consideration that a large number of plant transcripts with NMD features, particularly those with intron retention, are NMD-insensitive. Some of these transcripts are protected by remaining in the nucleus, avoiding the cytoplasmic NMD pathway, but others were found in association with ribosomes (Göhring et al., 2014 and references cited therein). In this thesis, proteins encoded by transcripts labeled as coding were found to be significantly longer than the proteins encoded by transcripts labeled as unproductive, but not all proteins from "coding" transcripts were longer than all proteins from "unproductive" transcripts. On the contrary, only the protein encoded by splice variants *BPM2.2* and *BPM2.11* was shorter than the shortest protein encoded by the "coding" transcript *BPM2.9*. Furthermore, most BPM2 protein isoforms had a similar domain composition: a complete MATH domain and a truncated BTB domain. This included all but the longest two proteins, encoded by *BPM2.12*, *BPM2.4* and *BPM2.14*, and the shortest protein, encoded by *BPM2.2* and *BPM2.11*. Meaning, the 11 remaining splice variants encode proteins with a complete MATH domain and a truncated BTB domain regardless of the predicted coding potential (Fig. A2). These results indicate that not all proteins encoded by splice variants predicted to be NMD substrates are genuinely truncated in comparison to the proteins encoded by other variants, underlining the importance of an experimental validation of the predicted coding potential.

5.2. Heat and cold stress RNA libraries

Since reads from specific developmental stages couldn't be distinguished due to the pooling of samples prior to RNA library construction, potential developmental functions were only explored experimentally. Instead, potential association of *BPM2* splice variants with heat and cold stress was explored by mapping reads from libraries related to temperature stress to the AtRTD3 transcriptome. While the percentage of mapped reads from all libraries was excellent (>99%), the mapping quality (mapq) scores were quite low (Fig. 13, Fig. 15D). In theory, mapq scores are the Phred-transformed probabilities of the primary mapping being incorrect, with reads that align equally well to different positions of the reference producing mapq = 0 (Schmidt et al., 2024; Li et al., 2008). The presence of several splice variants for most genes in AtRTD3 was probably the main cause of the low mapq scores, since most reads could map to other splice variants in adition to the correct one. Because this was an intrinsic property of AtRTD3, and because mapq scores were so low, the reads were considered to be uniquely mapped if their mapq scores were > 0.

Splice variant *BPM2.12* (TAIR *BPM2.1*) was found in all examined libraries, and was the only *BPM2* splice variant associated with both heat and cold stress. The novel variant *BPM2.9* was the only other *BPM2* variant found in either of the heat stress related libraries. Novel variants *BPM2.1* and *BPM2.15*, and variant *BPM2.14* (TAIR *BPM2.5*) were found in cold stress related libraries (Fig. 15). All variants except *BPM2.12* were only represented by one or two reads, which points to the possibility of other variants being present in the sampled tissues, but not being captured due to a smaller concentration or simply by chance, since the difference between one read and no reads is very small.

While *BPM2.12* was clearly the dominant variant, both in terms of the number of libraries it appeared in and the number of reads, real quantitative results and differential expression of these variants couldn't be obtained from the reads in these libraries. Firstly, simply the number of reads that mapped to them was a major obstruction. Secondly, there were no replicate libraries and no library that could represent the „control", since the only libraries that didn't have tissues exposed to some kind of stress were mutant libraries and tissue-specific libraries of tissues that weren't present in the heat and cold treated samples. Thirdly, even the tissues present in the heat- and cold stress libraries weren't consistent, with heat stress libraries containing seedlings in addition to rosettes. These issues contributed to the problems with finding control genes which could be used to normalize read counts – potential control genes had to be consistent across temperatures, tissues, and developmental stages. The control genes that were tested across these conditions (Škiljaica et al., 2022) were either not present in all libraries, or they were represented by very few reads, leading to differences in read counts that weren't consistent between libraries, and wildly different scaling factors depending on which of the proposed control genes were included. In the end, reads were only normalized to sequencing depth using TPM and RPKM normalization (Novogene, n.d.). While libraries used for AtRTD3 transcriptome assembly have clearly been excelent for capturing many different transcripts from a variety of tissues, developmental stages, abiotic and biotic stress conditions, as well as RNA degradation mutants, they simply weren't designed for quantitative analysis.

5.3. Expression of splice variants *BPM2.3*, *BPM2.8*, *BPM2.9* and *BPM2.15* in different tissues

In the experimental part of this thesis, expressions of variants *BPM2.3, BPM2.8, BPM2.9* and *BPM2.15* were analyzed. The selected tissues included the vegetative phase of development of adult plants (oval leaves of the rosette), the reproductive phase (flower buds and open flowers) and embryogenesis (ovules, zygotic embryos in the cotyledonary phase and somatic embryos in two developmental phases – induction and maturation). This selection was based on existing knowledge of the functional roles of the characterized MATH-BTB proteins from both the core clade and the expanded clade. Some of these roles are under the precise spatial and temporal control and some are very universal (Juranić et al., 2012; Leljak-Levanić et al., 2013.; Bauer et al., 2019).

Expression of the selected splice variants in embryonic tissues couldn't be compared to expression in other reproductive and vegetative tissues due to the use of different RNA extraction methods. Because embryonic tissues are difficult to obtain, harvest and isolate in larger quantities, Dynabeads™ mRNA DIRECT™ Micro Purification Kit (Invitrogen, Thermo Fisher Scientific) was used since it requires < 10 mg of tissue per sample. This kit specifically extracts mRNA, while the MagMAX™ Plant RNA Isolation Kit (Applied Biosystems, Thermo Fisher Scientific) used for other tissues with higher biomass per sample extracts all RNA molecules. Comparison of starting mRNA concentrations in solutions obtained using these methods isn't possible, and neither is the comparison of qPCR results.

5.3.1. Splice variant *BPM2.8*

One of the splice variants is *BPM2.8* (TAIR *BPM2.2*), which encodes a unique protein. It is the only one out of the 16 *BPM2* splice variants in AtRTD3 that hasn't been identified using PacBio Iso-Seq, and was instead added from the short-read assembly AtRTD2 (Zhang et al., 2022). The previously designed specific qPCR primers for *BPM2.8* (Škiljaica, 2022) never produced any amplicons in the experiments within this thesis nor in any earlier experiments. Because of this, new specific primers designed to detect the variant using standard PCR, with the expected amplicon size 310 bp, were used for both standard PCR and qPCR. The new primers resulted in the amplification of two fragments in all tissues – one about 350 bp, and an additional one about 500 bp long (Fig. 17). Melting curves obtained in qPCR experiments using these same primers had two peaks and thus confirm the presence of two different fragments (Fig. 22). These two fragments might come from novel splice variants that are currently still not described. The assumption that neither of the amplified fragments come from *BPM2.8* is supported by the fact that they weren't amplified using the previously designed specific PCR primers (Škiljaica, 2022), in addition to the fragment sizes being different from the expected *BPM2.8* fragment. The full-length sequences of these potential novel splice variants will be obtained in additional experiments within the project that this thesis was a part of. Even though expression of the potential two novel variants was confirmed in all tested tissues, qPCR showed a reduced expression in a mature somatic embryo sample, although the difference wasn't statistically significant. In a parallel study, Frlin (2024) found no expression of these variants in seedlings, and found it to be stress-related.

