

**Charles University Faculty of Science**

Study programme: Experimental Plant Biology



**Mgr. Božena Klodová**

**Studying the dynamics of gene expression and the role of NAC complex in male gametophyte development**

Doctoral thesis

Supervisor: RNDr. Jan Fíla, Ph.D.

Consultant: prof. RNDr. David Honys, Ph.D.

Prague, 2024

### **Author's declaration**

I hereby declare that the work presented in this doctoral thesis is entirely my own (with the exceptions stated in the presented manuscripts) and was composed independently. None of this work, nor any significant portion of the text, has been presented for any other defence proceedings. I also confirm that all sources used have been appropriately cited or recognized. Artificial intelligence (ChatGPT 4.0) was used as a grammar editor of the text.

Prague, 11. 1. 2024

Božena Klodová

### **Prohlášení autorky**

Tímto prohlašuji, že práce předložená v této doktorské práci je zcela má vlastní (s výjimkami uvedenými v předložených rukopisech) a byla sestavena samostatně. Žádná část této práce ani žádná podstatná část textu nebyla prezentována v rámci jakéhokoli jiného obhajovacího řízení. Potvrzuji také, že všechny použité zdroje byly řádně citovány nebo uznány. Umělá intelligence (ChatGPT 4.0) byla využita k editaci gramatiky textu.

Praha, 11. 1. 2024

Božena Klodová

### **Supervisor's declaration**

I hereby verify that Božena Klodová's stated contribution to this thesis accurately reflects her actual involvement.

Prague, 11. 1. 2024

Jan Fíla

### **Prohlášení školitele**

Tímto potvrzuji, že uvedený příspěvek Boženy Klodové k této práci přesně odráží její skutečnou účast.

Praha, 11. 1. 2024

Jan Fíla

## ACKNOWLEDGMENTS

First and foremost, I extend my deepest gratitude to my supervisor, Dr. Jan Fíla, for his invaluable guidance, help, teaching, and supervision throughout this journey. I would also like to express my sincere appreciation to my consultant and the Principal Investigator of our lab, Prof. David Honys. His mentorship and the myriad of opportunities he provided have been crucial in opening doors for me in the scientific world. His guidance has been a cornerstone of my development as a researcher. Special thanks go to Dr. Said Hafidh, whose assistance with experimental solutions and scientific discussions has been instrumental and his support in my transition to the adult life of a scientist has been invaluable. I am deeply grateful to my lab mate, Zahra Kahrizi, for her endless help, trust, and emotional support as well as to all my PhD colleagues, Vinod Kumar, Karel Raabe, Janto Pieters, Elnura Turataeva and Alena Náprstková for their constant support. I am very thankful to all the members of the Laboratory of Pollen Biology for their endless assistance and patience.

My gratitude also goes to Dr. Ljudmilla Timofejeva, who has always been an inspiration for my scientific career. Further, I would like to acknowledge Dr. Petra Procházková Schrupflová for posing challenges and helping me advance in my bioinformatics journey with her projects. I am deeply appreciative of Prof. David Twell and Dr. Jorg D. Becker for introducing me to the world of pollen gene expression giving me a direction and advice I will carry with me throughout my career.

I extend my thanks to Dr. Sotirios Fragkostefanakis, Dr. Uwe Bodensohn, and Dr. Mark Aurell Schoettler for providing me with the opportunity to visit their laboratories during my Ph.D. stays. These experiences have significantly enriched my scientific knowledge. Further acknowledgement goes to Dr. Kateřina Malinská and Matěj Drs for their invaluable help and advice with microscopy, and to Samuel Haluška for introducing me to the world of Marchantia biology.

My heartfelt gratitude goes to my entire family, whose unwavering faith in my capabilities have been the backbone of my journey. Finally, I would like to thank all my friends. Their role in my life extends far beyond the confines of this academic endeavour.

This work was made possible through the generous support of the Czech Ministry of Education, Youth and Sports [LTC18043, LTC20050], the Czech Science Foundation [18-02448S, 19-01723S, 21-15856S, 21-15841S, 23-07000S, 22-29717S and 24-10653S], and Mobility Plus Program with DAAD (DAAD-23-06). The work was also supported from European Regional Development Fund-Project “Centre for Experimental Plant Biology” (No. CZ.02.1.01/0.0/0.0/16\_019/0000738) and Johannes Amos Comenius Programme, OP JAK „TANGENC“ (Towards Next Generation Crops) [CZ.02.01.01/00/22\_008/0004581].

## Table of Contents

<b>Abstract</b>	<b>5</b>
<b>Abstrakt</b>	<b>6</b>
<b>Frequently used abbreviations</b>	<b>7</b>
<b>Introduction</b>	<b>8</b>
Conserved, demanding and crucial: the importance of precise gene expression regulation	8
Regulatory points of gene expression in plants	9
Gene expression regulation in male gametophyte development	17
Role of nascent polypeptide associated complex in translation	22
<b>Results:</b>	<b>24</b>
Regulatory dynamics of gene expression in the developing male gametophyte of <i>Arabidopsis</i>	24
GOLEM: distribution of Gene regulatOry eLEMents within the plant promoters	54
A decade of pollen phosphoproteomics	55
The beta subunit of nascent polypeptide associated complex plays a role in flowers and siliques development of <i>Arabidopsis thaliana</i>	60
The functional conservation of beta subunits of the nascent polypeptide associated complex in liverworts and angiosperms	75
Beta subunits of the nascent polypeptide associated complex play role in protein folding and chloroplast transport in <i>A. thaliana</i>	128
<b>General discussion</b>	<b>153</b>
Challenges of studying long non-coding RNAs in the male gametophyte	153
Translatomics filling the gap in translation dynamics	155
Dissecting the pollen gene expression in sperm cells	156
Pollen desiccation, the third peak in microgametogenesis expression	158
The online resources for the analysis of gene expression in male gametophyte	159
The role and functional conservation of the Nascent polypeptide associated complex in plants	162
<b>Conclusion</b>	<b>166</b>
<b>Bibliography</b>	<b>168</b>

## Abstract

This Ph.D. thesis presents a comprehensive study of gene expression regulation within the male gametophyte development focusing on transcriptional, translational, and post-translational levels. The research introduces an online tool, GOLEM, designed for the visualization and analysis of motif distribution within the DNA of various plant species. Utilizing RNA-seq data, GOLEM enables the study of gene expression across different tissues and developmental stages, as well as a comparative analysis across the evolution of plant lineages. Through an integrated multi-omics approach, combining transcriptomic and proteomic data, the thesis enriches the understanding of gene expression dynamics in male gametophyte development, identifying significant trends and categorizing gene families based on their expression patterns. This multifaceted dataset provides a valuable resource for future functional genomics studies. Additionally, prospects and challenges of studying gene expression regulation in male gametophyte are discussed.

Furthermore, the thesis explores the regulatory potential of the nascent polypeptide-associated complex (NAC) protein family in male gametophyte development, particularly their role in translation during pollen tube growth. Experiment results indicate that NAC $\beta$  subunit knock-down causes defects in pollen tube growth, suggesting its essential function in translation. Moreover, the expression patterns of the NAC complex align with those of translation-related genes, implying its potential involvement in the translation machinery, an association previously noted in animal systems. The research posits that the NAC complex may be part of an extensive network of chaperones involved in protein folding and sorting in *Arabidopsis thaliana*. The research extends to examine the conservation of NAC function across different plant species, evidenced by successful functional complementation experiments between angiosperms and liverworts. Overall, the thesis contributes to a deeper understanding of the regulation governing gene expression in plant reproduction and highlights the critical role of the NAC protein family in plant development.

## Abstrakt

Tato doktorská práce představuje komplexní studii regulace genové exprese během vývoje samčího gametofytu se zaměřením na transkripční, translační a post-translační úroveň. Práce dále představuje online nástroj GOLEM, který je navržen pro vizualizaci a analýzu distribuce DNA regulačních motivů v genomech různých rostlinných druhů. S využitím transkriptomických dat GOLEM umožňuje studovat genovou expresi v různých pletivech a vývojových stádiích, stejně jako komparativní analýzu napříč různými fylogenetickými skupinami rostlin. Integrací transkriptomických a proteomických dat práce zásadně zlepšuje porozumění dynamice genové exprese ve vývoji samčího gametofytu, identifikuje významné regulační trendy a kategorizuje rodiny genů na základě jejich exprese. Tento komplexní soubor dat poskytuje cenný zdroj pro budoucí genomické studie. Kromě toho jsou v práci diskutovány perspektivy a výzvy dalšího studia regulace genové exprese v samčím gametofytu.

Dále práce zkoumá regulační potenciál rodiny proteinů asociovaných s nascentním polypeptidem (NAC) ve vývoji samčího gametofytu, zejména jejich roli v translaci během růstu pylové láčky. Výsledky experimentů naznačují, že snížená exprese NAC $\beta$  podjednotky způsobuje defekty v růstu pylové láčky. Navíc exprese komplexu NAC během vývoje samčího gametofytu ukazuje stejný trend jako exprese genů souvisejících s translací. To dále naznačuje jeho potenciální roli v translaci, která byla dříve pozorována u živočišných modelů. Na základě výsledků předpokládáme, že komplex NAC může být součástí rozsáhlé sítě chaperonů zapojených do skládání a třídění proteinů do buněčných kompartmentů huseníčku rolního (*Arabidopsis thaliana*). Výzkum dále zkoumal, jak je funkce komplexu NAC konzervovaná v evolučně vzálených rostlinných druzích. Funkční komplementace mezi krytosemenným huseníčkem rolním (*Arabidopsis thaliana*) a játrovkou porostnicí mnohotvárnou (*Marchantia polymorpha*) poukazuje na zástupnou funkci NAC proteinů u těchto dvou organismů. Celkově práce přispívá k hlubšímu pochopení regulace řídicí genovou expresi v rostlinné reprodukci a zdůrazňuje klíčovou roli rodiny proteinů NAC ve vývoji rostlin.

## Frequently used abbreviations

BCP	bicellular pollen
BiFC	bimolecular fluorescence complementation
CDS	coding sequence
GFP	green fluorescent protein
lncRNA	long non-coding RNA
MPG	mature pollen grain
NAC	nascent polypeptide associated complex
PT	pollen tube
PTM	post-translational modification
RFP	red fluorescent protein
ROS	reactive oxygen species
SC	sperm cell
TCP	tricellular pollen
TE	transposable element
TF	transcription factor
UNM	uninucleate microspore
VN	vegetative nucleus
Y2H	yeast-two hybrid

# Introduction

## Conserved, demanding and crucial: the importance of precise gene expression regulation

Genes represent the core information within the cells of all living organisms. Through gene expression, the information encoded in DNA is realized, leading to the production of various functional molecules, including proteins and RNAs. The necessity for precise regulation of gene expression is underscored by the need for fine control over the timing and location of expression, which contributes to energy conservation and ensures adaptability during growth, development, and in response to environmental stimuli. This precision is especially crucial for sessile organisms like plants that face the complexities of a constantly changing environment.

Gene expression, a process shared by all living organisms, encompasses transcription, which generates various types of RNAs, and translation, which converts the information from mRNAs into functional entities such as peptides or proteins. Regulation starts from epigenetic DNA modifications, which affect the availability of DNA for transcription initiation. RNA is regulated at all levels of its synthesis, including co-transcriptional and post-transcriptional checkpoints. During these stages, the heterogenous nuclear RNA (hnRNA) is spliced and further modified to become ready to function as a non-coding RNA or mRNA that can be translated into proteins, stored, or degraded. Regulation continues at the level of translation, and after the protein is synthesized, it undergoes various post-translational processes. These involve folding, post-translational modifications, targeting to the final compartments or degradation (Fig. 1). Consequently, the synergistic coordination of these processes—precise timing and a complex regulatory network across multiple levels—enables continuous control over gene expression, exerting a significant influence on every aspect of the existence of living organisms.

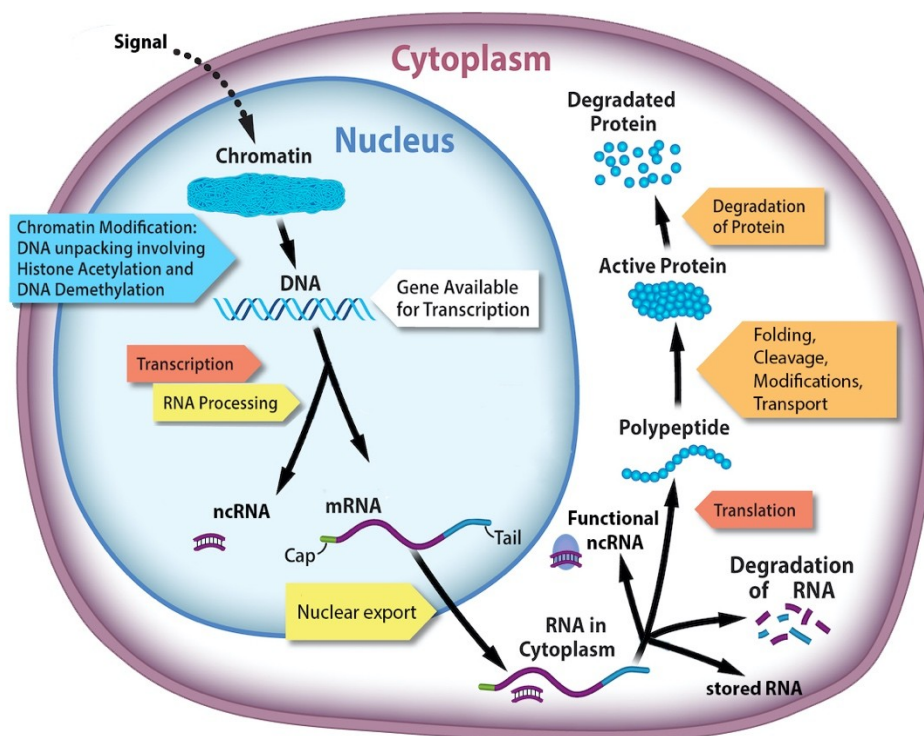


Figure 1 Overview of gene expression and its regulatory points.



## Regulatory points of gene expression in plants

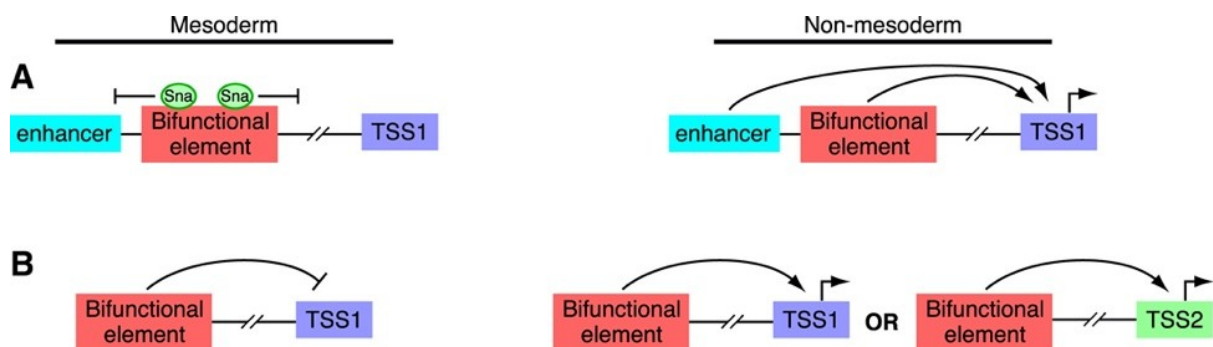
### Transcription initiation

To initiate transcription, a multiprotein complex, including RNA polymerase, assembles at the transcription start site (TSS). This process is facilitated by transcription factors (TFs), which are capable of binding to specific DNA sites, thereby enabling protein-DNA and protein-protein interactions. Thus, the proper context of DNA is essential for transcription initiation, encompassing both binding motifs for transcription factors and other components of the transcription initiation complex. Collectively, these cis-acting elements (CREs) comprise the cis-regulatory modules (CRMs), which may consist of modules necessary for transcription initiation or function as regulatory elements with the potential to modulate transcription levels in response to internal signals transmitted from the cytoplasm via specific transcription factors. This modulation determines the specific cell type, timing, and expression level of the gene in question. CREs form sequence-specific motifs that are recognized by specific TFs, the primary regulators of eukaryotic transcription. Pioneer TFs are essential for nucleosome reorganization and the recruitment of chromatin remodelling complexes, which alter chromatin structure to change DNA strand accessibility (Spitz and Furlong 2012). Unlike non-pioneer TFs, they can bind to their target binding sites (TFBSs) in the inactive chromatin state on nucleosomes (Zaret and Carroll 2011). The mechanisms by which pioneer TFs recognize DNA sequences vary among candidates, encompassing DNA methylation, histone variants, and histone modification (Bartke et al. 2010; Strahl and Allis 2000; H. Zhu, Wang, and Qian 2016). This recognition process leads to the creation of accessible chromatin regions (ACRs), which are then bound by non-pioneer TFs, culminating in CRM activation. In plants, various pioneer TFs are responsible for meristem and flower development, including SEPALLATA3 and APETALA1 (Pajoro et al. 2014).

The overall rate and activity of transcription are determined by the combined activities of various CRMs. These can be categorized as core promoters, enhancers, silencers, and multifunctional sequence elements (Schmitz, Grotewold, and Stam 2022). The core promoter, typically extending around 100 base pairs and situated near the transcription start site (TSS), facilitates transcription initiation, often through the binding of general TFs (Haberle and Stark 2018). In eukaryotes, several general binding motifs such as the TATA box, CCAAT box, and TC elements are generally required for transcription initiation (Das and Bansal 2019). However, plant promoters often lack these motifs, relying instead on a more complex and diverse array of cis-elements (Mejía-Guerra et al. 2015). Enhancers, silencers, and insulators, bound by TFs and cofactors, can respectively increase or decrease the transcription initiation rate, often in a tissue-, condition-, or development-specific manner. The location of enhancers and silencers relative to the TSS can vary significantly, sometimes exceeding 1Mbp, and their influence on the target gene can be mediated by chromatin interactions (Pang and Snyder 2020). TFs binding to these sites can function cooperatively or additively, and a single gene may be influenced by multiple enhancers and silencers, which can act complementarily, antagonistically, or redundantly (Schmitz, Grotewold, and Stam 2022). The effects of enhancers and silencers are not absolute and can be modulated by factors such as tissue specificity, leading to variable impacts on transcription initiation determined by the presence of various TFs. For instance, one study on fruit fly *D. melanogaster* reported that hundreds of studied silencers could also act as enhancers in a different cellular context (Figure 2) (Gisselbrecht et al. 2020). Unfortunately, such study is

yet to be conducted in plants. Consequently, these elements can be categorized as multifunctional sequence elements, with a potential to serve as both enhancers and silencers depending on the other factors and circumstances (Ngan et al. 2020; Ozadam et al. 2020). Such example is in the Jasmonic acid transcriptional response which is mediated by the master TF MYC2. The ME2 enhancer, situated in the intergenic area of the MYC2 locus, exhibits bifunctional regulatory effects on the transcription of MYC2. It acts as a positive regulator of MYC2 expression during the immediate jasmonic acid (JA) response, while it assumes a repressive role in regulating MYC2 expression if the JA response is prolonged (H. Wang et al. 2019).