The other type of melting curve (Fig. 22) had the classic markers of non-specific amplification caused by primer dimerization – it had a lower Tm (77 °C), was present in some technical replicates regardless of biological sample as well as in no-template controls, and it inhibited the amplification of other fragments. The formation of primer dimers was also apparent in electrophoresis gels of all standard PCR reactions as diffuse bands around 100 bp (Kilobaser, n.d.; Creative BioMart, n.d., Chauhan, 2019), visible in Figures 16-19.

5.3.2. Splice variant *BPM2.9*

The next selected splice variant was the novel *BPM2.9*, which also encodes a unique protein. *BPM2.9* was associated with heat stress in the transcriptome mapping done in the bioinformatic part of this thesis. The expected amplicon size of this variant with the newly designed specific primers was 565 bp, but it was only present in cotyledonary zygotic embryos and somatic embryos in the maturation phase, and the dominant fragment in all tissues was around 610 bp. The dominant fragment was potentially smaller in zygotic and somatic embryos, but it was difficult to determine due to overheating of the electrophoresis machine causing some uneven migration of DNA fragments. Additional unexpected fragments were present in some tissues: in open flowers, ovules and both somatic embryo samples there was a band around 750 bp, in zygotic embryos there was one around 900 bp, and in mature somatic embryos there was one around 300 bp (Fig. 18).

Multiple fragments that might point to additional splice variants are further supported by different peaks of the melting curves obtained in qPCR experiments (Fig. 23, Fig. 24). Unlike *BPM2.8*, the expression of variants amplified by specific *BPM2.9* primers for qPCR was maintained throughout embryogenesis (Fig. 21C, Table 7).

To determine whether the fragments visible in the gel are actually specifically amplified and whether they belong to the *BPM* family, the fragment around 610 bp from ovules was sequenced, and the results showed that the fragment is from a novel *BPM2* splice variant. The actual size of the fragment was 625 bp and the sequence corresponded to the expected *BPM2.9* sequence with an additional 60 bp insertion (Leljak-Levanić, personal communication). In further studies, specific primers should be designed based on the sequence of the insertion, and after amplification of 3’- and 5’-ends with universal primers for *BPM2* splice variants, the complete sequence of the novel variant should be obtained. This same process should be repeated for all fragments which were reproducibly obtained in the performed experiments.

5.3.3. Splice variants *BPM2.3* and *BPM2.15*

The highest number of unexpected fragments, in addition to the expected 1221 bp for *BPM2.3* and 876 bp for *BPM2.15*, were found to be amplified with the primer pair specific for these two variants. *BPM2.3* and *BPM2.15* aren't yet annotated in TAIR and NCBI databases, but they correspond to the variant whose partial sequence was described by Pali (2020). These two variants encode the same protein so they were unified in the expression analysis. Variant *BPM2.15* was present in all tested tissues, but not in all biological replicates of oval rosette leaves, while *BPM2.3* was present in all seven tissues except in ovules. The absence of *BPM2.3* in ovules is strange, since it is present in embryos, which make up the majority of the ovule. This result could have been caused by the experimental methods and errors in tissue sampling. Ovules were isolated from the silique by vortexing which might have led to the contamination of isolated ovules with the surrounding tissue and preferential amplification of the fragment of about 1100 bp, which was present in all tissues and probably represents a novel splice variant. Another unexpected phenomenon recorded in embryonic samples were the 12-fold higher expression levels of *BPM2.3* and *BPM2.15* in somatic embryo samples in both stages compared to ovules (Fig. 21A, Table 7). Somatic embryos have been noted to have certain morphological and histological abnormalities potentially caused by changes in tissue and organ determination (Godel-Jędrychowska et al., 2021 and references cited therein). A comparison of expressions of *BPM2.3* and *BPM2.15* in zygotic and somatic embryos at equivalent stages of development might elucidate the potential role of these variants in tissue and organ determination during embryogenesis.

Oval leaves of the rosette had significantly higher expression levels of *BPM2.3* and *BPM2.15* splice variants compared to flower buds and open flowers (Fig. 20A), although the difference was not as extreme as the one recorded in embryonic samples. Oval rosette leaves were also the only tissue that contained a fragment around 990 bp in standard PCR (Fig. 19). Analyzing the whole sequence of the transcript that produced this fragment might confirm a potentially interesting novel splice variant that could be involved in rosette development.

5.3.4. Alternative splicing in developmental transitions

Transitions from totipotency to pluripotency, and from maternal to zygotic control of development, have been shown to be associated with the appearance of specific transcript isoforms in animal models, namely in *Mus musculus* (Zhang et al., 2024). The results of this thesis – appearance of fragments representing potential new splice variants, some present in all tissues and some tissue-specific, along with trends and changes in expression between different tissues that are different for each examined variant – suggest that alternative splicing is involved in developmental transitions in plants as well, such as the transition from juvenile to adult phase, vegetative adult to reproductive phase, and reproductive phase to development of a new sporophyte (embryogenesis). Wheat MATH-BTB proteins TaMAB1, TaMAB2 and TaMAB3, although they aren't splice variants transcribed from the same gene, do show a controlled domination of related MATH-BTB proteins during fertilization. While TaMAB3 is ubiquitously expressed, the egg cell-derived TaMAB1 is immediately down-regulated after fertilization and cannot be detected in any other tissue of wheat. TaMAB2 is a zygotic-induced gene that is also very tissue-specific and only expressed in zygotes and in two-celled embryos (Leljak-Levanić et al, 2013). The function of TaMAB2 was determined based on its interaction partners (Bauer et al., 2019), but additional extensive research is needed to determine whether TaMAB1 and TaMAB2 perform different developmental functions, and to examine if a similar change in dominant expression and a potential difference in developmental function is present in Arabidopsis *BPM2* splice variants.

However, the continuation of this research in the near future should focus on detecting as many novel splice variants as possible, with the primary goal being finding those with the MATH domain diverging from the conserved, core clade form found in splice variants characterized up to this point. This would elucidate whether alternative splicing is a mechanism which most plants could use to compensate for the small number of *MATH-BTB* genes and increase the variety of their functions to more closely resemble the multiplicity and diversity of *MATH-BTB* genes in grasses.

6. Conclusion

Searching the newest and most comprehensive *Arabidopsis thaliana* transcriptome to date, AtRTD3, revealed that the six *BPM* genes encode 56 different transcript isoforms: 15 encoded by *BPM1*, 16 by *BPM2*, 11 by *BPM3*, 6 by *BPM4*, 2 by *BPM5* and 6 by *BPM6*.

The 16 splice variants encoded by *BPM2* – *BPM2.1*, *BPM2.2*, *BPM2.3*, *BPM2.4*, *BPM2.5*, *BPM2.6*, *BPM2.7*, *BPM2.8*, *BPM2.9*, *BPM2.10*, *BPM2.11*, *BPM2.12*, *BPM2.13*, *BPM2.14*, *BPM2.15* and *BPM2.16* – differ in their combinations of TSS, TES and splice junctions. They encode nine different proteins that differ in lengths and domain compositions.

Alternative splicing of the *BPM2* gene has a role during the plant development. Expression of splice variants *BPM2.3* and *BPM2.15* varied significantly between different tissues, being more highly expressed in oval leaves of the rosette than in flower buds and open flowers, and having a lower expression level in ovules compared to somatic embryos in the induction phase. Expression levels of *BPM2.3* and *BPM2.15* as well as *BPM2.9* in somatic embryos in the maturation phase are similar to somatic embryos in the induction phase, while the expression level of *BPM2.8* appeared reduced in the maturation phase.