The Plant Transcription Factor Database v5.0 (Jin et al. 2017) lists approximately 2,300 TFs described in the Arabidopsis genome, which account for around 8% of all protein-coding genes (as detailed in the Araport11 assembly (Cheng et al. 2017)). These TFs recognize TFBSs, small DNA sequences typically 6 to 12 bases in length, which may be represented over a thousand times within an organism's genome. Additionally, the binding specificity of eukaryotic TFs is often nonspecific and requires additional mechanisms such as clustering of binding sites or TF oligomerization to enhance specificity (Podlaha and Zhang 2009). Currently, Arabidopsis TFs are classified into 58 families, predominantly based on their DNA-binding motifs or conserved dimerization domains.



**Figure 2** Model of regulation of the bifunctional element by the Snail TF (*Sna*) in fruitfly *D. melanogaster*. A. *Sna* binds the element and causes short-distance repression on near-by enhancers, if present. B. In the absence of *Sna*, the bifunctional element acts as a silencer or enhancer depending on the cellular context. (Modified from: Gisselbrecht et al. 2020).

The binding capacity of transcription factors is also affected by the structural and chemical states of the DNA. For example, the configuration of chromatin and the organization of nucleosomes can impact the accessibility of DNA binding sequences for enhancers and transcription factors. Moreover, the methylation status of DNA sequences can either promote or inhibit the transcription of genes. Consequently, successful transcription relies on a combination of multiple factors: the conformational state of DNA, the accessibility of silencers and enhancers, the availability of transcription machinery, the composition of the DNA sequence, including the presence of TFBSs, and the presence and accessibility of the transcription factors themselves, which may include protein-protein interactions with additional enhancers or silencers.

Transcription factor expression is tightly regulated and can be significantly influenced by environmental cues, developmental stages, or tissue-specificity in plants. Comprehending the interplay between transcription dynamics and the presence of specific motifs within

particular gene groups expressed in specific conditions sheds light on their functional categorisation. Such understanding can also reveal phylogenetic relationships and the evolutionary context of protein functions in plant evolution. To explore this potential, we have developed an online motif analysis tool called GOLEM (Gene regulatOry eLEMents, unpublished). This tool is designed to analyse the distribution of motifs within promoter regions across genes categorized by their expression during different developmental stages, covering both sporophytic and male gametophytic tissues. The inclusion of multiple plant species in GOLEM further allows to track the presence and specificity of these motifs throughout plant evolution.

### **Co-transcriptional and post-transcriptional regulation**

Upon completion of transcription, RNA undergoes processing to become functional non-coding RNA (i.e., short non-coding RNA including siRNA and miRNA, long non-coding RNA (lncRNA), transfer RNA (tRNA), ribosomal RNA (rRNA)), mRNA with protein-coding potential, or is subjected to degradation. Post-transcriptional regulation is highly conserved across animals; however, plants have evolved several distinctive regulatory mechanisms (Roy and Arnim 2013).

Nascent RNA is processed both co-transcriptionally and post-transcriptionally. The initial co-transcriptional step involves the addition of a methyl-7-guanosine (m7G) cap to the 5' end of the hnRNA. Post-transcriptionally, the polyadenylation of the 3' end occurs (Eaton and West 2020). These processes are characteristic of genes transcribed by RNA polymerase II (Pol-II), which transcribes both protein-coding and certain non-coding RNAs. The eukaryotic genome encodes additional polymerases: the highly conserved polymerase I, which transcribes ribosomal RNAs (except the 5S rRNA, Russell and Zomerdijk 2006), and polymerase III, which is responsible for the transcription of t-RNA and 5S rRNA (Dieci et al. 2013). In plants, Pol IV produces most small non-coding RNAs (siRNAs), whereas Pol V generates the majority of long non-coding RNAs (lncRNAs) and is involved in the AGO4 pathway of *de novo* DNA methylation (Wierzbicki, Haag, and Pikaard 2008). RNAs transcribed by these polymerases usually undergo different processing from the typical Pol-II products. These variations in RNA processing based on polymerase type are reviewed in Yang et al. (2023).

Occurring mostly co-transcriptionally, occasionally post-transcriptionally, RNA splicing is another key regulatory point for functional RNAs. The splicing plays a crucial role in plant biology, with over 90% of pre-mRNAs undergoing splicing to produce a mature mRNA form. Notably, up to 60% of these mRNAs can generate multiple isoforms. The splicing process, highly conserved across eukaryotes, is facilitated by an ensemble of ribonucleoproteins (snRNPs) and RNA binding proteins (RBPs). These molecules orchestrate the recognition and removal of introns via sequential phosphodiester transfer reactions (Meyer, Koester, and Staiger 2015). Moreover, splicing factors have been observed to influence translational processes (Moore and Proudfoot 2009). Specifically, the association of splicing factors and other nuclear-acquired proteins with newly spliced mRNA favours its selection by polysomes. Consequently, these new mRNAs are preferentially translated over older transcripts that have shed these associated components (Moore and Proudfoot 2009).

Alternative splicing (AS) serves as a vital mechanism for modulating gene expression and enhancing cellular transcriptome and proteome complexity without an increase in gene number. This process involves non-canonical mRNA splicing, facilitating the creation of a diverse array of transcripts from a singular genetic sequence. External stimuli can affect splicing patterns, underscoring its adaptability to environmental changes. For instance, this adaptability is crucial for flowering regulation, as evidenced by the temperature dependent AS of the Flowering locus M (Balasubramanian et al. 2006; Jiao and Meyerowitz 2010). Furthermore, the non-coding RNA COOLAIR, which significantly impacts the expression of Flowering locus C (FLC), is regulated by AS, with class I variants exerting a strong repressive effect on FLC, while class II variants are associated with high FLC expression (Marquardt et al. 2014; Swiezewski et al. 2009; Z. W. Wang et al. 2014). The influence of AS extends to a myriad of other developmental and stress response regulations, both biotic and abiotic, as reviewed by Shang, Cao, and Ma (2017).

RNA binding proteins (RBPs) are pivotal in post-transcriptional regulation and RNA modification. The original classification of RBPs is based on their RNA-binding domains (RBDs), which are conserved motifs in the DNA sequence. Some of the conserved eucaryotic RBDs include the RNA recognition motif (RRM), K homology (KH) motif, or zinc finger (ZF) motif. The Arabidopsis genome encodes additional RBPs not found in mammals, suggesting a higher level and more complex regulation of gene expression in plants, likely due to their sessile nature and the need for fine-tuned regulatory processes. The evolution of additional RBPs in plants allows their specialization. In Arabidopsis, highly abundant groups of RBPs include WD40s, kinases, RNases, pentatricopeptide repeats (PPRs), and the ZnF-CCHC and ZnF-C3H families. However, many more RBDs with non-canonical domains possibly remain to be discovered.

RBPs function through several molecular mechanisms (Prall, Sharma, and Gregory 2019). For instance, RNA chaperones, categorized as RBPs, assist in RNA folding. RNA is a dynamic molecule capable of adopting various tertiary structures. Chaperones aid RNAs in achieving thermodynamically stable structures, which may be essential for optimal function, stability or for enabling further RBP binding (Herschlag 1995; Rajkowitsch et al. 2007; Woodson 2010). The interaction between RNA and chaperones is typically transient and facilitated by electrostatic forces (Mayer et al. 2007), with RNA folding often being ATP-independent. mRNA molecules may misfold or refold in response to environmental changes and cellular signals (Ivanyi-Nagy et al. 2008; Tompa and Kovacs 2010). For example, the UNIVERSAL STRESS PROTEIN (AtUSP) is implicated in cold tolerance, stabilizing misfolded RNAs during cold stress (Melencion et al. 2017). Various other RBPs, including cold-shock domain proteins, glycine-rich proteins, and RNA helicases, have been identified as RNA chaperones active under stress conditions (Ganie 2020; Kim et al. 2013; Kwak et al. 2011; Nidumukkala and Khareedu 2019).

RBPs also play role in RNA editing and modification. RNA editing alters the nucleotide sequence of RNA molecules from their original DNA templates. A common form of RNA editing involves the deamination of cytosines to uracils, facilitated by RBPs such as DYW domain containing PPR proteins (Zehrmann et al. 2009). RNA editing adds a layer of specificity and serves as a potential corrective mechanism for DNA damage, for instance caused by UV radiation, and is crucial for organelle development and function (Hao et al.

2021). RNA modifications involve chemical alterations, for instance methylation of adenosine bases, performed by methyltransferases such as METTL3 in Arabidopsis (Bhat et al. 2020). RNA editing and modification collectively diversify the cell's transcriptome.

### **Translation regulation and mRNA storage**

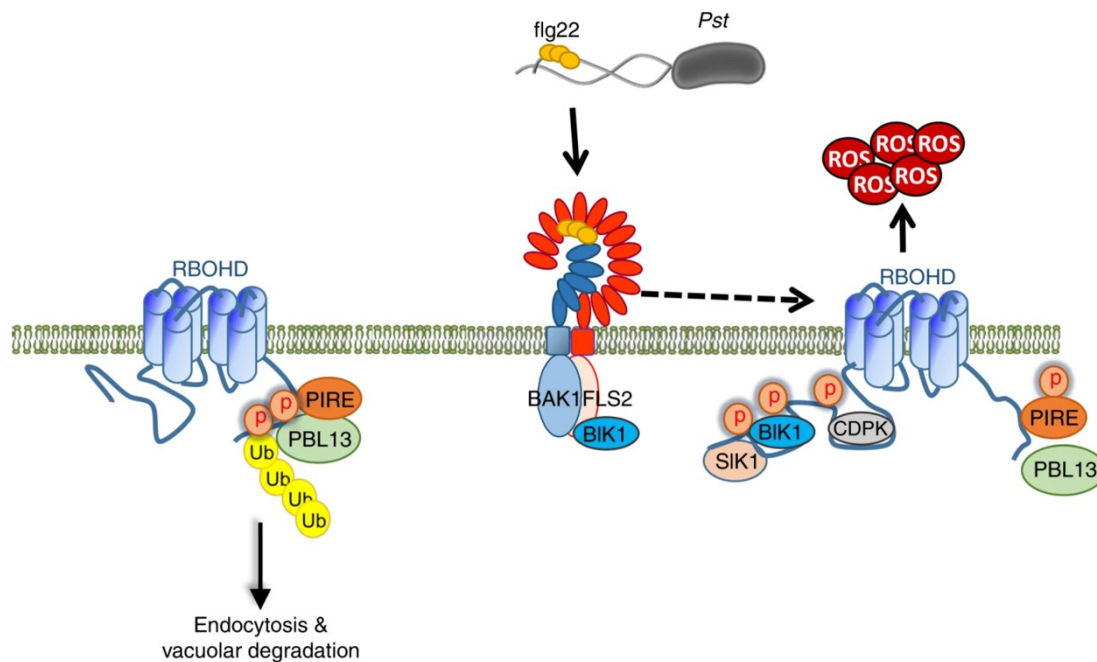
Translation is a highly conserved and energetically demanding process. In plants, the rate of translation varies according to substrate availability, metabolic status of plant cells, and environmental factors. While the total number of ribosomes in a cell remains relatively stable, the proportion of actively engaged ones in translation dynamically shifts from approximately 20 to 60%, with a peak during the daylight hours (Missra et al. 2015; Pal et al. 2013). Thus, the rate of translation impacts plant growth and development as it adapts to external stimuli and is also influenced by internal factors during developmental processes such as seed and pollen development, which require dynamic shifts in translation efficiency.

The molecular mechanisms regulating translation are numerous. The translatability of mRNAs can be affected by their structural features, such as the sequences of the 5' untranslated region (UTR) and the 3'UTR, as well as the 5' end cap and 3' polyadenylation sequence. The structure of mRNA and the binding of specific RNA-binding proteins (RBPs) can also modulate translation. Interactions between mRNA and other cytoplasmic proteins or regulatory RNAs may direct mRNA to degradation through pathways such as miRNA-targeted degradation or sequestration into stress granules or processing bodies (P-bodies). Translation machinery can also utilize internal ribosomal entry sites (IRES) to recruit the 40S ribosome subunit, independent of the start codon, 5' cap, and 3' polyadenylation (Dinkova et al. 2005; Miller, Wang, and Treder 2007). Upstream open reading frames (uORFs) in untranslated regions can also regulate translation, particularly in response to environmental conditions, by diverting the machinery from the main mRNA ORF and influencing the synthesis of short peptides (Bazin et al. 2017; Kurihara et al. 2018). Additionally, the sequence context surrounding the start codon is crucial in determining the frequency of mRNA translation initiation. Consensus contexts differ in dicots and monocots and they can influence the rate of translation or cause leaky scanning of the start codon (Gupta et al. 2016; Kozak 1987, 2002). Furthermore, covalent RNA modifications, such as m<sup>6</sup>A, m<sup>5</sup>C, or m<sup>7</sup>G, can alter mRNA structure, RBP binding sites, and the translation initiation complex, thus affecting translation rates (David et al. 2017; Scutenaire et al. 2018; H. Zhang et al. 2019).

Translation is also directly regulated by various factors, including the modulation of initiation and elongation factors in conjunction with accessory proteins. The activity of eIF4E and its binding proteins, for example, is regulated through complex phosphorylation cascades (Sesma, Castresana, and Castellano 2017; Waskiewicz et al. 1997). Phosphorylation of eIF2 $\alpha$  by GCN2 kinase following in stress conditions represents another modulatory pathway (Lageix et al. 2008; Sormani et al. 2011; Zhang et al. 2008). mRNA can be also stored in special cytoplasmic structures. One of those are stress granules, which sequester mRNAs during stress conditions, such as heat or hypoxia, releasing them again after conditions normalize (Nguyen et al. 2016; Sorenson and Bailey-Serres 2014). This temporal mRNA sequestration is also significant in developmental processes like pollen maturation or seed development, as evidenced by the presence of specific inactive ribosome mRNA

complexes called monosomes in EDTA/puromycin-resistant particles (EPP) (Hafidh et al. 2018). The role of this mRNA storage in pollen will be discussed later.

Nonsense-mediated decay (NMD) is a well-characterized translation quality control mechanism in plants, targeting mRNAs with premature stop codons to prevent the production of potentially malfunctioning proteins. It also plays a crucial role in the regulation of alternative splicing, as some isoforms are selectively degraded by NMD at specific times. NMD components are localized to P-bodies, which are cytoplasmic mRNP granules containing decapping enzymes and ribonucleases like ARGONAUTE1 for mRNA degradation (Maldonado-Bonilla 2014).



**Figure 3** Model of regulation of RBOHD by phosphorylation and polyubiquitination. The C-terminus of RBOHD undergoes phosphorylation and ubiquitination by PBL13 and PIRE, respectively, when the plant is not under infection, leading to decreased stability of RBOHD via its degradation in the vacuole. The recognition of flg22 activates RBOHD through phosphorylation at the N-terminal, triggering a release of reactive oxygen species (ROS) into the apoplast and initiating ROS-dependent defence mechanisms. Additionally, PIRE also undergoes phosphorylation during immune reaction activation. (modified from Lee et al. 2020).

### Post-translational modifications

An additional layer regulating protein function is provided by post-translational modifications (PTMs). PTMs refer to the addition of modifying groups to one or more amino acids, including for example acetyl, glycosyl, methyl, or phosphoryl groups, leading to altered protein properties. These modifications can be either reversible or irreversible, Reversible PTMs include acetylation, phosphorylation, glycosylation, and ubiquitination, which can be reversed by cellular enzymes like deacetylases, phosphatases, glycosidases, and deubiquitinating enzymes. These modifications act as regulatory switches within the cell. Irreversible modifications include proteolytic cleavage of part of the polypeptide chain and some non-enzymatic modifications like certain types of glycations, which can accumulate under conditions of metabolic imbalance or oxidative stress and may be a cause of cardiovascular diseases in human (Wu, Jankowski, and Jankowski 2022). In plants, the

role of glycation has not been extensively studied, but was reported to play role in some stress responses (Rabbani, Al-Motawa, and Thornalley 2020). Overall, the PTMs are critical for regulating protein function, stability, localization, and interaction with other molecules. PTMs can modify enzyme activities, alter signalling pathways, and govern the lifecycle of proteins. Consequently, they contribute to a more complex and dynamic proteome, introducing an additional level of regulation essential for plant growth, development, and stress responses. Phosphorylation is one of the most prevalent PTMs in plants, characterized by a balance in the activities of various kinases and phosphatases (Schulze 2010). Approximately 5% of the Arabidopsis genome encodes kinases, indicating a much higher complexity than in mammals (Zulawski et al. 2014). Phosphorylation governs numerous processes, including for instance plant defence, RNA metabolism, carbon metabolism, cold acclimation, and the regulation of Light Harvesting Complex II by light stimuli (reviewed in Arsova, Watt, and Usadel 2018).

Ubiquitination, another PTM mechanism, involves the addition of a small, 76-amino-acid peptide. The cascade starts with the activation of ubiquitin, a small protein, by an E1 ubiquitin-activating enzyme. This activated ubiquitin is then transferred to an E2 ubiquitin-conjugating enzyme. Finally, an E3 ubiquitin ligase facilitates the transfer of ubiquitin from the E2 enzyme to a lysine residue on the target protein. The specificity of this process is largely determined by the E3 ligase, as there are many different E3 ligases that recognize various substrates. It is estimated that around 7% of Arabidopsis coding genes are implicated in ubiquitination, which affects a broad spectrum of processes including plant growth and development, hormone signalling, photomorphogenesis, and stress responses (reviewed in Friso and Van Wijk 2015). Next example are small ubiquitin-like modifiers (SUMOs), ubiquitin-like polypeptides. Sumoylation occurs through the covalent attachment of the C-terminal Gly of SUMO to accessible Lys residues on target proteins. For example, in Arabidopsis, SUMO1/2 plays role in activating temperature acclimation by regulating the heat stress factor 1 family, playing role in long-term warm stress acclimation (Hammoudi et al. 2021). One protein can also undergo multiple PTMs and the PTMs can form a consequent signalling cascade to enhance its effects. For instance, phosphorylation can work as a signal for polyubiquitination and protein degradation. Such example is plant NADPH oxidase RBOHD, which is required for reactive oxygen species production during plant immunity reaction. RBOHD can become phosphorylated at the C-terminus which then enhances level of RBOHD polyubiquitination by PIRE E3 ubiquitin ligase and its degradation in the resting state, when no infection signal is present (fig. 3, Lee et al. 2020).

Overall, more than 400 types of PTMs have been described in the Swiss-Prot database (Khoury, Baliban, and Floudas 2011). A comprehensive understanding of their regulatory roles in plants still requires an extensive functional study to unravel the complexity of this regulatory network.

### **Regulation via non-coding RNAs**

In the latest annotated version of the Arabidopsis thaliana genome, Araport11, there are 5,178 non-coding RNAs (ncRNAs) identified. With the continuous advancement and refinement of detection methods, this number is likely to grow. ncRNAs can be categorized into structural and regulatory classes. Structural ncRNAs include tRNA, rRNA, snRNA, and snoRNAs and typically maintain housekeeping functions. Regulatory ncRNAs are often

classified as small ncRNAs, long ncRNAs (lncRNAs), circular ncRNAs, and derived ncRNAs (Bhogireddy et al. 2021).

The modes of action through which ncRNAs can influence gene expression are diverse. miRNAs and siRNAs can target specific mRNAs for degradation. The small RNA, within a protein complex, binds to complementary sites on the target RNA, leading to degradation by the RISC-AGO1 complex (L. Peters and Meister 2007). They can also modulate translation through a mechanism of translational repression (Aukerman and Sakai 2003). This repression typically occurs when the binding interaction between the mRNA and target is weak, with the extent of the effect being further influenced by the number of bound miRNAs; a higher number of bound miRNAs generally results in more efficient translational repression (Cuellar and McManus 2005; Doench and Sharp 2004; Zeng, Wagner, and Cullen 2002). In cases of very weak complementarity, miRNAs can accelerate the decay of the target mRNA by mediating the removal of the polyadenyl tail, although this mechanism has not yet been reported in plants (Wu, Fan, and Belasco 2006). To date, approximately 180 miRNAs have been identified in the Arabidopsis genome, resulting in 80 distinct miRNA families (Ambros et al. 2003; Meyers et al. 2008). Functional analyses have linked these to roles in various processes, including meristem initiation, leaf development, vascular development, flowering, and the development of stress tolerance (Dong, Hu, and Zhang 2022; Ruiz-Ferrer and Voinnet 2009).

siRNAs share several functional mechanisms with miRNAs. Ta-siRNAs can regulate gene expression at both the transcriptional and post-transcriptional levels, predominantly through mRNA cleavage driven by complementarity (Peragine et al. 2004; Vazquez et al. 2004). Ra-siRNAs can modulate transcription by altering the epigenetic status of DNA, either through DNA methylation or by H3K9 histone methylation, resulting in the silencing of specific DNA regions (Hamilton et al. 2002; Zilberman, Cao, and Jacobsen 2003). One mechanism, RNA-directed DNA methylation (RdDM), is vital for silencing transposable elements and de novo DNA methylation and has also been implicated in abiotic stress resistance in plants (Chinnusamy and Zhu 2009).

lncRNAs are characterized as transcripts with no coding potential that are typically longer than 200 nucleotides (Liu et al. 2012; Wang and Chekanova 2017). They can be further classified based on their relationship to protein-coding transcripts, including intronic ncRNAs (incRNAs) derived from introns, long intergenic ncRNAs (lincRNAs) originating from intergenic regions, and natural antisense transcripts (NATs) transcribed from the antisense strand of genes (Budak, Kaya, and Cagirici 2020). lncRNAs can influence gene expression at multiple levels through various mechanisms (Palos et al. 2023). They can modulate chromatin accessibility to transcription factors through chromatin remodelling, achieved by recruiting epigenetic modifiers (Csorba et al. 2014; Ying Wang et al. 2018). Some lncRNAs can alter chromatin topology by inducing the formation of chromatin loops (Ariel et al. 2014). Others can directly associate with regulatory elements, facilitating promoter binding or acting as scaffolds for ribonucleoproteins (Seo et al. 2017; Yuqiu Wang et al. 2014). Additionally, certain lncRNAs can affect splicing and isoform selection of target mRNAs, such as by inducing R-loop formation, a DNA-RNA hybrid that causes transcriptional pausing and promotes the production of alternatively spliced isoforms (Bardou et al. 2014; Conn et al. 2017). Post-transcriptionally, lncRNAs can participate in miRNA-mediated regulatory



pathways, including RdDM (Wierzbicki, Haag, and Pikaard 2008). lncRNAs transcribed by RNA polymerases IV and V play a significant role in RdDM and can sequester complementary miRNAs, preventing the degradation of their targets through a mechanism known as target mimicry (Franco-Zorrilla et al. 2007). lncRNAs can also regulate translation by increasing the association of polysomes with their target mRNAs, leading to higher protein production (Jabnour et al. 2013). They have been implicated in a wide range of processes, including plant flowering, developmental transition, stress response, pollen development, nodulation or photomorphogenesis (reviewed in Daniela Cordeiro, Canhoto, and Correia 2022; Datta and Paul 2019).

Recently the ncRNA world have attracted attention of many research groups delivering new studies every day, however it will take time before the full extent of their regulatory functions will be thoroughly established. In this thesis, we will bring more insight into the potential of lncRNAs in the regulation of pollen development.

## **Gene expression regulation in male gametophyte development**

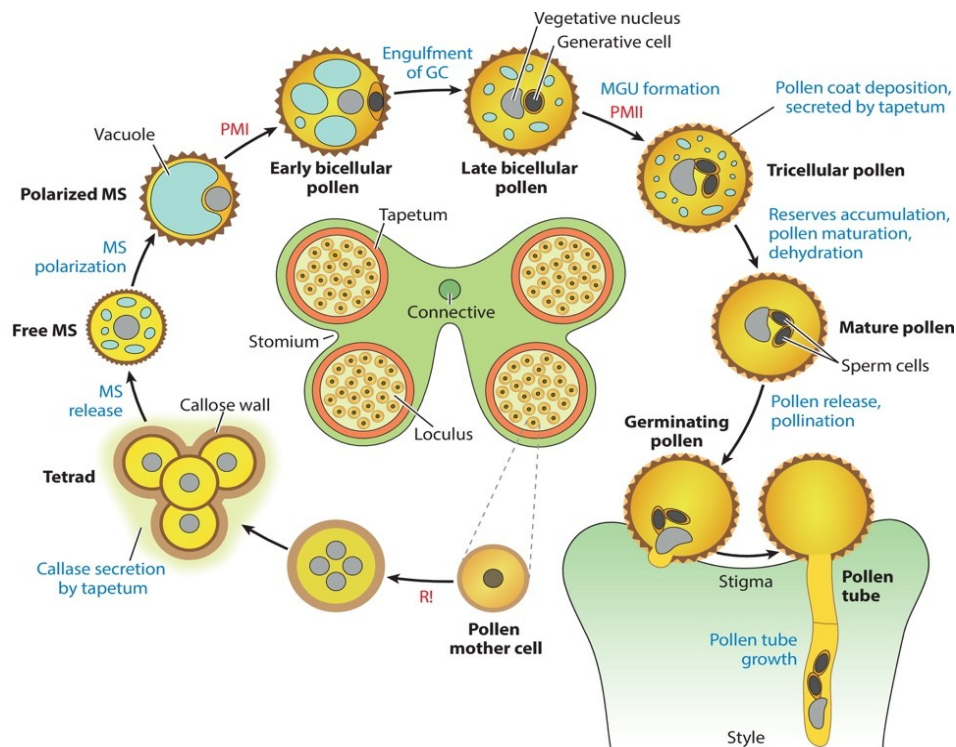
### **Sexual reproduction in plants**

In the life cycle of land plants (Embryophyta), the alternation of generations is a fundamental process involving two distinct phases: the diploid sporophyte and the haploid gametophyte. The life cycle of land plants alternates between two generations of different ploidy levels. The diploid sporophyte stage reproduces via meiosis in sporangia, producing haploid spores. These spores germinate into the haploid gametophyte. In this stage, gametes are produced through mitosis within specialized reproductive structures. The fusion of male and female gametes forms a diploid zygote, marking the beginning of the new sporophyte generation. This cycle is a key aspect of plant reproduction and development. This prevalence of stage difference between land plant groups, with bryophytes like mosses showing dominant gametophytes and vascular plants, including ferns and seed plants, possess a dominant sporophyte phase. In seed plants, the gametophyte is greatly reduced and fully dependent on the sporophyte.

The female gametophyte, formed by the embryo sac, consists of only a few cells in the polygonum-type plants (around 60% of angiosperms including *Arabidopsis thaliana*). In this group, the embryo sac includes the egg cell, 2 synergid cells, central cell, and 3 antipodal cells. The male gametophyte, pollen grain, contains two or three cells depending on the plant species. The bicellular mature pollen is shed from the anthers for instance by the model species *Nicotiana glauca* and consists of vegetative and generative. The generative cells undergo mitosis during the pollen tube growth to produce two sperm cells. The tri-cellular pollen, which is formed by the model plant *Arabidopsis thaliana*, undergoes second mitosis already during its development before shedding, including one vegetative cell and two sperm cells. After landing on stigma, the pollen germinates a pollen tube, which delivers the sperm cells through pistils to the ovule where it bursts, enabling fertilisation. The egg cell is fertilized by one of the sperm cells to form the diploid zygote, and the central cell, upon fertilization by the second sperm cells, develops into the triploid endosperm, to provide nutrients to the developing embryo. This phenomenon is termed the double-fertilisation (Hamamura, Nagahara, and Higashiyama 2012).

### **Microgametogenesis**

Microgametogenesis represents the developmental process, where the mature pollen is produced (Fig. 4). During the development, pollen assumes its unique gametophytic identity by synthesizing specific RNAs and proteins and is metabolically active in creating stores of reserves. Microgametogenesis begins when microspores, produced through reductive meiotic division of the microsporocyte (pollen mother cell), are released from a tetrad, each forming an individual uninucleate microspore (UNM) (De Storme and Geelen 2013). The UNM contains a single nucleus and a large vacuole. Following asymmetric mitotic division, known as pollen mitosis I, bicellular pollen (BCP) is formed. BCP consists of a large vegetative cell and a small generative cell. During the maturation of BCP, the sperm cell migrates into the vegetative cell, creating a complex cell-in-cell structure, completing the early stage of microgametogenesis (S. D. Russell and Jones 2015). The subsequent developmental phase is initiated with pollen mitosis II, where the sperm cell divides into two generative cells resulting in tricellular pollen (TCP) in *Arabidopsis thaliana* (Berger and Twell 2011; S. D. Russell and Jones 2015). In plants with bi-cellular pollen, the pollen mitosis II takes place only after pollen tube germination. The final step involves the maturation of TCP to produce a mature pollen grain (MPG). The mature pollen then undergoes complete desiccation, and its metabolic activity significantly decreases (Vogler, Konrad, and Sprunck



**Figure 4** Pollen development occurs in the anther, beginning with microsporogenesis, where the microsporocyte divides meiotically to create four microspores (MSs) forming a tetrad. These MSs are encased in callose, which is later broken down and the MSs are released and the microgametogenesis is started. First, each MS divides asymmetrically, forming a larger vegetative cell and a smaller generative cell. The generative cell, housed within the vegetative cell, divides again to form two sperm cells. This cell division, pollen mitosis II (PMII), can happen before or after pollen matures, leading to mature pollen that is shed either bicellular or tricellular, depending on the species. This process, and the close association of the vegetative nucleus with the sperm cells, forms the male germ unit. Upon landing on stigma, the mature pollen germinates and the growth of pollen tube and the delivery of the sperm cells is initiated. Abbreviation: R! meiosis, MS microspore, PMI pollen mitosis I, PMII pollen mitosis II, GC generative cell, MGU male germ unit. Modified from (Hafidh and Honys 2021)

2015). During this phase, protein synthesis is reduced, and mRNAs are stored as inactive ribonucleoprotein particles (RNPs) (Hafidh et al. 2018; Lin et al. 2014). In the inactive state, the pollen grains can withstand environmental conditions and survive until it is transferred to the stigma — the female receptive organ (Edlund, Swanson, and Preuss 2004; Footitt and Cohn 2001). Upon contact with a compatible stigma, pollen activation is initiated by rehydration and metabolic activity restart (Heslop-Harrison 1987). Pollen tube germination is initiated by utilizing the mRNA and protein reserves (McCormick 2004). During this process, massive translation is reinstalled, and complex regulation of pollen tube growth is initiated. The pollen tube can grow only less than millimetre, as observed for instance in *Petunia* species (Kato, Watanabe, and Hoshino 2022), however in some species, such as maize, up to 30 cm journey is required to deliver the sperm cells to the ovules, growing in rate up to 1cm per hour (Barnabas and Fridvalsky 1984). The climax of the pollen development happens upon a pollen tube burst delivering the sperm cells to the female ovules resulting in the double fertilisation.

Male gametophyte development is characterized by dynamic physiological changes throughout its maturation. Transitioning from highly metabolically and translationally active stages to a desiccated dormant state, and again later on transitioning once again to its full activity possesses a challenge and a unique model for studying the regulation of gene expression. These dynamic changes require a rapid and precise modulation of gene expression, including transcription, translation, and post-translational regulation.

The issues addressed in this thesis focus on the regulation of gene expression during the development of the male gametophyte. Consequently, the following introductory chapters delve into the aspects of gene regulatory mechanisms in male gametophyte development in greater detail. While they do not exhaustively cover all levels of regulation, they highlight the key regulatory processes that are relevant to the research findings presented in this thesis.

### **Transcription factors and their binding sites influencing male gametophyte development**

Gene expression in male gametophyte, particularly during specific phases of microgametogenesis, is highly enriched for certain mRNAs and proteins that can be categorized as pollen-specific. Comparative studies between gametophytic and sporophytic tissues have indicated that over 10% of genes are expressed uniquely in pollen (Honys and Twell 2004; David Twell, Oh, and Honys 2006). Our recent study on gene expression dynamics in developing pollen identified expression of approximately 26,000 genes across 4 microgametogenesis developmental stages, including a list of newly detected genes, which are highly expressed and potentially specific to microgametogenesis. However, functional studies to describe their molecular function are missing so far (Klodová et al. 2023).

*Arabidopsis* sperm cells exhibit a unique transcriptional profile (Borges et al. 2008). A prime example of the role of male gametophyte-specific proteins is the transcription factor DUO1, an R2R3 MYB transcription factor, which influences pollen sperm cell differentiation and establishes male gametophyte identity. DUO1, detectable after the asymmetric division of the microspore, is not present in any sporophytic tissue, indicating highly selective transcriptional regulation. DUO1 regulates its targets, which include transporters, transcription factors, signalling membrane proteins, and histones, by altering their

transcription levels. While most DUO1 targets appear to be sperm cell-specific, genes like DNA1 and DAU1 also exhibit sporophytic activity. Since DUO1 mutations do not affect their sporophytic expression, alternative regulatory mechanisms are likely to be involved in their regulation in non-gametophytic tissues. A mutation in DUO1 results in the generative cell failing to divide, leading to infertile pollen and subsequent failure of fertilization.

Some regulatory elements and promoters associated with the expression of male gametophyte-specific genes have been characterized. Pioneering studies identified the 5' flanking regions of two tomato genes, LAT52 and LAT59, which can drive pollen-specific expression in several plants, including *Arabidopsis thaliana* (D. Twell et al. 1991). These studies pinpointed small regions within the proximal promoters of LAT52 and LAT59 as responsible for specific expression, identifying three key motifs: the PB core motif (GTGA), the POLLEN1LeLA52 motif (AGAAA), and the "TCCACCATA" motif in the core LAT52 promoter in tomato (Bate and Twell 1998; D. Twell et al. 1991; D Twell et al. 1989). Other male gametophyte-specific motifs, such as the Lily Generative cell-specific (LGC1) motif GGCTGAATTT, have been identified in multiple plant species. This motif is recognized by Germline Restrictive Silencing Factors (GRSF), leading to the silencing of genes responsible for male gametophyte development in sporophytic tissues (Haerizadeh, Singh, and Bhalla 2006). A similar silencing role in sporophytic tissues was observed in *Arabidopsis* for an 88bp sequence, including the motif GAATATTCCT, in the promoter of the ACA7 gene (Hoffmann et al. 2017). However, comparisons of the expression of pollen-specific genes to those in seedlings did not reveal any significant differences in their distribution. The regulatory region of DUO1 (ROD1) is conserved among eudicots (B. Peters et al. 2017) and includes several CREs. DUO1 expression is specifically driven by the repeated DNGTGGV motif. Positive feedback regulation of DUO1 expression is facilitated through the YAACYGY motif (B. Peters et al. 2017). In rice, *in silico* methods based on expression data from Affymetrix arrays of rice sperm cells have yielded a list of putative sperm cell-specific motifs (Sharma et al. 2011). While these identified motifs require further empirical validation, omics and bioinformatics tools offer powerful approaches to identify candidate CREs.

Pollen-specific regulatory elements and transcription factors collectively fine-tune the dynamic gene expression in male gametophyte development, representing crucial aspects for studying the underlying molecular mechanisms of pollen development.

### **Role of alternative splicing and differential isoform usage in male gametophyte development**

Alternative splicing (AS) is a mechanism that expands the complexity of the transcriptome. It is often highly tissue-specific, and many AS events, resulting in novel isoform variants, may be uniquely present in pollen or restricted to certain stages of pollen development. The application of high-throughput methods such as RNA-sequencing brought the possibility to study AS on a whole transcriptome level. A pioneering study compared the dry pollen transcriptome with that of 3-week-old seedlings (Loraine et al. 2013), finding that while the majority of dominant isoforms were consistent across both seedlings and pollen, there were some genes with differential isoform preferences between the two tissues. The study also discussed methodological limitations that might contribute to the under-detection of pollen-specific isoforms (Loraine et al. 2013). Subsequent research validated 8 out of 9

differentially spliced transcripts between seedlings and pollen identified in the RNA-seq data using a semi-quantitative PCR approach (Estrada et al. 2015).