Existence of additional, undescribed splice variants was supported by unexpected fragments produced in standard PCR reactions with all primer pairs, and by melting curves obtained after qPCR with specific primers for *BPM2.8* and *BPM2.9*. Further exploration of these variants, especially if they will contain diverse MATH domains, would support alternative splicing as a potential mechanism for non-grass species to achieve the multiplicity and functional diversity of MATH-BTB proteins.

7. References

Bauer, N., Škiljaica, A., Malenica, N., Razdorov, G., Klasić, M., Juranić, M., Močibob, M., Sprunck, S., Dresselhaus, T. and Leljak Levanić, D. (2019) The MATH-BTB protein TaMAB2 accumulates in ubiquitin-containing foci and interacts with the translation initiation machinery in *Arabidopsis*. *Frontiers in Plant Science*. 10, 1469. doi:10.3389/fpls.2019.01469.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *genesis*. 53 (8), 474–485. doi:10.1002/dvg.22877.

Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. and Hochreiter, S. (2015) msa: an R package for multiple sequence alignment. *Bioinformatics*. 31 (24), 3997–3999. doi:10.1093/bioinformatics/btv494.

Charif, D. and Lobry, J.R. (2007) SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: U. Bastolla, M. Porto, H.E. Roman, and M. Vendruscolo (eds.). *Structural Approaches to Sequence Evolution*. Berlin, Heidelberg, Springer Berlin Heidelberg. pp. 207–232. doi:10.1007/978-3-540-35306-5_10.

Chen, Y., Chen, L., Lun, A.T.L., Baldoni, P.L. and Smyth, G.K. (2024) *edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets*. doi:10.1101/2024.01.21.576131.

Cheng, C., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*. 89 (4), 789–804. doi:10.1111/tpj.13415.

Demir, T. (2021) *Optimization of Arabidopsis thaliana (L.) Heynh. zygotic embryo isolation for subsequent use in quantitative gene expression analysis*. Master thesis. University of Zagreb, Faculty of Science, Department of Biology.

Entizne, J.C., Guo, W., Calixto, C.P.G., Spensley, M., Tzioutziou, N., Zhang, R. and Brown, J.W.S. (2020) *TranSuite: a software suite for accurate translation and characterization of transcripts*. doi:10.1101/2020.12.15.422989.

Fitch, W.M. (1966) An improved method of testing for evolutionary homology. *Journal of Molecular Biology*. 16 (1), 9–16. doi:10.1016/S0022-2836(66)80258-9.

Frlin, M. (2024) *Differential expression of arabidopsis BPM2 gene splicing variants in response to temperature stress*. Master thesis. Zagreb, University of Zagreb, Faculty of Science, Department of Biology.

Gingerich, D.J., Gagne, J.M., Salter, D.W., Hellmann, H., Estelle, M., Ma, L. and 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 (19), 18810–18821. doi:10.1074/jbc.M413247200.

Gingerich, D.J., Hanada, K., Shiu, S.-H. and 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*. 19 (8), 2329–2348. doi:10.1105/tpc.107.051300.

Godel-Jędrychowska, K., Kulińska-Łukaszek, K. and Kurczyńska, E. (2021) Similarities and Differences in the GFP Movement in the Zygotic and Somatic Embryos of Arabidopsis. *Frontiers in Plant Science*. 12, 649806. doi:10.3389/fpls.2021.649806.

Göhring, J., Jacak, J. and Barta, A. (2014) Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-nediated decay in *Arabidopsis*. *The Plant Cell*. 26 (2), 754–764. doi:10.1105/tpc.113.118075.

Grau-Bové, X., Ruiz-Trillo, I. and Irimia, M. (2018) Origin of exon skipping-rich transcriptomes in animals driven by evolution of gene architecture. *Genome Biology*. 19 (1), 135. doi:10.1186/s13059-018-1499-9.

Horn, M., Collingro, A., Schmitz-Esser, S., Beier, C.L., Purkhold, U., Fartmann, B., Brandt, P., Nyakatura, G.J., Droege, M., Frishman, D., Rattei, T., Mewes, H.-W. and Wagner, M. (2004) Illuminating the evolutionary history of Chlamydiae. *Science*. 304 (5671), 728–730. doi:10.1126/science.1096330.

Irimia, M. and Blencowe, B.J. (2012) Alternative splicing: decoding an expansive regulatory layer. *Current Opinion in Cell Biology*. 24 (3), 323–332. doi:10.1016/j.ceb.2012.03.005.

Ivanić, A. (2022) *Somatic embryogenesis efficiency in DNA methylation mutants of Arabidopsis thaliana L.* Master thesis. Zagreb, University of Zagreb, Faculty of Science, Department of Biology.

Jagić, M., Vuk, T., Škiljaica, A., Markulin, L., Vičić Bočkor, V., Tokić, M., Miškec, K., Razdorov, G., Habazin, S., Šoštar, M., Weber, I., Bauer, N. and Leljak Levanić, D. (2022) BPM1 regulates RdDM-mediated DNA methylation via a cullin 3 independent mechanism. *Plant Cell Reports*. 41 (11), 2139–2157. doi:10.1007/s00299-022-02911-9.

Juranić, M. and Dresselhaus, T. (2014) Phylogenetic analysis of the expansion of the *MATH-BTB* gene family in the grasses. *Plant Signaling and Behavior*. 9 (4), e28242. doi:10.4161/psb.28242.

Juranić, M., Srilunchang, K., Krohn, N.G., Leljak-Levanić, D., Sprunck, S. and Dresselhaus, T. (2012) Germline-specific MATH-BTB substrate adaptor MAB1 regulates spindle length and nuclei identity in maize. *The Plant Cell*. 24 (12), 4974–4991. doi:10.1105/tpc.112.107169.

Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H.Q., Barta, A. and Brown, J.W.S. (2012) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Research*. 40 (6), 2454–2469. doi:10.1093/nar/gkr932.

Lawrence, M., Gentleman, R. and Carey, V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*. 25 (14), 1841–1842. doi:10.1093/bioinformatics/btp328.

Lechner, E., Leonhardt, N., Eisler, H., Parmentier, Y., Alioua, M., Jacquet, H., Leung, J. and Genschik, P. (2011) MATH/BTB CRL3 receptors target the homeodomain-leucine zipper ATHB6 to modulate abscisic acid signaling. *Developmental Cell*. 21 (6), 1116–1128. doi:10.1016/j.devcel.2011.10.018.

Leljak Levanić, D., Horvat, T., Martinčić, J. and Bauer, N. (2012) A novel bipartite nuclear localization signal guides BPM1 protein to nucleolus suggesting its cullin3 independent function. *PLoS ONE*. 7 (12), e51184. doi:10.1371/journal.pone.0051184.

Leljak-Levanić, D., Juranić, M. and 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 (3), 267–285. doi:10.1007/s00497-013-0229-4.

Lespinet, O., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Research*. 12 (7), 1048–1059. doi:10.1101/gr.174302.

Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences*. 100 (1), 189–192. doi:10.1073/pnas.0136770100.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 34 (18), 3094–3100. doi:10.1093/bioinformatics/bty191.

Li, H., Ruan, J. and Durbin, R. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*. 18 (11), 1851–1858. doi:10.1101/gr.078212.108.

Liu, Y., Zhang, M., Wang, R., Li, B., Jiang, Y., Sun, M., Chang, Y. and Wu, J. (2022) Comparison of structural variants detected by PacBio-CLR and ONT sequencing in pear. *BMC Genomics*. 23 (1), 830. doi:10.1186/s12864-022-09074-7.

Liyanage, K., Samarakoon, H., Parameswaran, S. and Gamaarachchi, H. (2023) Efficient end-to-end long-read sequence mapping using minimap2-fpga integrated with hardware accelerated chaining. *Scientific Reports*. 13 (1), 20174. doi:10.1038/s41598-023-47354-8.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15 (12), 550. doi:10.1186/s13059-014-0550-8.

Marchler-Bauer, A. and Bryant, S.H. (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*. 32 (Web Server), W327–W331. doi:10.1093/nar/gkh454.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*. 39 (Database), D225–D229. doi:10.1093/nar/gkq1189.

Marín, I. (2015) Origin and diversification of meprin proteases. *PLOS ONE*. 10 (8), e0135924. doi:10.1371/journal.pone.0135924.

Martín, G., Márquez, Y., Mantica, F., Duque, P. and Irimia, M. (2021) Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biology*. 22 (1), 35. doi:10.1186/s13059-020-02258-y.

Minor, D.L., Lin, Y.-F., Mobley, B.C., Avelar, A., Jan, Y.N., Jan, L.Y. and Berger, J.M. (2000) The polar T1 interface is linked to conformational changes that open the voltage-gated potassium channel. *Cell*. 102 (5), 657–670. doi:10.1016/S0092-8674(00)00088-X.

Miškec, K. (2019) *The role of conserved BPM1 protein domains MATH, BTB and SPOP in interaction with DMS3, RDM1 and HB6 proteins*. Master Thesis. Zagreb, University of Zagreb, Faculty of Science, Department of Biology.

Oelmüller, R., Peškan-Berghöfer, T., Shahollari, B., Trebicka, A., Sherameti, I. and Varma, A. (2005) MATH domain proteins represent a novel protein family in *Arabidopsis thaliana*, and at least one member is modified in roots during the course of a plant–microbe interaction. *Physiologia Plantarum*. 124 (2), 152–166. doi:10.1111/j.1399-3054.2005.00505.x.

Pagès, H., Aboyoun, P., Gentleman, R. and DebRoy, S. (2017) *Biostrings: efficient manipulation of biological strings*. doi:10.18129/B9.BIOC.BIOSTRINGS.

Pali, D. (2020) *Alternative splicing of BPM1, BPM2 and BPM3 genes under temperature stress conditions*. Master Thesis. Zagreb, University of Zagreb, Faculty of Science, Department of Biology.

Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 35 (3), 526–528. doi:10.1093/bioinformatics/bty633.

Pintard, L., Willis, J.H., Willems, A., Johnson, J.-L.F., Srayko, M., Kurz, T., Glaser, S., Mains, P.E., Tyers, M., Bowerman, B. and Peter, M. (2003) The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature*. 425 (6955), 311–316. doi:10.1038/nature01959.

Reddy, A.S.N., Marquez, Y., Kalyna, M. and Barta, A. (2013) Complexity of the alternative splicing landscape in plants. *The Plant Cell*. 25 (10), 3657–3683. doi:10.1105/tpc.113.117523.

Schmidt, M.R., Barcons-Simon, A., Rabuffo, C. and Siegel, T.N. (2024) Smoother: on-the-fly processing of interactome data using prefix sums. *Nucleic Acids Research*. 52 (5), e23–e23. doi:10.1093/nar/gkae008.

Scotti, M.M. and Swanson, M.S. (2016) RNA mis-splicing in disease. *Nature Reviews Genetics*. 17 (1), 19–32. doi:10.1038/nrg.2015.3.

Staiger, D. and Brown, J.W.S. (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. *The Plant Cell*. 25 (10), 3640–3656. doi:10.1105/tpc.113.113803.

Stogios, P.J., Downs, G.S., Jauhal, J.J., Nandra, S.K. and Privé, G.G. (2005) Sequence and structural analysis of BTB domain proteins. *Genome Biology*. 6 (10), R82. doi:10.1186/gb-2005-6-10-r82.

Škiljaica, A. (2022) *The role of MATH-BTB family proteins TaMAB2 and AtBPM1 in plant development and stress response*. Doctoral thesis. Zagreb, University of Zagreb, Faculty of Science, Department of Biology. https://urn.nsk.hr/urn:nbn:hr:217:744869.

Škiljaica, A., Jagić, M., Vuk, T., Leljak Levanić, D., Bauer, N. and Markulin, L. (2022) Evaluation of reference genes for RT-qPCR gene expression analysis in *Arabidopsis thaliana* exposed to elevated temperatures. *Plant Biology*. 24 (2), 367–379. doi:10.1111/plb.13382.

The Galaxy Community, Abueg, L.A.L., Afgan, E., Allart, O., Awan, A.H., et al. (2024) The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*. doi:10.1093/nar/gkae410.

Torre, D., Francoeur, N.J., Kalma, Y., Gross Carmel, I., Melo, B.S., et al. (2023) Isoform-resolved transcriptome of the human preimplantation embryo. *Nature Communications*. 14 (1), 6902. doi:10.1038/s41467-023-42558-y.

Vuk, T. (2023) *The role of BPM1 protein in de novo DNA methylation mechanism during development of Arabidopsis thaliana L.* Doctoral thesis. Zagreb, University of Zagreb, Faculty of Science, Department of Biology. https://urn.nsk.hr/urn:nbn:hr:217:895020.

Wang, J., Chitsaz, F., Derbyshire, M.K., Gonzales, N.R., Gwadz, M., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J. and Marchler-Bauer, A. (2023) The conserved domain database in 2023. *Nucleic Acids Research*. 51 (D1), D384–D388. doi:10.1093/nar/gkac1096.

Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M. and Wang, X.-Z. (2015) Mechanism of alternative splicing and its regulation. *Biomedical Reports*. 3 (2), 152–158. doi:10.3892/br.2014.407.

Weber, H., Bernhardt, A., Dieterle, M., Hano, P., Mutlu, A., Estelle, M., Genschik, P. and Hellmann, H. (2005) Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZ-MATH protein family. *Plant Physiology*. 137 (1), 83–93. doi:10.1104/pp.104.052654.

Weber, H. and Hellmann, H. (2009) *Arabidopsis thaliana* BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *The FEBS Journal*. 276 (22), 6624–6635. doi:10.1111/j.1742-4658.2009.07373.x.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 13 (1), 134. doi:10.1186/1471-2105-13-134.

Zapata, J.M., Martínez-García, V. and Lefebvre, S. (2007) Phylogeny of the TRAF/MATHdomain. In: H. Wu (ed.). *TNF Receptor Associated Factors (TRAFs)*. New York, NY, Springer New York. pp. 1–24. doi:10.1007/978-0-387-70630-6_1.

Zhang, H., Wang, Y., Hu, Z., Wu, Y., Chen, N., Zhu, Y., Yu, Y., Fan, H. and Wang, H. (2024) Zygotic splicing activation of the transcriptome is a crucial aspect of maternal-to-zygotic transition and required for the conversion from totipotency to pluripotency. *Advanced Science*. 11 (14), 2308496. doi:10.1002/advs.202308496.