Pollen development is notably vulnerable to high temperature stress (Bokszczanin and Fragkostefanakis 2013; Giorno et al. 2013). In a study comparing two tomato cultivars, Moneymaker and Red Setter, the role of AS in stress response within pollen was investigated. The study uncovered 141 transcripts specific to heat shock regulation and 22 transcripts repressed by heat shock. Subsequent research might address the differences in AS related to heat stress tolerance between pollen and sporophytic tissues to uncover potential pollen-specific heat shock tolerance mechanisms.

The understanding of AS role in pollen development remains limited. Therefore, this thesis addresses this gap by identifying changes in AS and differential intron usage (DIU) across the four stages of microgametogenesis, highlighting its regulatory significance and presenting a list of potential candidates for functional study.

### **Translational regulation and mRNA storage in male gametophyte development**

After pollen activation, rapid pollen tube growth is initiated. It has been shown that the PT germination relies predominantly on translation, with transcriptional repression having minimal impact (Čapková-Balatková, Hrabětová, and Tupý 1980). In various species, inhibition of translation leads to a swift arrest of pollen tube elongation (Mascarenhas 1993). Regulation of translation and mRNA storage during pollen development are critical for successful fertilization. Pollen development can be divided into three distinct stages, differentiated by metabolic activity. During microgametogenesis, pollen is metabolically very active and specific RNAs and proteins are created for storage. Translation is characterized by both the high activity of polysome-translated mRNAs and the sequestering of monosome-bound mRNAs, some of which are stored immediately after transcription in the early stages of microgametogenesis (Honys et al. 2009). Monosome structures are associated with various RNA-binding proteins (RBPs), including certain translation initiation and elongation factors, ALBA proteins, and PUM proteins (Hafidh et al. 2018). After pollen maturation, the grains desiccate and enter a quiescent stage, where translation nearly ceases, and a higher ratio of monosomes prevails over active polysomes (Honys et al. 2009). Upon pollen tube germination, the rapid mobilization of reserves and re-initiation of translation are critical. Translation reactivation is marked by a high number of active polysomes and the release of stored mRNAs from EPP particles (Honys et al. 2009). Impaired or less efficient translation machinery correlates with slower pollen tube growth and, consequently, a reduced chance of successful sperm delivery to the ovules. Interestingly, the mechanisms of translational arrest and re-initiation are not exclusive to developing pollen but are also observed in seed development, which undergoes a similar transition from activity to dormancy and back to re-activity, including the presence of inactive monosomes (Bai et al. 2020).

The dynamics of mRNA translation can be inferred through a multi-omics approach. By analysing the four stages of microgametogenesis at both the transcriptome and proteome levels, we have constructed a putative map of gene clusters based on the timing of their translation. The results of such analysis are presented in this thesis.

### **Posttranslational modifications in male gametophyte development**

The cytoskeleton is essential for pollen tube growth. Actin filaments facilitate the movement of organelles within the growing tube, while microtubules are thought to be responsible for transporting sperm cells.  $\alpha$ -tubulin can undergo post-translational modifications in various plant tissues (Wang et al. 2004). Some modifications appear to be tissue-specific; for example, the tyrosinated isoform  $\alpha 6$  tubulin is pollen-specific, whereas the polyglutamylated form of  $\alpha$ -tubulin is predominantly found in leaves (Wang et al. 2004). Transglutaminases (TGases), a class of calcium-dependent enzymes, catalyse the formation of bonds between acyl acceptor glutamyl residues and amine donors, which can create cross-links among proteins. In pollen, TGase induces the formation of actin aggregates, affects enzyme activity, and alters the binding of myosins to actin filaments, thereby potentially regulating pollen tube growth (Del Duca et al. 2009).

Another significant PTM for pollen development is glycosylation. For example, O-glycosylation of extensins is important for cell-wall assembly. Loss of function of HYP O-ARABINOSYLTRANSFERASE (HPAT) responsible for O-arabinosylation of extensins results in shorter pollen tubes, and defects in fertilisation (Beuder et al. 2020; MacAlister et al. 2016). Whereas N-glycosylation was reported to play role in pollen ovule recognition. Mutations in genes encoding a uridine diphosphate (UDP)-glycosyltransferase superfamily protein (TURAN) and a dolichol kinase (EVAN), which are both essential for the initial assembly of core N-glycans, have been observed to obstruct the bursting of the pollen tube (PT) and the release of sperm cells (Lindner et al. 2015).

Phosphorylation is another key PTMs influencing pollen development. For instance, phosphorylation of transcription factor WRKY34 by MAPKs MPK3/MPK6 is pivotal at early pollen developmental stages. The absence of phosphorylation leads to diminished pollen viability, germination, and tube growth (Guan et al. 2014). RBHOD, the reactive oxygen species (ROS) producer during immunity response, was reported to be also regulated by phosphorylation during its role in anther development. ROS have been suggested to act as a signalling molecule that facilitates the programmed cell death (PCD) of tapetal tissue post-meiosis (Hu et al. 2011; Xing and Zachgo 2008). RBOHD undergoes phosphorylation during specific stages of anther development, which could amplify ROS presence in tapetal cells and encourage tapetal PCD (Z. Zhang et al. 2017).

High-throughput phosphoproteomic analyses have shed light on the phosphorylation dynamics during pollen activation in tobacco, identifying phosphorylation sites on 301 phosphoproteins across three pollen stages: mature pollen, pollen activated for 5 minutes, and pollen activated for 30 minutes (Fíla et al. 2016). These proteins fall into 13 functional categories, including those involved in transcription, translation, protein processing, and signal transduction, and they can be grouped based on phosphorylation dynamics. For instance, several initiation factors such as eIF4G, eIF4B, and eIF5B become dephosphorylated upon the pollen reactivation (Fíla et al. 2016). Another group, represented by the nascent polypeptide associated complex proteins, was identified as being strictly phosphorylated upon pollen activation. This family has been chosen for functional characterization, which is addressed in this thesis.

To compile and discuss recent advancements in high throughput phosphoproteome dynamics within male gametophyte tissues, we have reviewed the existing literature in an article. We compared findings from various plant species to identify conserved mechanisms

and analysed similarities between the phosphoproteomes of pollen tubes and root tips, suggesting potential common regulatory elements for directional tip growth.

### **Role of nascent polypeptide associated complex in translation**

The nascent polypeptide associated complex (NAC) family in *Arabidopsis thaliana* comprises five  $\alpha$  subunits (NAC $\alpha$ ; NAC $\alpha$ 1: At3g12390, NAC $\alpha$ 2: At3g49470, NAC $\alpha$ 3: At5g13850, NAC $\alpha$ 4: At4g10480, NAC $\alpha$ 5: At1g33040) and two  $\beta$  subunits (NAC $\beta$ ; NAC $\beta$ 1: At1g73230, NAC $\beta$ 2: At1g17880). NAC $\beta$  and NAC $\alpha$  subunits are present in all eukaryotes. The number of subunits varies among plant species; for instance, the basal land plant, liverwort *Marchantia polymorpha* has only one NAC $\beta$  and one NAC $\alpha$  subunit. In contrast, species like *Physcomitrium patens* or *Oryza sativa* encode multiple NAC $\alpha$  subunits with 5 genes both, and NAC $\beta$  subunits with 5 genes present in *Physcomitrium* and 3 genes present in *Oryza*. The nucleotide and protein structures of NAC subunits are highly conserved across evolution, particularly within the NAC domain—a domain postulated to facilitate NAC $\beta$  and NAC $\alpha$  dimerization in eukaryotes (Beatrix, Sakai, and Wiedmann 2000; Wiedmann et al. 1994). NAC $\alpha$  subunits are also present in *Archaea*. However, no homologues of NAC proteins have been identified in bacterial genomes. There, the sequentially and structurally unrelated cytoplasmic chaperone Trigger factor possibly compensates for NAC functions (Deuerling et al. 1999).

NAC proteins have been observed binding to ribosomal subunits near the exit tunnel, acting as an initial cytoplasmic chaperone engaging with nascent polypeptides in various organisms, including yeast, human, and fruit fly *Drosophila melanogaster* (Hsieh et al. 2020; Wiedmann et al. 1994). The role of NAC complex is possibly vital, as silencing of NAC subunits in animals leads to embryonic lethality in *Mus musculus*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (Bloss et al. 2003; Deng and Behringer 1995; Markesich et al. 2000). However, the detailed molecular mechanisms of NAC function in plants remain to be elucidated. In a phosphoproteome enrichment analysis of activated pollen in tobacco, NAC $\alpha$  subunits were specifically phosphorylated during pollen reactivation (Fíla et al. 2016). Pollen development and pollen tube growth are critically dependent on translational efficiency and regulation. With their role in protein folding and the potential to broadly affect translation, particularly post-translational regulation during pollen activation, NACs have become promising subjects for research due to their involvement in pollen development and possibly in plant physiology in general. We thus aimed to describe its molecular mechanics in plants. Due to its remarkably high sequence conservation and its potentially critical housekeeping role in translation, we have also included a comparative study between two evolutionarily distant plant species—*Arabidopsis thaliana* and *Marchantia polymorpha*—to investigate the functional conservation of NAC subunits in plants.

It should be noted that the term NAC also refers to a group of transcription factors, derived from the initials of its members NAM, ATAF1/2, and CUC2 (Aida et al. 1997). The name is coincidental and not connected to the NAC complex, which is unrelated to these transcription factors.

## Results:

### Regulatory dynamics of gene expression in the developing male gametophyte of *Arabidopsis*

**My contribution:** In collaboration with Prof. David Twell from the University of Leicester (UK), we expanded upon their original 2004 article, which described transcription dynamics in developing pollen (Honys and Twell 2004). We integrated RNA-seq and LC MS/MS methods to encompass dynamic changes at the levels of transcripts, isoforms and proteins. My role encompassed all the bioinformatic analyses, with the exception of processing the raw LC MS/MS data. My responsibilities included processing the RNA-seq data, deducing differential gene and isoform expression analyses, categorizing and functionally annotating genes of interest from both RNA and protein datasets, and conducting k-means clustering analysis for transcriptome and proteome data comparison. I authored the complete manuscript text and generated all figures, incorporating corrections and suggestions from co-authors

Current status: published in Plant Reproduction. IF<sub>2022</sub>: 3.4, cited 4x (SpringerLink)





## Regulatory dynamics of gene expression in the developing male gametophyte of *Arabidopsis*

Božena Klodová<sup>1,2</sup> · David Potěšil<sup>3</sup> · Lenka Steinbachová<sup>1</sup> · Christos Michailidis<sup>1</sup> · Ann-Cathrin Lindner<sup>6</sup> · Dieter Hackenberg<sup>5,8</sup> · Jörg D. Becker<sup>4,6</sup> · Zbyněk Zdráhal<sup>3,7</sup> · David Twell<sup>5</sup> · David Honys<sup>1</sup>

Received: 27 May 2022 / Accepted: 26 September 2022 / Published online: 25 October 2022  
© The Author(s) 2022, corrected publication 2023

### Abstract

Sexual reproduction in angiosperms requires the production and delivery of two male gametes by a three-celled haploid male gametophyte. This demands synchronized gene expression in a short developmental window to ensure double fertilization and seed set. While transcriptomic changes in developing pollen are known for *Arabidopsis*, no studies have integrated RNA and proteomic data in this model. Further, the role of alternative splicing has not been fully addressed, yet post-transcriptional and post-translational regulation may have a key role in gene expression dynamics during microgametogenesis. We have refined and substantially updated global transcriptomic and proteomic changes in developing pollen for two *Arabidopsis* accessions. Despite the superiority of RNA-seq over microarray-based platforms, we demonstrate high reproducibility and comparability. We identify thousands of long non-coding RNAs as potential regulators of pollen development, hundreds of changes in alternative splicing and provide insight into mRNA translation rate and storage in developing pollen. Our analysis delivers an integrated perspective of gene expression dynamics in developing *Arabidopsis* pollen and a foundation for studying the role of alternative splicing in this model.

**Keywords** *Arabidopsis* · Male gametophyte · RNA-seq · Proteome · Microgametogenesis

Communicated by Prem L Bhalla .

✉ David Twell  
twe@leicester.ac.uk

✉ David Honys  
honys@ueb.cas.cz

Božena Klodová  
klodova@ueb.cas.cz

David Potěšil  
david.potesil@ceitec.muni.cz

Lenka Steinbachová  
steinbachova@ueb.cas.cz

Christos Michailidis  
Christos@ueb.cas.cz

Ann-Cathrin Lindner  
Ann-Cathrin.Lindner@gmx.net

Dieter Hackenberg  
dieter.hackenberg@kws.com

Jörg D. Becker  
jbecker@itqb.unl.pt

Zbyněk Zdráhal  
zdrahal@sci.muni.cz

<sup>1</sup> Institute of Experimental Botany of the Czech Academy of Sciences, Rozvojová 263, 165 02 Prague 6, Czech Republic

<sup>2</sup> Department of Experimental Plant Biology, Faculty of Science, Charles University, Viničná 5, Praha 2 128 00, Czech Republic

<sup>3</sup> Mendel Centre for Plant Genomics and Proteomics, Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

<sup>4</sup> Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa (ITQB NOVA), Av. da República, 2780-157 Oeiras, Portugal

<sup>5</sup> Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK

<sup>6</sup> Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

<sup>7</sup> National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

<sup>8</sup> KWS SAAT SE & Co. KGaA, Grimsehlstraße 31, 37574 Einbeck, Germany

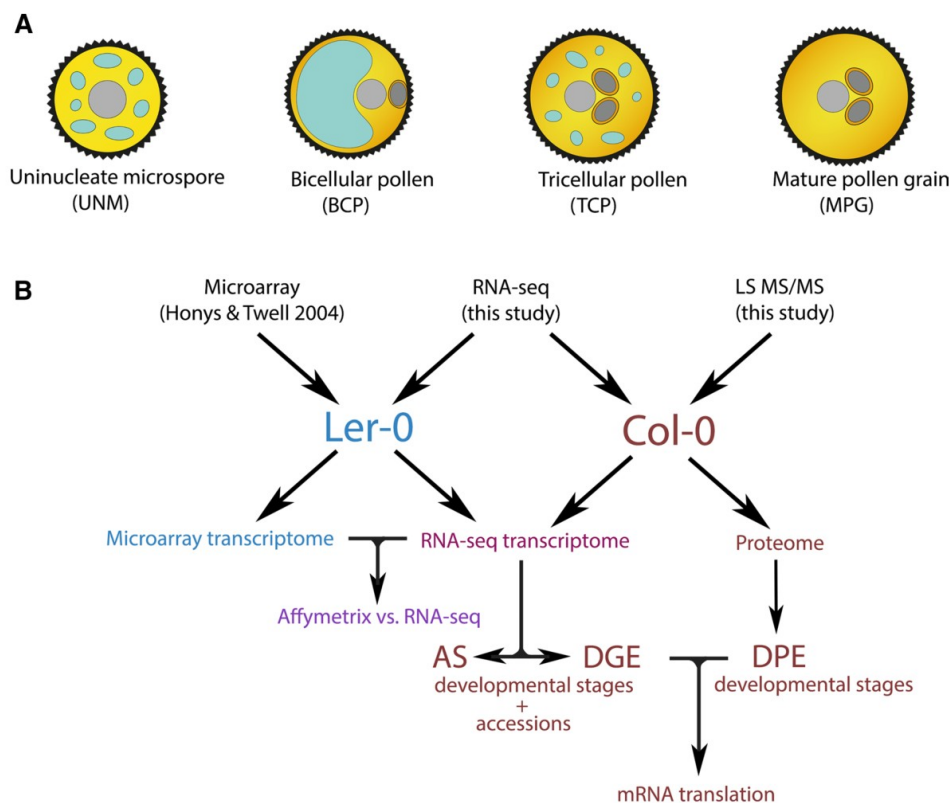
## Introduction

The life cycle of land plants alternates between haploid gametophyte and diploid sporophyte generations. In seed-bearing plants, male and female gametophytes are reduced to only a few cells supported by the maternal sporophyte and the male gametophytes are dispersed as pollen grains. The extreme reduction of the angiosperm male gametophyte to only three cells requires the regulation of gene expression in a short developmental window to enable double fertilization and seed set. Understanding the dynamics of gene expression in pollen is therefore central for understanding reproductive development, its evolution and role in crop productivity (reviewed in Xu et al. 2011; Raggi et al. 2020).

Male gametophyte development is comprised of two main phases; microsporogenesis, in which diploid microsporocytes undergo meiosis to form tetrads of haploid microspores and microgametogenesis, wherein

microspores develop into pollen grains (Fig. 1A). This study is focussed on the second phase, during which uninucleate microspores (UNM) expand and become polarized with the nucleus positioned near the cell wall. Polarized microspores divide asymmetrically at pollen mitosis I (PMI) to form bicellular pollen (BCP), which is comprised of a large vegetative cell and small generative cell. In approximately 30% of angiosperms, including *Arabidopsis thaliana*, the generative cell divides again at pollen mitosis II (PMII) to form tricellular pollen (TCP). Prior to release as a mature pollen grain (MPG), the male gametophyte is partially dehydrated (reviewed in Hackenberg and Twell 2019; Hafidh and Honys 2021). Metabolic re-activation of pollen on the female stigma results in the outgrowth of pollen tubes (PT), which are guided to ovules to deliver twin sperm cells (reviewed in Johnson et al. 2019).

The *Arabidopsis thaliana* mature pollen transcriptome has been analysed using microarray and serial analysis of gene expression (SAGE) methods (Becker et al. 2003; Honys and Twell 2003, 2004; Lee and Lee 2003; Pina et al.



**Fig. 1** Overview of datasets and analyses. **A** Key stages in *Arabidopsis* microgametogenesis. **B** Datasets used, and analyses executed in this project. AS=alternative splicing, DGE=differential gene expression, DPE=differential protein expression

2005). Microarray studies estimated *Arabidopsis* pollen transcriptome complexity to be about 6,000 genes with around 10% pollen-specific genes (Twell et al. 2006). Cell wall, cytoskeleton, signalling and vesicle transport gene ontology (GO) categories are over-represented compared with vegetative tissues, whereas transcription, translation and some metabolic pathways are under-represented. In a landmark study using Affymetrix ATH1 genome arrays, 13,977 *Arabidopsis* male gametophyte-expressed mRNAs were identified, of which 9.7% were considered male gametophyte-specific (Honys and Twell 2004). The developmental transcriptome of pollen undergoes a phase shift in transcript abundance involving a decrease in abundant early transcripts and a corresponding increase of late transcripts. These trends are associated with a reduction of cellular activity and the preferential expression of specific transcript groups during the late developmental phase (Honys and Twell 2003). The results also support the broad division of developmental gene expression into an early, more sporophyte-like phase and a late, more gametophyte-specific phase (Mascarenhas 1990). RNA sequencing (RNA-seq) has provided further insight into the pollen transcriptome of *Arabidopsis* with around 500 newly detected pollen-expressed genes and 2000 previously unannotated splicing events (Loraine et al. 2013). In a recent study, RNA-seq datasets from various organs and gametes of ten plant species were analysed to establish missing components of organogenesis and gamete development in an evolutionary context. There was conservation of the male transcriptome among angiosperms and enrichment of genes with unknown function suggesting undiscovered functions in reproductive development (Julca et al. 2021). Recently, single-cell RNA-seq analysis of microgametogenesis in maize highlighted phase shifts associated with meiotic prophase and the transition from uninucleate microspores to bicellular pollen (Nelms and Walbot 2022).