Zhang, R., Calixto, C.P.G., Marquez, Y., Venhuizen, P., Tzioutziou, N.A., Guo, W., Spensley, M., Entizne, J.C., Lewandowska, D., ten Have, S., Frei dit Frey, N., Hirt, H., James, A.B., Nimmo, H.G., Barta, A., Kalyna, M. and Brown, J.W.S. (2017) A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*. 45 (9), 5061–5073. doi:10.1093/nar/gkx267.

Zhang, R., Kuo, R., Coulter, M., Calixto, C.P.G., Entizne, J.C., et al. (2022) A high-resolution single-molecule sequencing-based Arabidopsis transcriptome using novel methods of Iso-seq analysis. *Genome Biology*. 23 (1), 149. doi:10.1186/s13059-022-02711-0.

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, J.W., White, K.P. and Schulman, B.A. (2009) Structures of SPOP-substrate complexes: insights into molecular architectures of BTB-Cul3 ubiquitin ligases. *Molecular Cell*. 36 (1), 39–50. doi:10.1016/j.molcel.2009.09.022.

Ziegelbauer, J., Shan, B., Yager, D., Larabell, C., Hoffmann, B. and Tjian, R. (2001) Transcription factor miz-1 is regulated via microtubule association. *Molecular Cell*. 8 (2), 339–349. doi:10.1016/S1097-2765(01)00313-6.

Zollman, S., Godt, D., Privé, G.G., Couderc, J.L. and 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*. 91 (22), 10717–10721. doi:10.1073/pnas.91.22.10717.

Online sources

Chauhan, D.T. (2019) *What are Primer Dimers? - A Beginner's Guide*. 15 April 2019. Genetic Education. https://geneticeducation.co.in/what-are-primer-dimers-a-beginners-guide/ [Accessed: 23 June 2024].

Dynabeads® mRNA DIRECT™ Micro Kit User Guide (2012). Catalog Number 61021 Revision 004, *Invitrogen*, *Thermo Fisher Scientific*. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/DynabeadmRNADIRECTMicro_UG_Rev004_20120514.pdf [Accessed: 23 June 2024].

How to choose Normalization methods (TPM/RPKM/FPKM) for mRNA expression (n.d.). *Novogene*. https://www.novogene.com/us-en/resources/blog/how-to-choose-normalization-methods-tpm-rpkm-fpkm-for-mrna-expression/ [Accessed: 21 June 2024].

Kolde, R. (2019) *pheatmap: Pretty Heatmaps*. https://cran.r-project.org/web/packages/pheatmap/index.html. [Accessed: 15 June 2024].

MagMAX™ Plant RNA Isolation Kit User Guide (2016). version no. MAN0016311 Revision A.0, *Applied Biosystems*, *Thermo Fisher Scientific*. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0016311_MagMAX_PlantRNA_Isol_UG.pdf [Accessed: 23 June 2024].

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/. [Accessed: 15 June 2024].

Real-time Quantitative PCR (n.d.). *Creative BioMart*. https://www.creativebiomart.net/resource/principle-protocol-real-time-quantitative-pcr-367.htm [Accessed: 21 June 2024].

The Pain of Primer Dimer (n.d.). *Kilobaser*. https://kilobaser.com/post/the-pain-of-primer-dimer [Accessed: 21 June 2024].

Thompson, J., Brett, C., Neuhaus, I. and Thompson, R. (2022) *DGEobj.utils: differential gene expression (DGE) analysis utility toolkit*. https://cran.r-project.org/web/packages/DGEobj.utils/index.html.

8. Biography

I was born in the year 2000 in Zagreb, Croatia. I completed my high school education in 2019 at the V. Gymnasium in Zagreb. In my senior year of high school, I participated in the state biology competition in the knowledge category. I placed above the 90th percentile in all subjects at the state graduation (advanced level mathematics, advanced level Croatian language, advanced level English language, physics and biology) and above 99th percentile in biology.

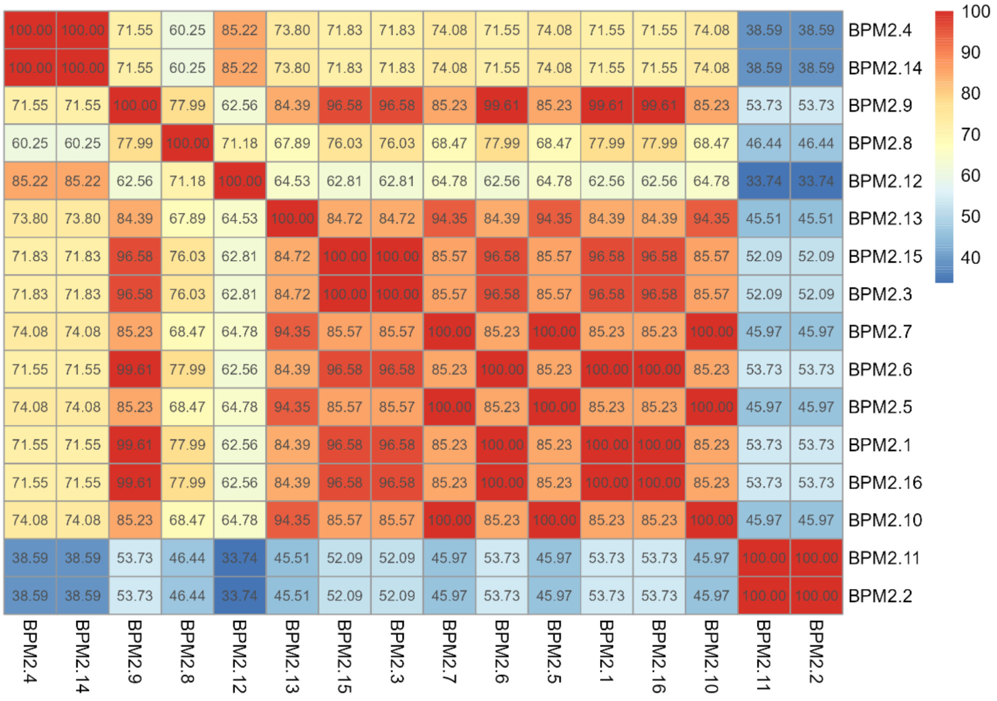
After that, in 2019, I enrolled in the undergraduate university programme of Molecular Biology at the Department of Biology, Faculty of Science, University of Zagreb. During my undergraduate studies, I volunteered at the Croatian Museum of Natural History on Night of Museums 2020. I completed an internship in the Laboratory for Molecular Genetics at the Ruđer Bošković Institute, where I was involved in the project Evolution in the dark (EVODARK) lead by Helena Bilandžija, PhD. I was one of the co-authors on two conference publications as a part of the EVODARK project (Grgić, M., Weck, R.G., Keresteš, G. and Bilandžija, H. (2022a) Cave dwelling Physella sp. as a model system for studying the loss of pigmentation. Congress of the European Society for Evolutionary Biology: Book of Abstracts. pp. 461–462; Grgić, M., Weck, R.G., Keresteš, G. and Bilandžija, H. (2022b) The pleiotropic effects of melanin loss in cave snails Physella sp. ARPHA Conference Abstracts. 5, e86885. doi:10.3897/aca.5.e86885). I received the STEM scholarship from the Croatian Government for all three years of my undergraduate study.