Studies of the pollen proteome with methods such as gel free liquid chromatography tandem mass spectrometry (LC MS/MS) platforms, have been used to identify changes in protein levels in distinctive conditions, or developmental stages. Studies focussed on mature pollen or pollen tubes of various plants including *Arabidopsis*, lily, tomato, rice and olive have revealed enrichment of proteins connected to metabolism, energy generation and cell structure (reviewed in Fila et al. 2017). Developmental studies of pollen for two solanaceous crops identified 1821 proteins in tomato (Chaturvedi et al. 2013) and 3888 proteins in tobacco (Ischebeck et al. 2014). Both studies reported dynamic changes in metabolic pathways and identified groups of proteins specific for each developmental stage. In another pollen proteomic study of tomato,

groups of mRNAs were identified that differ in the timing of translation under heat stress (Keller et al. 2018).

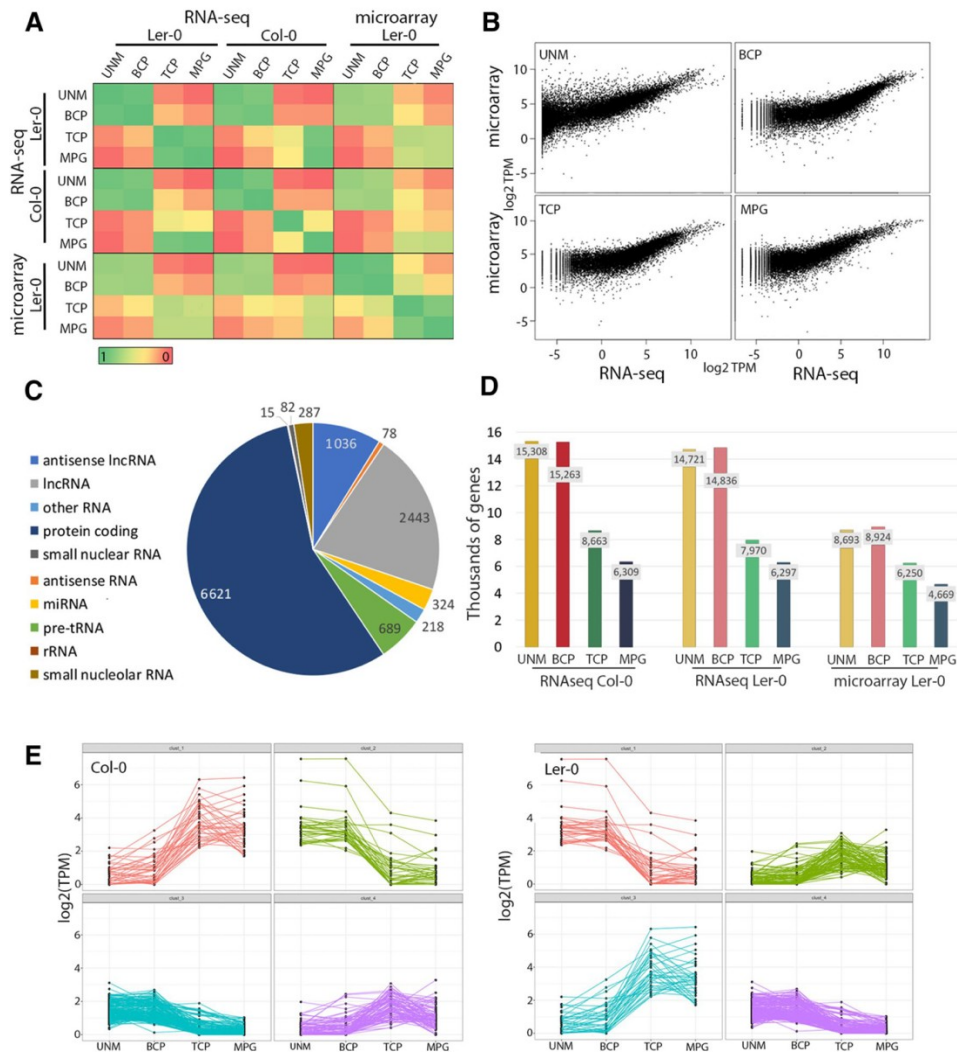
We have analysed the transcriptome and proteome of the *Arabidopsis* male gametophyte at four developmental stages (Fig. 1A). RNA-seq was used to compare transcriptomic data for Columbia-0 (Col-0) and Landsberg erecta (Ler-0) accessions, and LC-MS/MS proteome data was generated for Col-0 (Fig. 1B). The enhanced resolution and sensitivity of RNA-seq is highlighted by comparative analysis with Affymetrix ATH1 microarray data for identical developmental stages (Honys and Twell 2004; Fig. 1B). Mapping of pollen RNA-seq transcriptomes to the *Arabidopsis* transcriptome (TAIR10) identified hundreds of new mRNA alternative splicing events. Our analyses provide a map of pollen transcriptome dynamics and a catalog of alternative splicing events for two *Arabidopsis* accessions. Proteome analysis provides further insight into post-transcriptional fate during *Arabidopsis* microgametogenesis. Collectively, our study integrates global patterns of gene expression in developing pollen and provides a perspective of transcriptomic variability between two *Arabidopsis* accessions.

## Results

### RNA-seq analysis improves resolution in pollen developmental transcriptomics

We examined transcript profiles throughout male gametophyte development for the *Arabidopsis* Col-0 and Ler-0 accessions using RNA-seq. RNA-seq data were obtained from pure populations of isolated microspores and pollen at four developmental stages: unicellular microspores (UNM), bicellular pollen (BCP), tricellular pollen (TCP), and mature pollen grains (MPG). Three biological replicates were used for each stage. RNA-seq reads were mapped to 33,988 TAIR10 annotated gene models (Berardini et al. 2015) and gene expression was calculated with TPM normalization (Supplementary File 1). In previous analysis using Affymetrix ATH1 Genome Arrays, expression profiles were determined for the same four pollen developmental stages of Ler-0 (Honys and Twell 2004). The ATH1 array harboured probes sets for 22,591 gene models based on the *Arabidopsis* Genome Initiative annotation (GEO accession number: GPL198) and the majority of these (93%, 21,038 gene models) corresponded to genes that mapped in the new RNA-seq datasets.

The level of similarity between RNA-seq and microarray data was evaluated by comparison of Pearson correlation coefficients. We observed a positive correlation between corresponding datasets regardless of the platform used (Fig. 2A). RNA-seq and microarray data for Ler-0 were the most similar ( $r > 0.74$ ) for UNM and BCP stages,



**Fig. 2** Comparison of RNA-seq and pollen microarray expression data. **A** Pearson correlation ( $r$ ) shows high similarity between profiling methods. Colour bar reflects  $r$  values. **B** Scatter plot of TPM values (RNA-seq) and signal values (microarray) for the 21,038 gene models present in both analyses showing higher dynamic range for RNA-seq data. **C** Distribution of 12,939 transcripts detected only by RNA-seq. **D** RNA-seq and microarray data show similar trends in the numbers of expressed genes. RNA-seq data shows numbers of genes

with expression above 3 TPM. The microarray shows genes with reliable detection call (8). **E** Expression profiles of lncRNAs in Col-0 (left) and Ler-0 (right) represented by k-means clustering. Major profiles are early expression of lncRNAs in UNM and BCP and late expression in TCP and MPG. Each line represents a single lncRNA and all genes with summed expression for three biological replicates above 3 TPM are plotted

with lower values ( $r > 0.59$ ) for TCP and MPG. The same trend was seen for Col-0 RNA-seq and Ler-0 microarray data, but with lower correlation coefficients at all stages (Fig. 2A). For both accessions, RNA-seq data showed greater similarity between early developmental stages than to either of the late developmental stages and vice versa,

in accord with previous analysis (Hony and Twell 2004; Fig. 2A). Scatter plot comparisons illustrate the similarity between RNA-seq and microarray datasets (Fig. 2B). On the other hand, the S-shaped skewing of the nonlinear regression for genes with extremely high or low expression

(Fig. 2B), highlights the higher dynamic range expected for RNA-seq data (Marioni et al. 2008).

We compared the number of genes with detectable expression at each developmental stage with both methods and for both accessions (Fig. 2D). Previous microarray analysis identified 21,038 expressed genes in Ler-0 according to their MAS5.0 present call threshold, representing 8693 genes in UNM, 8924 in BCP, 6250 in TCP and 4669 genes in MPG (Twell et al. 2006). As expected, RNA-seq detected greater numbers of genes for both accessions. In Ler-0, 14,721 genes were detected in UNM, 14,836 in BCP, 7970 in TCP and 6297 in MPG. The numbers of genes in Col-0 were slightly higher, at 15,308 in UNM, 15,263 in BCP, 8663 in TCP and 6309 in MPG. For both datasets, there was substantial reduction in the number of expressed genes at late developmental stages.

RNA-seq data provided evidence for 26,916 expressed genes in one or more stages of developing Ler-0 pollen, which almost doubled the number (13,977) detected by microarray analysis (out of the total 22,591 probes used; Honys and Twell 2004). Of the 12,939 newly detected genes, approximately half (6621, 51.1%) are protein coding RNAs, while a substantial fraction (3479, 26.9%) are long non-coding RNAs (lncRNAs). The remaining 2839 expressed genes (22%) were other non-coding RNAs, such as pre-tRNAs or small nucleolar RNAs (Fig. 2C).

The expression profiles of newly detected pollen-expressed genes were compared to that of all expressed genes. The average expression of newly detected genes was 17.6 TPM in Col-0 and 16.3 TPM in Ler-0, whereas average expression for all gene models (33,988 genes) was 29.4 TPM for both accessions. For newly detected protein-coding genes, expression was higher than for non-coding genes and their average expression was similar to that for all gene models at 27.5 and 25.6 TPM in Col-0 and Ler-0, respectively. GO analysis of the newly detected protein-coding genes revealed only three enriched biological process terms, 'regulation of protein localization to cell surface' (40 genes), 'regulation of double fertilization forming a zygote and endosperm' (41 genes) and 'unclassified' (1758 genes).

We further analysed lncRNAs as a novel transcript category that could not be studied in previous work (Honys and Twell 2004). Out of 3479 lncRNAs, 2443 were annotated as long non-coding RNAs and 1036 as antisense lncRNAs or natural antisense transcripts (NATs), transcribed from the opposite strands of either protein-coding or non-coding genes. The average expression signals of all lncRNAs were low in both accessions (2.5 TPM in Col-0; 2.6 TPM in Ler-0), but a significant number of lncRNA genes were expressed at each stage. In both accessions the number of lncRNAs peaks at BCP stage and declines thereafter. In Col-0 there were 333 (UNM), 347 (BCP), 217 (TCP) and 139 (MPG) lncRNAs with expression values above 3 TPM

and in Ler-0 270 (UNM), 301 (BCP), 174 (TCP) and 137 (MPG). The reduced numbers of lncRNAs expressed at later developmental stages resembles the trend for coding RNAs and distinct early and late lncRNA clusters were apparent (Fig. 2E).

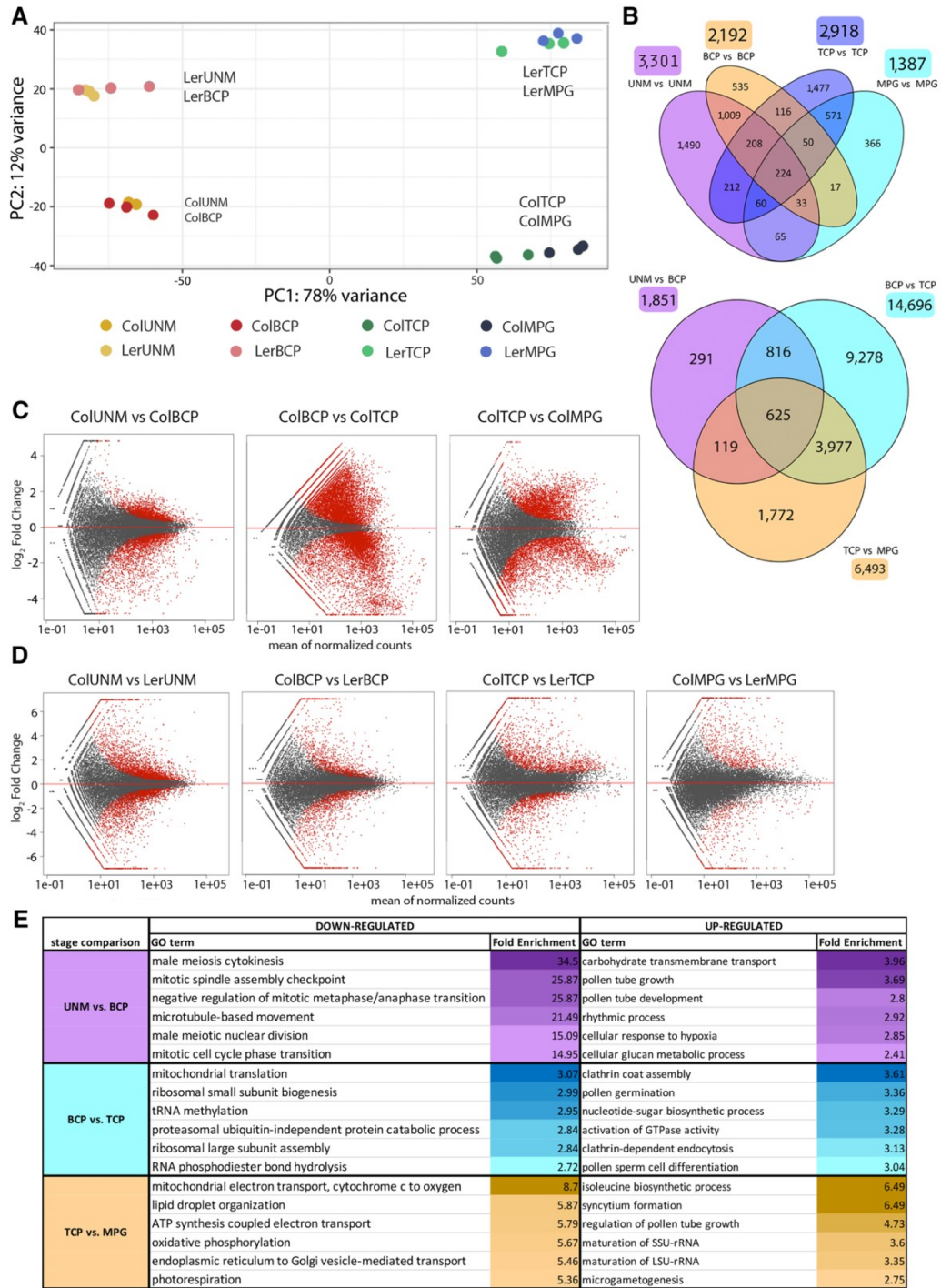
In summary, the quantification of gene expression in developing pollen by RNA-seq analysis is in close accord with previous microarray analyses in Ler-0 (Honys and Twell 2004) but delivers new information about both protein coding and non-coding transcripts including data for different *Arabidopsis* accessions.

### Developmental transcriptome profiles reflect changes in numerous biological processes

Correlation analysis of gene expression profiles across developmental stages and between platforms (RNA-seq and microarray) unsurprisingly showed the highest similarity between adjacent early (UNM-BCP) and late (TCP-MPG) stages (Fig. 2A). Principal component analysis (PCA) of RNA-seq data identified four clusters, highlighting the greater similarity between accessions rather than between early and late pollen developmental stages (Fig. 3A). The similarity between early and late developmental stages was also apparent by hierarchical clustering, with two main branches according to accession (data not shown).

We investigated differential gene expression (DGE) with the DESeq2 package and set thresholds of statistical significance to  $\log_2$  fold change  $< -1$  or  $> 1$  and adjusted  $p$  value  $< 0.05$ . In Col-0, there were 1851 differentially expressed genes (DEGs) in the two early stages, with 547 up-regulated in UNM and 1304 up-regulated in BCP (Fig. 3B, D, Supplementary File 2). The major developmental shift was during BCP-TCP transition, resulting in 14,696 DEGs. While 8433 transcripts were up-regulated in BCP, only 5763 were up-regulated in TCP (Fig. 3D, Supplementary File 2), confirming the reduced transcriptome complexity after BCP stage (Honys and Twell 2004). In two late developmental stages, 6493 DEGs were identified, with 2915 transcripts up-regulated in TCP and 3578 in MPG (Fig. 3D, Supplementary File 2). There was substantial overlap in DEGs between developmental stages, with 3977 DEGs shared between BCP-TCP and TCP-MPG transitions and 625 DEGs shared across all stage transitions (Fig. 3B). Conversely, the majority of DEGs between BCP and TCP stages were unique, further supporting the hypothesis of a major transition between these stages. Similar trends, namely the major developmental shift of 13,349 DEGs between BCP and TCP stages, were also observed in Ler-0.

The extent and dynamics of gene expression changes were also examined in individual developmental stages between accessions. The numbers of genes expressed at each stage were similar for the two accessions (Fig. 2D). The



**Fig. 3** Differential gene expression throughout pollen development. **A** PCA plot of RNA-seq data. Four clusters separate early and late stages and distinguish Col-0 and Ler-0. **B** Venn diagram (upper panel) showing numbers of DEGs in pairwise comparisons of pollen stages of Col-0 and Ler-0. Venn diagram (below) of the total number of DEGs and the overlaps between stage transitions (Col-0). **C** MA plots of DEGs in stage transitions (Col-0). The most prominent changes occur during BCP→TCP transition. DEGs with  $\text{padj} < 0.05$  are in red. **D** MA plots of DESeq2 results between accessions at each stage. The highest number of DEGs are at UNM stage between Col-0 and Ler-0. DEGs with  $\text{padj} < 0.05$  are in red. **E**. Table of the six most enriched GO terms between stage transitions

numbers of DEGs between accessions declined by 57.8% from UNM to MPG, again reflecting the general decrease in overall transcriptome complexity. Surprisingly, more DEGs were observed in TCP than in BCP. The highest number of DEGs between accessions was found at UNM stage (3301 genes), followed by 2918 and 2192 DEGs in TCP and BCP, respectively, with 1387 DEGs in MPG (Fig. 3B, C). The significant overlap between DEGs in UNM and BCP (1009 genes) and between MPG and TCP (571 genes) highlights the reduced variability of DEGs within early- and late-stage clusters (Fig. 3B).