In 2022, I enrolled in the graduate university programme of Molecular Biology at the Department of Biology, Faculty of Science, University of Zagreb. In my first year of graduate study, I completed an internship in the plant tissue culture lab of the Division of Molecular Biology in my university with the mentor Dunja Leljak Levanić, PhD. That year I also received the STEM scholarship. In my second year of graduate study, I was a demonstrator for the plant related part of the practical classes for the class Animal and Plant Cell Culture at the Department of Biology, Faculty of Science, University of Zagreb.

9. Appendix

Table A1. Samtools flagstat mapping results. Mapping was done using Minimap2 with the preset PacBio/Oxford Nanopore read to reference mapping (-Hk19) (map-pb), K-mer size = 16 and disabled spliced alignment (mapping k16), and with the same settings but with K-mer size = 28 and minimizer window size = 19 (mapping k28).

	k16		k28	
Library	total mapped	primary mapped	total mapped reads	primary mapped
SRR23291381	1782394 (99.81%)	339785 (99.01%)	1768420 (99.81%)	339722 (98.99%)
SRR23291382	1718473 (99.86%)	340490 (99.29%)	1699864 (99.85%)	340317 (99.24%)
SRR23291390	1705847 (99.86%)	328031 (99.29%)	1688137 (99.86%)	327980 (99.28%)
SRR23291398	1841413 (99.70%)	362088 (98.49%)	1828354 (99.69%)	361982 (98.47%)
SRR23291399	1654823 (99.78%)	338111 (98.91%)	1634685 (99.77%)	338035 (98.89%)



	BPM2.4	BPM2.14	BPM2.9	BPM2.8	BPM2.12	BPM2.13	BPM2.15	BPM2.3	BPM2.7	BPM2.6	BPM2.5	BPM2.1	BPM2.16	BPM2.10	BPM2.11	BPM2.2
BPM2.4	100.00	100.00	71.55	60.25	85.22	73.80	71.83	71.83	74.08	71.55	74.08	71.55	71.55	74.08	38.59	38.59
BPM2.14	100.00	100.00	71.55	60.25	85.22	73.80	71.83	71.83	74.08	71.55	74.08	71.55	71.55	74.08	38.59	38.59
BPM2.9	71.55	71.55	100.00	77.99	62.56	84.39	96.58	96.58	85.23	99.61	85.23	99.61	99.61	85.23	53.73	53.73
BPM2.8	60.25	60.25	77.99	100.00	71.18	67.89	76.03	76.03	68.47	77.99	68.47	77.99	77.99	68.47	46.44	46.44
BPM2.12	85.22	85.22	62.56	71.18	100.00	64.53	62.81	62.81	64.78	62.56	64.78	62.56	62.56	64.78	33.74	33.74
BPM2.13	73.80	73.80	84.39	67.89	64.53	100.00	84.72	84.72	94.35	84.39	94.35	84.39	84.39	94.35	45.51	45.51
BPM2.15	71.83	71.83	96.58	76.03	62.81	84.72	100.00	100.00	85.57	96.58	85.57	96.58	96.58	85.57	52.09	52.09
BPM2.3	71.83	71.83	96.58	76.03	62.81	84.72	100.00	100.00	85.57	96.58	85.57	96.58	96.58	85.57	52.09	52.09
BPM2.7	74.08	74.08	85.23	68.47	64.78	94.35	85.57	85.57	100.00	85.23	100.00	85.23	85.23	100.00	45.97	45.97
BPM2.6	71.55	71.55	99.61	77.99	62.56	84.39	96.58	96.58	85.23	100.00	85.23	100.00	100.00	85.23	53.73	53.73
BPM2.5	74.08	74.08	85.23	68.47	64.78	94.35	85.57	85.57	100.00	85.23	100.00	85.23	85.23	100.00	45.97	45.97
BPM2.1	71.55	71.55	99.61	77.99	62.56	84.39	96.58	96.58	85.23	100.00	85.23	100.00	100.00	85.23	53.73	53.73
BPM2.16	71.55	71.55	99.61	77.99	62.56	84.39	96.58	96.58	85.23	100.00	85.23	100.00	100.00	85.23	53.73	53.73
BPM2.10	74.08	74.08	85.23	68.47	64.78	94.35	85.57	85.57	100.00	85.23	100.00	85.23	85.23	100.00	45.97	45.97
BPM2.11	38.59	38.59	53.73	46.44	33.74	45.51	52.09	52.09	45.97	53.73	45.97	53.73	53.73	45.97	100.00	100.00
BPM2.2	38.59	38.59	53.73	46.44	33.74	45.51	52.09	52.09	45.97	53.73	45.97	53.73	53.73	45.97	100.00	100.00

Figure A1. Identity matrix for proteins encoded by splice variants of *BPM2*. Alignment was performed using the *Muscle* algorithm and manually adjusted for the protein encoded by variant *BPM2.8*.

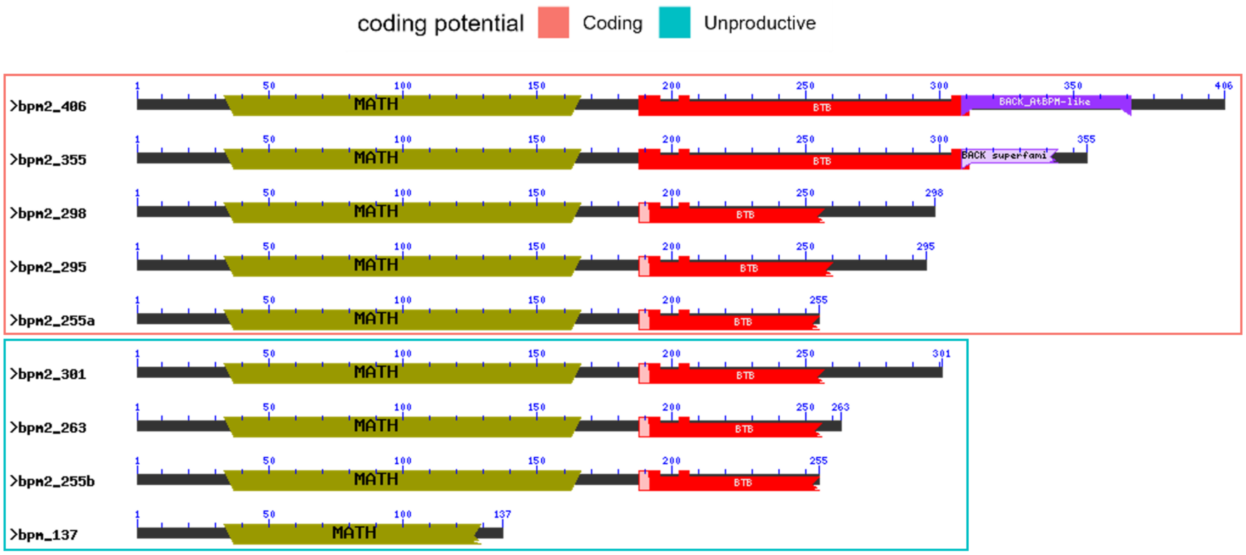


Figure A2. Conserved domains present in *BPM2* transcript variant translation products. Sorted by transcript coding potential and protein length. The names used for the proteins denote their lengths in aa. Figure made using Batch CD-Search (Wang et al., 2023; Marchler-Bauer et al., 2011; Marchler-Bauer and Bryant, 2004).

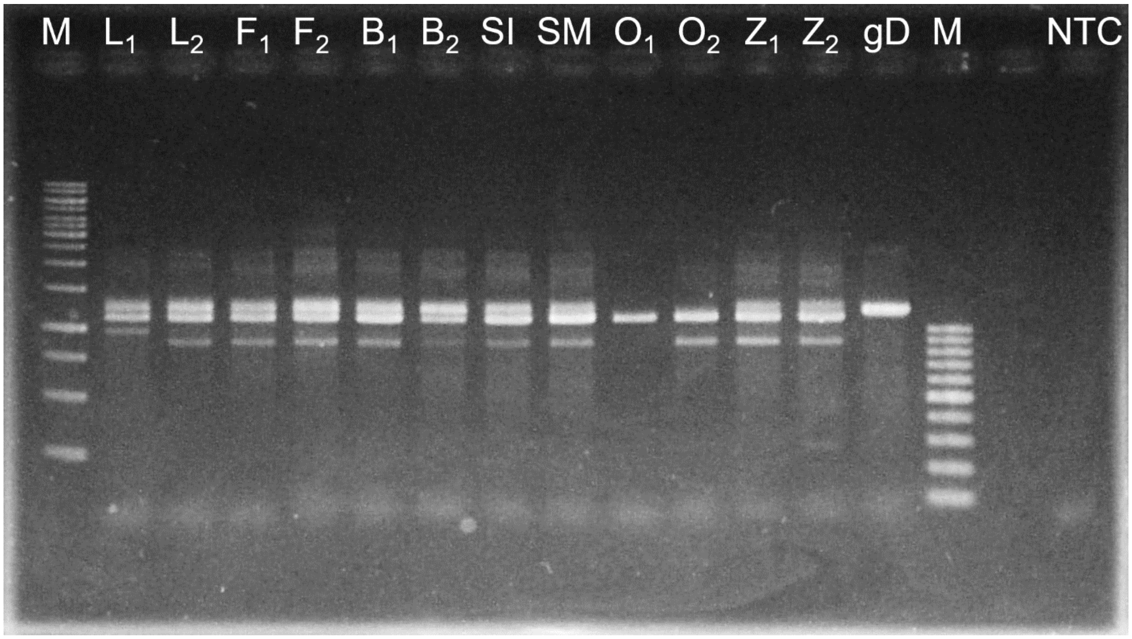


Figure A3. Agarose gel showing PCR products using primers **BPM2.3-15**_fw and BPM2_univ_rev on cDNA from oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), ovules (**O**), cotyledonary zygotic embryos (**Z**), somatic embryos in the induction phase (**SI**), somatic embryos in the maturation phase (**SM**), a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Index numbers mark different biological samples. **Ta = 61 °C, ET = 90 s**. Markers used (**M**) are GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific) on the left and GeneRuler 100 bp DNA Ladder (Thermo Fisher Scientific) on the right. Numbers denote band sizes in bp.

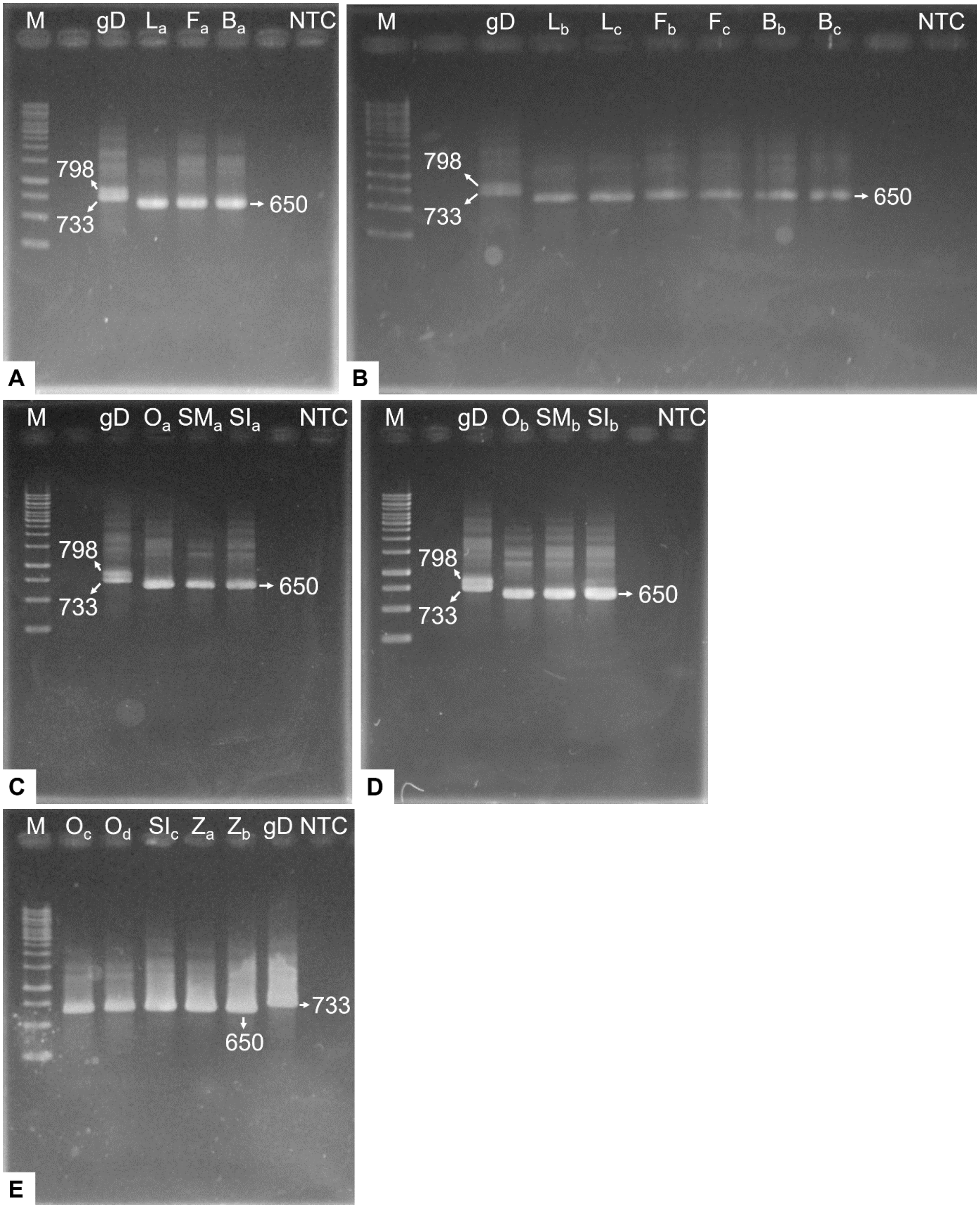

M gD L_a F_a B_a NTC
798
733
650
A
M gD L_b L_c F_b F_c B_b B_c NTC
798
733
650
B
M gD O_a SM_a SI_a NTC
798
733
650
C
M gD O_b SM_b SI_b NTC
798
733
650
D
M O_c O_d SI_c Z_a Z_b gD NTC
733
650
E

Figure A4. Agarose gel showing PCR products using primers **ACT3_fw** and **ACT3_rev** on cDNA used for standard PCR. Samples are cDNA from **A)**, **B)** oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), and **C)**, **D**, **E)** ovules (**O**), cotyledonary zygotic embryos (**Z**), somatic embryos in the induction phase (**SI**), somatic embryos in the maturation phase (**SM**). All gels have a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Index letters mark different biological samples. **Ta = 58 °C, ET = 60 s**. Marker used (**M**) is GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific). Numbers denote band sizes in bp.