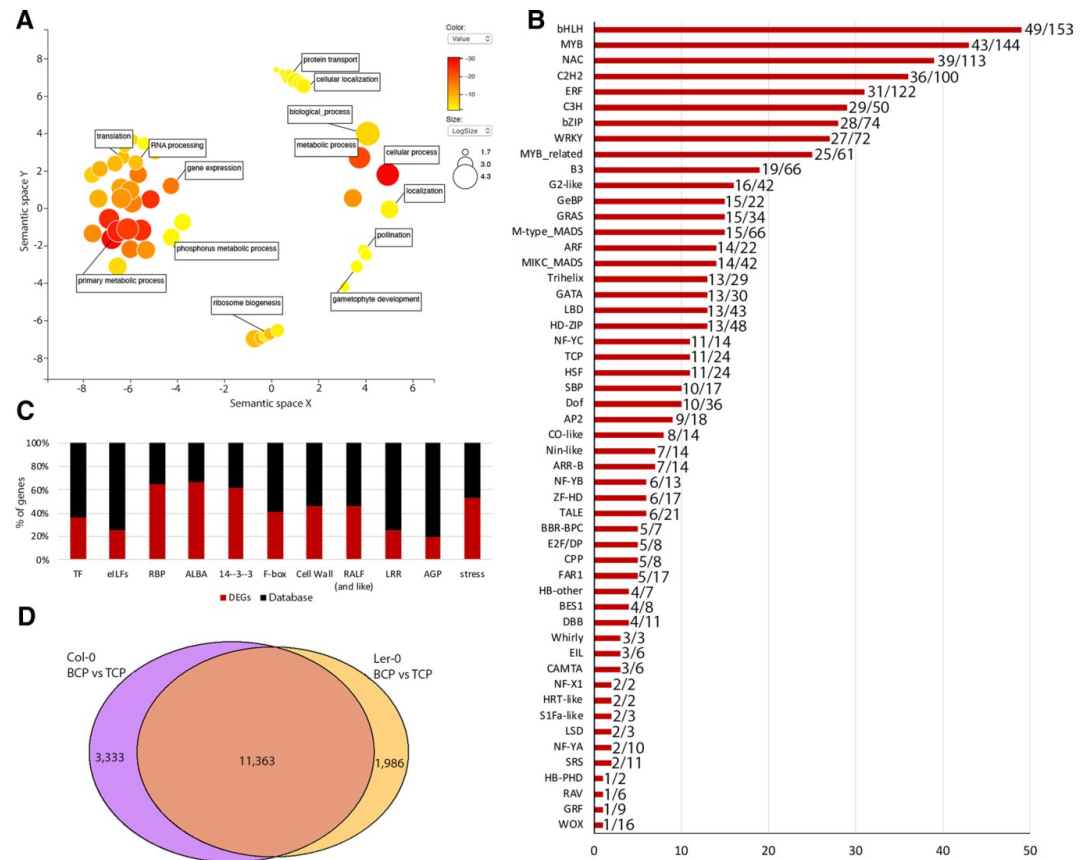
Next, we explored the potential biological significance of transcriptome changes by examining developmental shifts in GO categories (Fig. 3E). During UNM-BCP transition, 115 positively enriched GO categories were upregulated in UNM. These included terms associated with mitotic events (cyclin dependent protein phosphorylation, phase transition control or spindle organization) indicating preparation of microspores for cell division. Notably, 17 of 41 plant kinases are differentially expressed. Transcripts upregulated in BCP were mainly connected to glucan metabolism and transmembrane transport, including cell wall-associated transcripts. Remarkably, 22 transcripts associated with pollen tube development or pollen tube growth are also among upregulated DEGs. In more detail, 9 B-box type zinc finger and 20 EF hand domain proteins were among BCP upregulated DEGs with overrepresented protein domains. B-box transcription factors have diverse roles including flowering time regulation or stress tolerance and it has also been noted that they are expressed in pear pollen (Cao et al. 2017). The EF hand domain proteins include several calcium dependent protein kinases (CPK20, CPK16, CPK26, CPK6, CPK2), which continue to increase in expression to TCP stage and are reported to play a role during pollen tube growth (Yang et al. 2021).

A major transcriptomic shift was associated with BCP-TCP transition. GO analysis of the 13,349 (Ler-0) to 14,696 (Col-0) DEGs uncovered 325 enriched biological process terms. BCP stage was defined by upregulated transcripts connected to translation, from tRNA and ribosome biogenesis to mRNA maturation and protein folding. From TCP stage onwards, the enrichment of translation-related GO

categories was drastically reduced, and transcripts connected to pollination, pollen germination and pollen tube growth increased. For example, 40 transcripts were connected to pollen germination and 99 to pollen tube development. BCP-TCP transition was also characterized by upregulation of transcripts connected to vesicle transport including both exocytosis and clathrin-dependent endocytosis, while the signalling terms activation of GTPase activity, Rab protein signalling and Ras protein signalling were also enriched along with glycerophospholipid metabolic process. For most of these terms, more than 50% of total genes belonging to the terms were differentially expressed, indicating a major transcriptome shift linked to these processes. The most enriched GO terms for each stage transition are summarized in Fig. 3E.

Genes encoding regulatory proteins, involved in transcription, signalling cascades, protein modification and degradation were also modulated during BCP-TCP transition (Fig. 4A). Among transcription factors (TFs) 37% (626 genes from 52 TF families) showed altered expression, with 397 TFs upregulated in BCP and 229 in TCP (Fig. 4C, Supplementary File 3). Notable TF families, with at least 50% of members differentially expressed and a minimum of 75% of DE members upregulated in the BCP stage, were AP2, auxin response factors (ARF), BES1, CAMTA, CO-like, GeBP, homeobox-other (HB-other) and homeobox-PHD finger (HB-PHD). Other TF families, represented by less than 50% of members among DEGs and mainly upregulated in BCP, were B3, basic/helix-loop-helix (bHLH), GATA and homeodomain-zip (HD-zip). TFs upregulated at TCP stage belonged to EIL, NF-X1 and S1Fa-like gene families. Several large TF families, including basic leucine zipper (bZIP), C2H2, ethylene responsive factors (ERF), MYB and MYB-related, C3H, NAM, ATAF, and CUC (NAC), WRKY or MADS-box showed similar numbers of up- and down-regulated members at each stage, suggesting that their activity is modulated by exchange among members during BCP-TCP transition.

We further examined the expression dynamics of specific groups of genes. More than 60% (756 genes) of genes encoding RNA-binding proteins (RBP) were differentially expressed during BCP-TCP transition (Fig. 4C). Most RBPs were highly expressed during early stages, with 568 genes upregulated in BCP and only 190 with higher expression in TCP. Notable genes downregulated in TCP included four ALBA (acetylation-lowers binding affinity) superfamily protein genes (Náprstková et al. 2021) and seven encoding all subunits of nascent polypeptide-associated complex (Fila et al. 2020). Other markedly shifted groups included stress-associated transcripts, cell wall transcripts, 14-3-3 proteins, F-box proteins or RALFs (rapid alkalization-like factors), RALF-like groups and receptor kinases, as well as translation initiation factors (eIFs; Fig. 4C). Accordingly,



**Fig. 4** Functional categorization of differentially expressed genes in developing pollen. **A** ReviGO plot of enrichment analyses of BCP vs TCP DEGs indicating major transcript groups. **B** Numbers of DE TFs during BCP-TCP transition. Each bar shows how many DE TFs are

represented out of the total. **C** Percentage of DEGs in selected groups compared to reported numbers. **D** Venn diagram showing overlap between BCP-TCP DEGs in Col-0 and Ler-0

enriched KEGG pathways highlighted enrichment of processes related to mRNA metabolism, with most of the transcripts being upregulated in BCP (Fig. 5A, B, C). The large overlap in DEGs (11,306) between accessions highlights the conservation of developmental regulation during BCP-TCP transition in *Arabidopsis* (Fig. 4D).

The final developmental phase from TCP to MPG shows fewer changes compared with BCP-TCP transition and is enriched for 379 GO terms. The 2915 DEGs upregulated in TCP are enriched for energy and biosynthetic processes including ATP synthesis coupled proton transport and electron transport chain. The expression of these gradually declines from UNM to MPG, similar to those coding for vesicle transport- and exocytosis-related proteins. Transcripts connected to pollen, pollen tube germination and

pollen development (109 DEGs) are upregulated in TCP, while 36 DEGs upregulated in MPG are involved in pollen tube growth regulation and microgametogenesis. Notably, some MPG-upregulated transcripts are highly abundant with mean expression of 553 TPM, whereas average expression in MPG is only 29 TPM. A further 82 transcripts were included in the enriched term response to cold and 90 in response to abscisic acid. Of the 3,087 annotated DEGs upregulated in MPG, 26% (816 genes) are linked to stress response. In comparison with other stages, the upregulation of stress response transcripts is apparent. The list includes groups of genes responsive to salt, radiation, heat, cold, desiccation and chemical stimuli, as well as genes responsible for DNA repair, protein oxidation, signalling, transport and response to biotic stress.



Although translation-associated GO terms were also enriched, the overall expression of such transcripts was low when compared to similarly associated transcripts upregulated in earlier stages. For example, ribosome biogenesis is enriched in both BCP and MPG relative to TCP, but the mean expression of 127 genes in this category is 159 TPM in BCP and only 18.1 TPM in MPG. Therefore, a decline in the abundance of translation-related genes is apparent from BCP onward. Among GO terms shared between BCP and MPG, only four genes had higher expression in MPG than in BCP. These include ribosome biogenesis genes *At3g22510* and *REI1-LIKE1* (*At4g31420*; Cheong et al. 2021) and genes with translation functions, *SUI1*-family initiation factor (*At1g54290*; Bach-Pages et al. 2020) and *TMA7* (*At3g16040*; Fleischer et al. 2006).

MPG-upregulated transcripts also showed enrichment for the cellular location of their predicted products, with for example 67 apoplast, 109 cell wall and 521 plasma membrane proteins. The complete set of enriched GO terms for all stages is plotted in the Supplementary Fig. 1.

In summary, transcriptome dynamics of developing pollen impacts numerous biological processes with major changes during phase transition from early to late stages. Early stages are accompanied by enhanced transcription of genes encoding proteins involved in translation, mRNA, and protein processing and also in cell division. Late phase dynamics suggest preparation for pollen desiccation by enhanced transcription of stress-related genes, and later for pollen activation and pollen tube growth, by accumulation and storage of transcripts associated with vesicular transport, energy metabolism or pollen tube growth.

### There are no major transcriptome changes between Col-0 and Ler-0 during pollen development

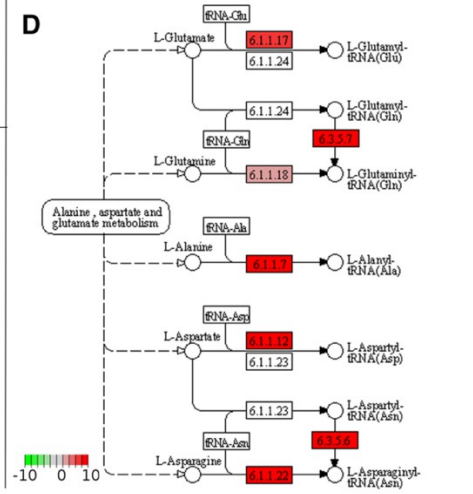
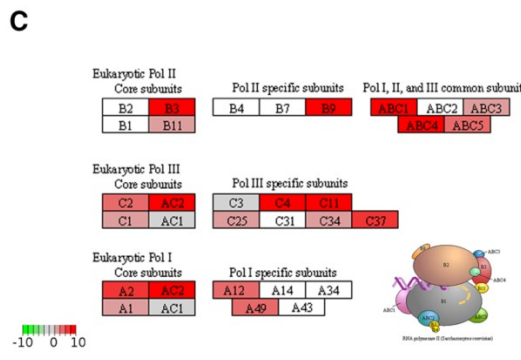
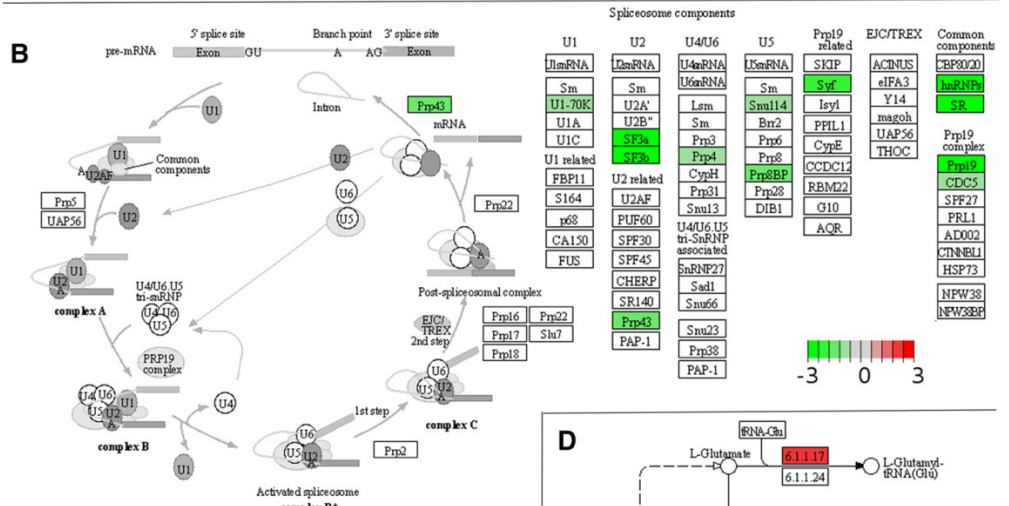
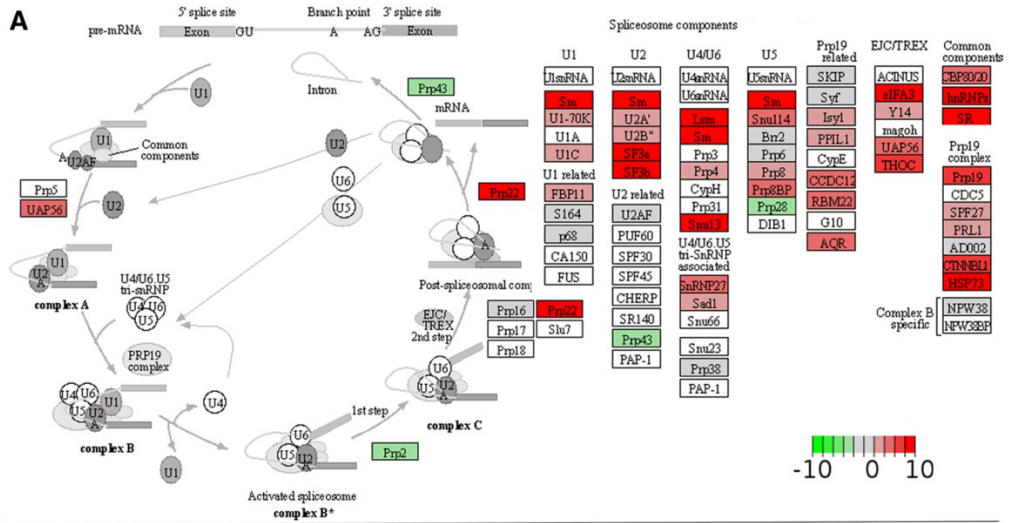
The analysis of developmental transcriptomes between Col-0 and Ler-0 accessions did not identify major differences in biological processes or metabolic pathways. Among processes with the highest fold-enrichment were GO terms related to DNA replication, but the associated transcripts were usually of low abundance (mean of 9.6 TPM in Col-0 and 2.2 TPM in Ler-0). The 13 enriched processes upregulated at UNM stage in Ler-0 included 6 DNA replication-associated terms, whereas transcripts representing replication connected GO terms were more abundant in Col-0 at TCP and MPG stages (Fig. 6A). Phosphorylation-associated transcripts as well as 10 transcripts related to regulation of pollen tube growth were also enriched in Col-0 at UNM stage. At MPG stage, Col-0 showed 51 enriched processes, whereas Ler-0 was enriched for transcripts involved in protein targeting to endoplasmic reticulum and the ubiquitin-dependent ERAD pathway.

Studies of in vitro pollen germination and tube growth have highlighted differences between *Arabidopsis* accessions. The carbohydrate type and concentration (Hirsche et al. 2017), pH (Rodriguez-Enriquez et al. 2013) and salt conditions (Azarov et al. 1990) of in vitro pollen germination media as well incubation temperature (Boavida and McCormick 2007) can lead to drastic differences in germination responses between Col-0 and Ler-0. The differential expression of sucrose-H<sup>+</sup> symporters are implicated (Sauer et al. 2004), specifically *AtSUC1* (*ARABIDOPSIS THALIANA SUCROSE-PROTON SYMPORTER 1*), which has reduced immunodetectable protein in Ler-0 pollen (Feuerstein et al. 2010). Similarly, we observed a log<sub>2</sub>fold difference of −4.4 for *AtSUC1* (*At1g71880*) at UNM stage and reduced signal in Ler-0 throughout development (mean expression 66.3 TPM in Col-0 and 24.44 TPM in Ler-0).

Genes encoding F-box proteins were the most abundantly represented group of DEGs between accessions, representing 2.5% to 5% of DEGs, with the greatest variability at early developmental stages (Fig. 6B). Further, the mean (217 TPM) and median (120 TPM) values were non-negligible with maximal expression of a DE F-Box transcript (*At4g27050*) reaching 2010 TPM in UNM. The expression statistics and AGIs of F-box genes at each stage are given in Supplementary File 4. Protein ubiquitination was connected to 35 differentially expressed F-box proteins. Notably, five F-box genes were exclusively expressed in Col-0 (*At2g04810*, *At3g47130*, *At5g38391*, *At5g42460*, *At5g56380*), while two were only expressed in Ler-0 (*At5g56440*, *At3g21170*). Although some F-box transcripts differ in expression between accessions, their developmental expression profiles were similar for most of the genes (Fig. 6C).