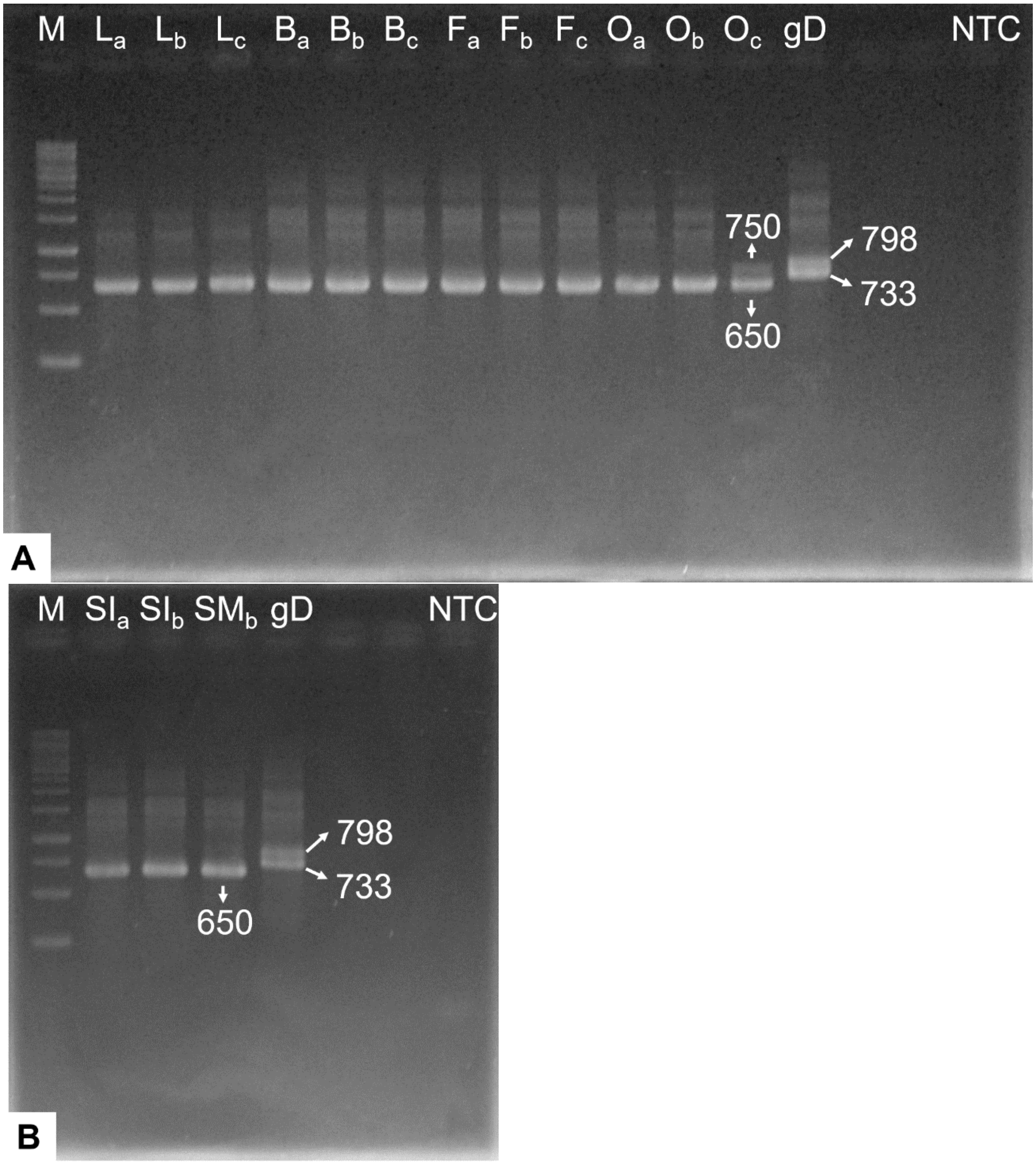

M L_a L_b L_c B_a B_b B_c F_a F_b F_c O_a O_b O_c gD NTC
750
798
733
650
A
M SI_a SI_b SM_b gD NTC
798
733
650
B

Figure A5. Agarose gel showing PCR products using primers **ACT3_fw** and **ACT3_rev** on cDNA used for qPCR. Samples are cDNA from **A)** oval rosette leaves (**L**), flower buds (**B**), open flowers (**F**), ovules (**O**), and **B)** somatic embryos in the induction (**SI**) and maturation phase (**SM**). Both gels have a genomic DNA sample from an *A. thaliana* (L.) Heynh. ecotype Col-0 plant (**gD**) and a no-template control (**NTC**). Index letters mark different biological samples. **Ta = 58 °C, ET = 60 s**. Marker used (**M**) is GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific). Numbers denote band sizes in bp. Sample O_c in gel A) shows possible gDNA contamination and wasn't used in downstream experiments.

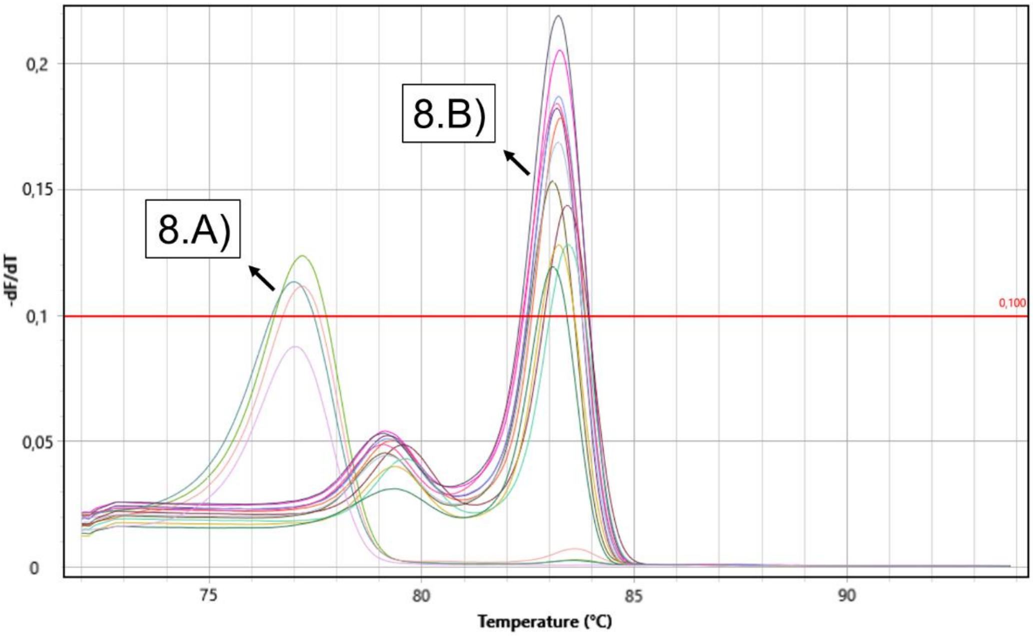


Figure A6. Melting curves of all amplicons obtained using primers BPM2.8_fw and BPM2_univ_rev for samples in which RNA was isolated using the Dynabeads™ mRNA DIRECT™ Micro Purification Kit (Invitrogen, Thermo Fisher Scientific). Created with Mic qPCR Cycler (Bio Molecular Systems) device and software.