### Pollen development is accompanied by isoform switches and differential exon usage

We used complementary methods to examine differential exon usage (DEU) across pollen developmental stages. First, we employed DexSeq using the scaled expression (TPM) of each isoform from RSEM analysis as input. Next, differential isoform usage (DIU) and its consequences were analysed using IsoformSwitchAnalyzeR with the same data input. In total, 1769 exons in 1132 genes showed significant DEU in Col-0 UNM-BCP transition. We detected DEU for 1769 exons in 1131 genes during BCP-TCP transition and 1037 exons in 588 genes for TCP-MPG transition. As expected, DIU events were much less common than DEU. DIU analysis identified 336 isoform switches in 204 genes during UNM-BCP transition, only 17 of which were DEGs. There were 837 isoform switches in 458 genes between BCP and TCP and 129 isoform switches in 78 genes during TCP-MPG transition (Supplementary File 5). We observed



◀**Fig. 5** KEGG pathway enrichment during BCP-TCP stage transition. KEGG pathway analysis shows upregulation of mRNA processing (A spliceosome, transcripts), transcription (C RNA polymerase) and translation (D tRNA synthesis) pathways in BCP compared to TCP. Spliceosome genes show peak expression in BCP, but protein expression peaks in TCP (B spliceosome, proteins). Plots were rendered by Pathview. Genes upregulated in TCP (green) or BCP (red) are highlighted. The scale indicates  $\log_2$  fold change of DEGs

a similar trend in the number of expressed isoforms to the number of expressed genes, which decreased from UNM to MPG (Fig. 7A). Only four transcript isoforms switched between all three stage transitions, with most being stage-specific events (Fig. 7B).

The isoform switches were further categorized as intron retention (IR), exon skipping (ES), alternative transcription start sites (ATSS), alternative transcription termination sites (ATTS), alternative 5' splice site (A5) and alternative 3' splice site (A3), multiple exon skipping (MES) or mutually exclusive exons (MEE). The distribution of alternative transcription events was comparable for all three stage transitions, with alternative transcription start sites and alternative transcription termination sites being the most represented followed by alternative 3' splice site and intron retention. Only a few isoforms switches resulted in multiple exon skipping (Fig. 7C, D). We analysed the consequences of isoform switches focussing on changes in coding potential, domain and signal peptide presence, intron retention or sensitivity to nonsense-mediated decay (NMD), that may drastically affect protein cellular localization or function. There were 90 switch consequences for 69 genes during UNM-BCP transition. This resulted in domain changes in 20 genes (18 gain, 2 loss), while 50 switches involved intron retention and five presence/absence of a signal peptide. Changes related to potential sensitivity to NMD, which was integrated from the presence of premature termination codons in isoform's sequence, was observed for 12 genes. In case of 3 gene, the NMD sensitivity led to decreased expression of the sensitive isoform. Similar results were observed throughout pollen development with the majority of DIUs involving intron retention (UNM-BCP and BCP-TCP) or resulting in a change in domain presence (TCP-MPG) (Fig. 7E).

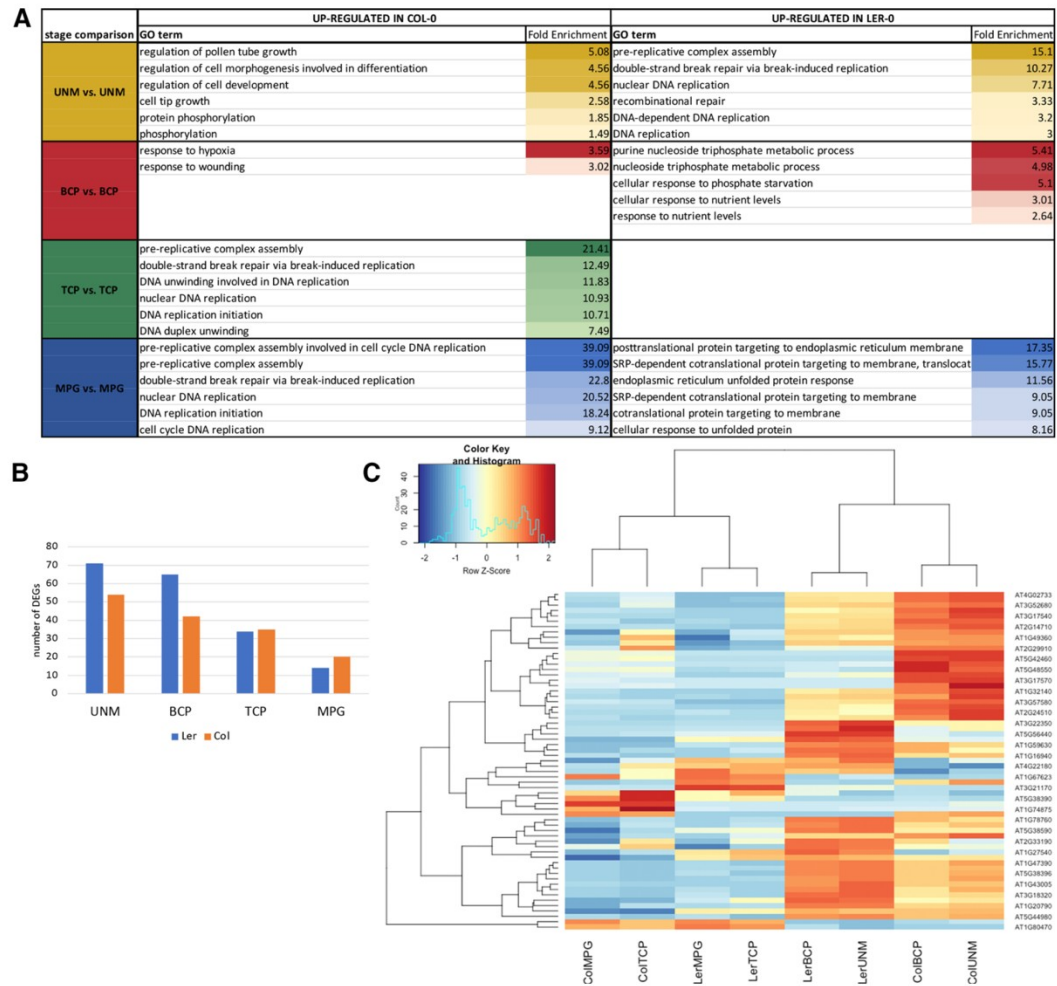
GO enrichment for biological function was calculated based on the annotation of genes with DEU and DIU and the most significant genes and those with switch consequences were investigated manually. The comparison of DEU and DIU for UNM-BCP transition resulted in the enrichment of genes involved in meiosis cytokinesis together with chromosome organization. DEU occurred in transcripts connected to cytoskeleton and vesicle trafficking (including vesicle transport along actin filament, cytoskeleton organization or vacuole transport). Involvement of developmental genes was represented by the terms pollen development and microgametogenesis. There were 11 genes with DIU involved in

posttranscriptional regulation of gene expression and 12 genes with a role in mRNA processing including splicing factors and ribonucleases.

Genes with switch consequences resulting in domain changes included VPS15 kinase (Xu et al. 2011; Liu et al. 2020), which is involved in autophagy (Fig. 8A). Interestingly, the kinase domain was lost from the transcript in BCP. Another candidate was the H<sup>+</sup>ATPase AHA3 (At5g57350; Bock et al. 2006; Robertson et al. 2004) with the isoform lacking ATPase domains dominant in BCP. Other candidates emerged when isoforms with high difference in isoform fraction (dIF) were examined. *THERMOSENSITIVE MALE STERILE 1 (TMS1)* isoform 2 (At3g08970.2; Yang et al. 2009; Ma et al. 2015) was only expressed in BCP, whereas isoform 1 was dominant in UNM. Similarly, histone deacetylase HDA5 (At5g61060), a salt stress-responsive protein also involved in flowering time regulation (Luo et al. 2015; Ueda et al. 2017), switched from isoform 1 to 2 during UNM-BCP transition.

During BCP-TCP transition in Col-0 there were changes in 836 isoforms of 458 genes and for 331 genes these switches had consequences. In total 67 isoform switches resulted in change in NMD sensitivity of which 27 NMD sensitive isoforms were downregulated. In Ler-0, 145 switch consequences were identified among the 548 isoforms of 304 genes. In total, 134 genes showed DIU in both accessions. Similar to the DGE analyses, the BCP-TCP transition accounts for the greatest number of isoform switches. The GO analysis of the genes with DIU showed enrichment in GTPase activity, membrane docking, mRNA splicing, protein transport and vesicle trafficking. Two most significant isoform switches resulted in the change of sensitivity to NMD via the dominance of the NMD-sensitive isoform of stress-responsive gene SAP13 (At3g57480, Zinc finger AN1 domain-containing stress-associated protein 13, Fig. 8B; Dixit et al. 2018) and of RPA2A (At2g24490, Replication protein A component responsible for transcriptional gene silencing; Aklilu et al. 2020; Xia et al. 2006). In AGAMOUS-LIKE MADS-box protein AGL65 (Adamczyk and Fernandez 2009) the BCP-prevailing isoform At1g18750.5 lacks the MADS-box domain, which is present in one of the most expressed isoforms (At1g18750.4) in TCP.

The TCP-MPG included 129 isoform switches with 30 causing the switch consequences. For example, 11 switches resulted in the change in NMD sensitivity and 8 NMD sensitive isoforms became downregulated during the TCP-MPG transition. Functionally, the progression is characterized by differential usage of isoforms enriched for transporters (17 in total). Examples include the functionally important PHOSPHOLIPASE A2 DELTA (At4g29470) which is essential for pollen germination and pollen tube growth (Kim et al. 2011) and the pollen-specific mechano-sensitive channel-like protein MSL8 (At2g17010) which has a protective role during



**Fig. 6** Functional categorization of differentially expressed genes between Col-0 and Ler-0. **A** Summary of the most enriched GO terms ( $\leq 6$ ) among DEGs between Ler-0 and Col-0 sorted by stage. **B** Number of F-box genes differentially expressed between Ler-0 and Col-0

at each stage. **C** Heatmap of candidate F-box gene expression. There are similar trends between early and late developmental phases for most of the genes in both accessions

pollen hydration (Hamilton et al. 2015). Other candidates with switch consequences include the shorter TCP dominant isoform 2 of phosphorylase At4g28940, which lacks a signal peptide (Fig. 8C), while an isoform switch in Glyoxalase GLXI-LIKE9 (At2g32090; Schmitz et al. 2018) results in the gain of the glyoxalase domains.

We found more differences in alternative splicing (AS) between Col-0 and Ler-0 during early development. There were 560 isoform switches in 330 genes in UNM and 650 in 375 genes in BCP, of which 150 and 180 had consequences

respectively. In late development, there were 289 switches in 153 genes in TCP and 216 switches in 117 genes in MPG. These isoform switches had consequences in 78 and 64 genes, respectively. Isoform switches between accessions were detected in 16 genes across all four stages. These data might reflect general differences in pollen isoform usage between accessions particularly in early developmental stages. In most cases, switches were connected to a specific developmental stage. One of the most significant switches in UNM resulted in a change of isoform usage for heat shock

protein THERMOSENSITIVE MALE STERILE 1 (Yang et al. 2009; Fig. 8D). Genes with isoform switches in BCP were enriched for RNA binding proteins and biological processes including translation and RNA processing. These categories included seven 60S and three 40S ribosomal proteins and eukaryotic translation initiation factor 3 subunits C and D. Another interesting candidate was the voltage dependent mitochondrial channel VDAC5 (At3g49920; Fig. 8E), which shows polymorphisms associated with flowering adaptation (Tabas-Madrid et al. 2018).

### Proteomic analyses of pollen development

Samples of four pollen developmental stages of Col-0 were processed and analysed by LC–MS/MS, with each stage represented by three biological replicates and two technical replicates. In summary, we identified 4965 protein groups (PGs) in the whole data set (Supplementary file 6). The highest numbers of PGs were identified in TCP (3197 PGs, PGs detected in four of six analyses per sample type using at least two peptides) and in MPG (3100 PGs), while the lowest number of PGs was detected in BCP (2716 PGs). In UNM samples, we identified 2963 PGs. Pearson correlation of the total of 6 replicates for each stage shows highest similarity between early stages (UNM and BCP) and a good correlation between replicates (Fig. 9A).

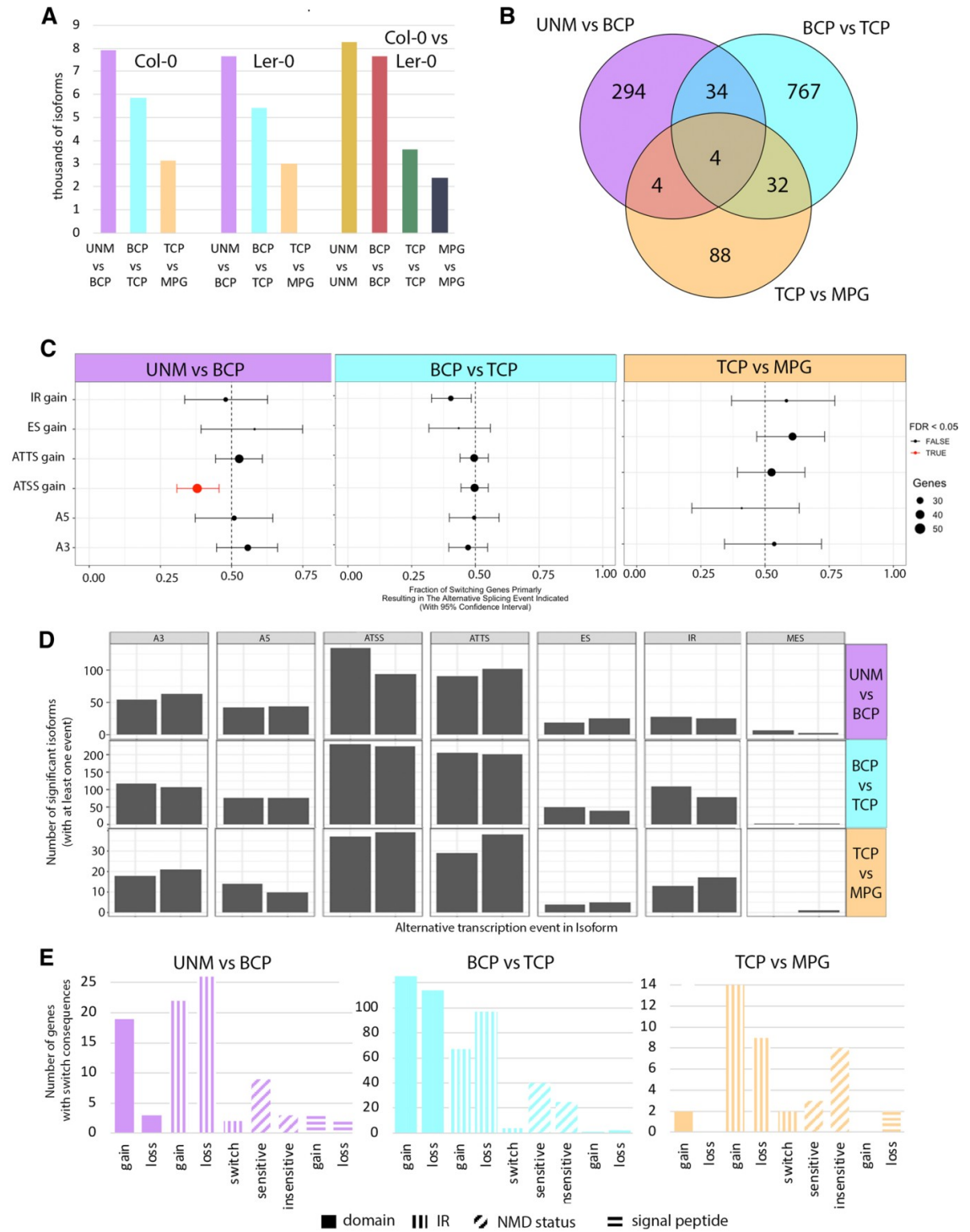
Differentially expressed proteins (DEPs) between adjacent stages were identified. The threshold for significant quantitative change was set as  $\log_2$  fold change (FC) < -1 or FC > 1, and adjusted *p* value < 0.05 (Supplementary File 7). Proteins with quantified replicates in only one sample of the two and zero in other were retained as qualitative changes. During UNM-BCP transition, 312 and 347 proteins were quantitatively upregulated, while 232 and 49 proteins were qualitatively upregulated in UNM and BCP, respectively (Fig. 9B, C). Both quantitative and qualitative changes were analysed together for functional consequences. The transition from UNM to BCP resulted in higher abundance of proteins involved in lipid storage, retrograde vesicular transport, and cell-wall organization. These groups included glycine-rich proteins, oleosins, expansins, and the cell wall regulators pectinesterases and pectinesterase inhibitors. Also, three subunits of the coatomer complex were up-regulated, epsilon-1 (At1g30630), zeta-2 (At3g09800) and zeta-3 (At4g08520) subunits (Cabada Gomez et al. 2020). Numerous proteins with a role in pollen germination and pollen tube growth showed enrichment. For example, three pollen coat proteins, extracellular lipase 4 (At1g75910; Updegraff et al. 2009), extracellular lipase 6 (At1g75930; Dong et al. 2016) and glycine rich protein 17 (At5g07530), and five pollen tube tip proteins including pectinesterase inhibitor 1 (At1g48020; Röckel et al. 2008), LOST IN POLLEN TUBE GUIDANCE 2 (At3g02810) 73), ANXUR1 (At3g04690;

Boisson-Dernier et al. 2009), pollen receptor like kinase 6 (At5g20690) and ATP-binding cassette G28 (At5g60740).

The largest number of DEPs were identified for the BCP-TCP transition with 1,048 and 946 proteins with quantitative up-regulation in TCP and BCP respectively (Fig. 9B, C, Supplementary File 7). There were 652 qualitatively upregulated proteins in TCP and 165 in BCP. Among BCP upregulated proteins, there was an enrichment in proteins involved in cell wall modification mainly consisting of pectinesterases (PME4, PME5, PPME1, PME67, PME49 and PME48) and pectinesterase inhibitors (PME43, PME28 and VGDH2). A further group of proteins containing the c11 pectin lyase fold domain represented connection to carbohydrate metabolism. Another enriched term was lipid storage, which was represented by four oleosins and one glycine-rich protein. Pollen tube growth regulating proteins also showed high fold enrichment as well as oxidation–reduction process proteins. Numerous processes connected to saccharide metabolism and utilization were enriched in BCP including sugar transport, pyruvate, pectin, sucrose, galactose, glutamate, fructose, malate, UDP-rhamnose metabolism as well as processes involved in gluconeogenesis as aspartate and glycerol metabolic processes. Biosynthesis of nucleotides is also represented.

At the beginning of the late developmental phase, proteins connected to mRNA processing, translation and protein modification were upregulated (Fig. 9D). For example, enriched processes included rescue of stalled ribosome, ribosome subunit assembly, regulation of translation, protein maturation, folding and targeting. There were 93 proteins possessing an RNA recognition motif, 15 proteins annotated as tRNA binding and 134 ribosomal proteins. Spliceosome was also observed as an enriched KEGG pathway (Fig. 5B). The BCP and TCP transcriptomes and corresponding proteomes overlapped for several spliceosome components including the PRP19 complex, which showed higher transcript abundance in BCP. Spliceosomal protein expression was then upregulated during BCP-TCP transition, suggesting uncoupling of transcription and translation between early and late stages (compare Fig. 5A and B). Also, we observed enrichment of proteins involved in posttranscriptional regulation of gene expression and negative regulation of transcription, which includes five histone deacetylases, the chromatin regulator ZUOTIN-RELATED FACTOR 1 (ZRF1) and Argonaute 4 (AGO4).

During the TCP-MPG transition 712 proteins were quantitatively upregulated in MPG, while 1011 were more abundant in TCP. A further 283 and 205 proteins were qualitatively upregulated in TCP and MPG respectively (Fig. 9B, C). The DEPs upregulated in TCP included the dominant categories connected to metabolism of fatty acids, nucleotides, and carbohydrates together with vesicle-mediated transport. Among DEPs upregulated in MPG, there are



◀**Fig. 7** Analysis of differential isoform usage (DIU) in Col-0 pollen development. **A** The number of isoforms passing statistical threshold (both gene and isoform expression > 3 TPM) declines with stage progression. **B** Venn diagram showing high DIU variability with only 4 isoforms switches in all 3 pairs of stages compared. **C** Enrichment of alternative isoform types during Col-0 stage transitions. Only ATSS gain during UNM-BCP transition is significantly enriched. **D** The distribution of alternative isoform types during Col-0 stage transitions is conserved during pollen development. ATSS and ATTS events are always most frequent. **E** Switch consequences of DIU. IR is most frequent during UNM-BCP and TCP-MPG transitions. Change of domain presence is the most frequent isoform switch consequence during BCP-TCP transition

proteins responsible for epigenetic post-transcriptional silencing, namely ARGONAUTE 1 (AGO1), AGO5, AGO9, SET DOMAIN PROTEIN 18 (SUVR2) or NUCLEAR RNA POLYMERASE D1B (NRPD1B). The enrichment of processes connected to chromatin remodelling, chromosome organization, gene silencing and siRNA production indicate epigenetic changes taking place during pollen maturation. The quantitative changes include proteins mainly connected to translation and protein metabolism together with mRNA transport, processing, and ribosome assembly. Interestingly, photosynthesis and plastid translation are among the upregulated protein groups. Quantitative changes included stress response proteins including 32 cold stress, 25 heat stress, 11 virus response and 42 defence response proteins. In general, the MPG proteome is characterized by expression of stress-related proteins and those associated with post-transcriptional and epigenetic regulation of transcription. In summary, early development is characterized by the greater abundance of proteins required for pollen development, and transport and metabolism, than in the later phase, which is followed by up-regulation of mRNA processing, translation regulation, stress proteins and epigenetic regulators.

Comparison of our proteomic dataset with a reference data set of 3491 proteins identified in mature *Arabidopsis* pollen (Grobei et al. 2009) revealed 2791 (79.9%) protein groups in common. Among these, 2345 proteins (83.9%) were detected in mature pollen (riBAQ > 1) with an average riBAQ value of 328.

### Major gene groups for active translation and mRNA storage emerge from the analysis of developmental transcriptomic and proteomic data

To explore the fate of transcripts expressed in developing pollen, we compared transcriptomic and proteomic data in Col-0. There were 4949 shared genes between datasets, and we focussed our analyses on these (Supplementary File 8). k-means clustering of  $\log_2$  scaled TPM (transcripts) and  $\log_2$  scaled riBAQ values (proteins) resulted in nine clusters covering possible relationships of expression profiles in both datasets (Fig. 10A, B, Supplementary File 9). The

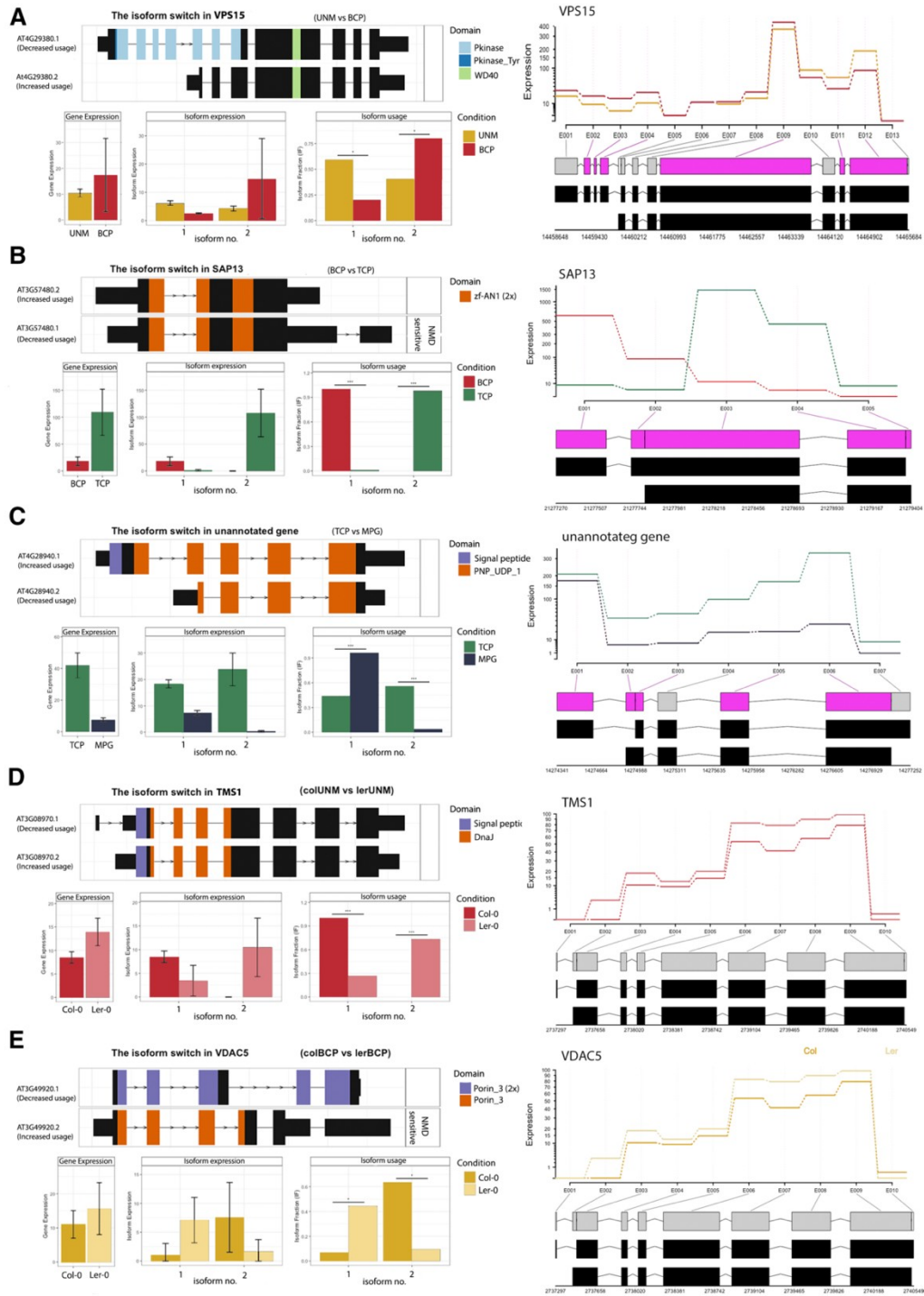
majority (61.6%) of transcripts were present in three clusters (T1; 1,083 genes; T2, 828 genes; T7, 1,137 genes) that declined in abundance during development, while the least represented clusters (T4, 185 genes; T8, 96 genes), which comprised only 5.7% of all transcripts, increased in abundance at late stages of pollen development.

The top three proteome clusters (P1, P5, P6) comprised almost half (46%, 2295 proteins) of all proteins. These either increased in abundance in later stages (P1, 695 proteins), or remained relatively stable (P5, 631 proteins; P6, 969 proteins). The three least represented clusters (P2, P4, P9), which comprised 20.5% of all proteins also showed contrasting profiles, although protein distribution among them was more uniform. P2 (247 proteins) showed abundant and stable expression, P4 (358 proteins) declining in late developmental stages, while P9 (408 proteins) increased in later stages, like P1 (Fig. 10B).

To determine the relationship between transcript and protein cluster pairs, we constructed an overlap matrix (Fig. 10C). To describe the fate of different transcript groups, we selected six patterns of transcriptomic/proteomic overlap that were named A to F (Fig. 10D, Supplementary File 10). Group A includes transcripts present in UNM and BCP, but with limited translation throughout pollen development (Fig. 10D (A)). Group B highlights transcripts with peak expression in early stages but maximum protein accumulation in TCP and MPG, indicating delayed translation (Fig. 10D (B)). Group C comprises transcripts accumulating mostly in later stages and showing limited translation (Fig. 10D (C)). These may include stored transcripts that are translated during pollen germination and tube growth. Groups D and E represent transcripts that are directly translated at early or late developmental stages, respectively (Fig. 10D (D, E)). Group F reflects transcripts with stable, stage-independent expression and direct translation (Fig. 10D (F)). Together, these patterns of overlap illustrate different possible modes of transcript usage.

We examined the biological and molecular processes enriched in each group (Fig. 10D). Group A was enriched for GO categories related to cytoskeleton organization. Other enriched categories included RNA metabolism and processing and posttranscriptional gene silencing by RNA. In addition to mRNA and cell wall connected genes, transcripts involved in meiotic cell cycle and mitotic cell cycle were also enriched in group A.

In Group B, transcripts from other GO categories connected to RNA processing were present, as well as those associated with ribosome biogenesis and translation, with exemplary terms: positive regulation of mRNA splicing via spliceosome, ribosomal small/large subunit biogenesis or protein folding, as well as posttranscriptional gene silencing'. KEGG pathway analysis also showed enrichment for RNA-associated pathways. The connection of group





◀**Fig. 8** Visualization of differential isoform usage (DIU) and differential exon usage (DEU) for selected transcripts. Isoform switches in selected genes are shown on the left, while difference in exon usage is shown on the right. The line plot represents exon usages for each condition (developmental stage), the lower part indicates significantly differentially expressed exons by pink colouring. **A** VPS15 isoform 2 lacking the kinase domain is dominant after UNM-BCP transition. **B** SAP13 isoforms switch during BCP-TCP transition. **C** AT4G28940.1 (unannotated) has a signal peptide and is only expressed in MPG. **D** ERDJ3A isoform 2 is only present in BCP of Ler-0. **E** A different VDAC5 isoform in BCP stage of each accession results in up-regulation of an isoform missing one of the porin domains and is NMD sensitive in Ler-0 DIU statistics and domain presence are shown on the left

B transcripts to RNA processing is further highlighted by enrichment for proteins containing RNA recognition motifs and/or WD domains.

Group C transcripts did not show enriched biological processes, but eight genes were connected to pollination and pollen tube growth, namely *ROP-INTERACTIVE CRIB MOTIF-CONTAINING PROTEIN 3 (RIC3)* and *ROP-INTERACTIVE CRIB MOTIF-CONTAINING PROTEIN 1 (RIC1)*, kinases *POLLEN RECEPTOR LIKE KINASE 6 (PRK6)* and *LOST IN POLLEN TUBE GUIDANCE 2 (LIP2)*, Liu et al. 2013, *CELLULOSE SYNTHASE-LIKE D1, AMINOPHOSPHOLIPID ATPASE 7 (ALA7)* ATPase, *INOSITOL-POLYPHOSPHATE 5-PHOSPHATASE 12 (SPTase12)* and DNAJ domain protein *THERMOSENSITIVE MALE STERILE 1 (TMS1)*. This supports the hypothesis, that group C transcripts are stored for later translation in pollen tubes. KEGG pathways enriched for group C transcripts were endocytosis and glycerophospholipid metabolism. For PFAM domain representation, three of 18 ANTH membrane binding domain proteins were present. Further, several genes were connected to cytoskeleton (microtubule organization), containing, for example, both developmentally regulated plasma membrane polypeptide (DREPP) family proteins *MICROTUBULE-ASSOCIATED PROTEIN 18 (MAP18)* and *MICROTUBULE-DESTABILIZING PROTEIN 25 (MDP25)*, which are reported to act as actin-severing proteins in pollen tubes (Qin et al. 2014).

Group D transcripts included only 15 genes. Two of these genes play a role in the initiation of DNA replication and four are connected to regulation of DNA conformation change.

Group E transcripts comprise 85 genes and do not show enrichment for biological processes, but metabolic pathways and glutathione metabolism were enriched among KEGG pathways.

Group F transcripts comprise 589 genes. Biological process enrichment as well as KEGG pathways point to a connection to energy metabolism, represented by terms such as glutamate catabolic process, citrate metabolic process, mitochondrial ATP synthesis coupled electron transport,

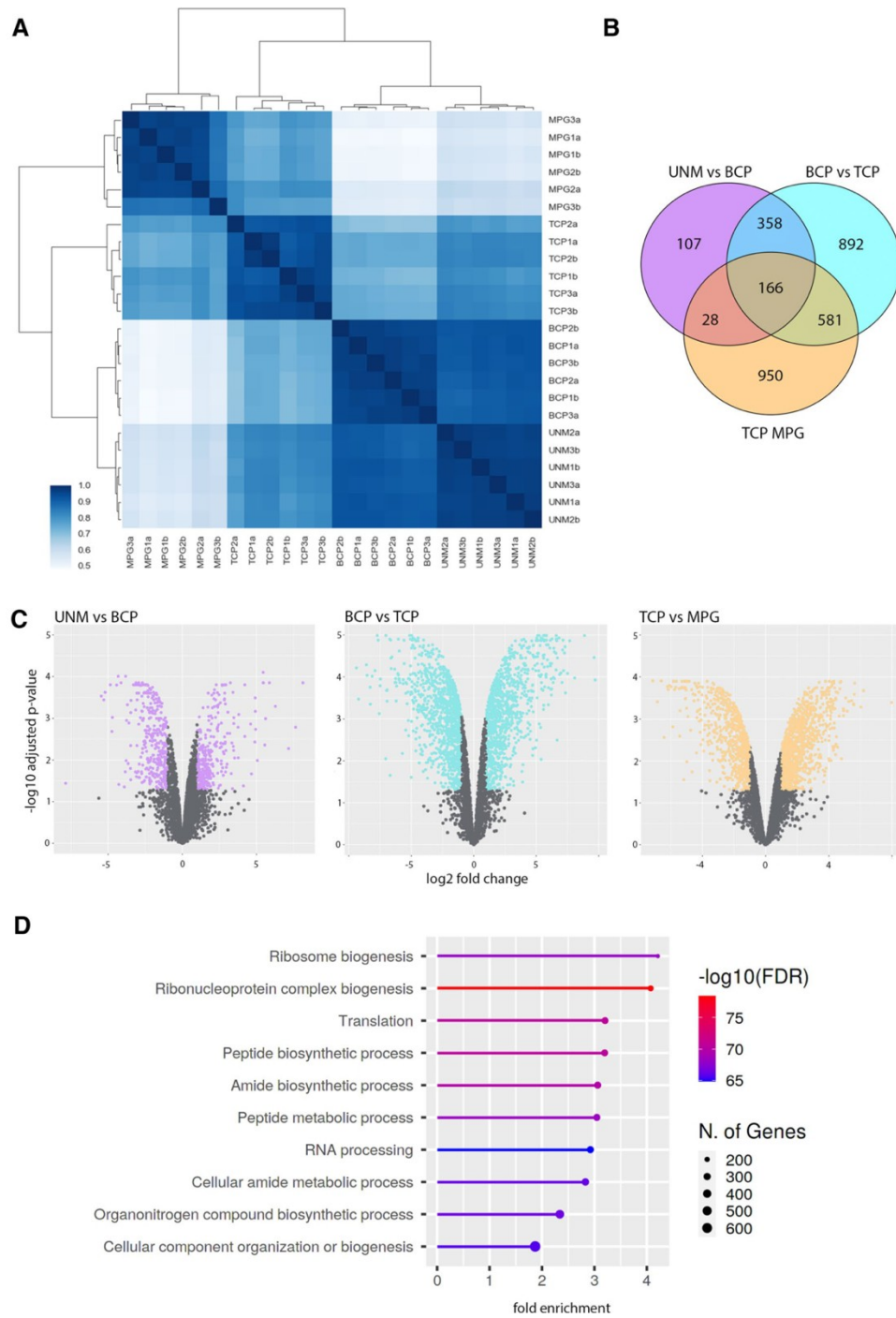
UDP-D-xylose biosynthetic process, glyoxylate cycle or gluconeogenesis. Further connected terms include ribosome assembly and cytoplasmic translational initiation. Identified protein domains included the ribosomal protein L12 family, ribosomal protein P1/P2, N-terminal domain and ribosomal protein L2, domain 2 and five 14-3-3 protein homologues were present.

## Discussion

We used RNA-seq to determine the transcriptomes of isolated microspores and developing pollen at four developmental stages for two *Arabidopsis* accessions. This enabled a more comprehensive qualitative and quantitative analysis, when compared with microarray analysis for the same four developmental stages (Honys and Twell 2004). Our analysis on both platforms in the Ler-0 accession was extended to include Col-0, for which we broadened our study by integrating proteomic data. This allowed us to compare pollen transcriptome dynamics between accessions and to extend the regulatory levels studied.

RNA-seq data was strongly correlated with microarray data suggesting high reproducibility of these methods. RNA-seq provided data for over 33,000 genes, a 1.5-fold increase in the number of genes compared to ATH1 arrays (Honys and Twell 2004). Among these, 6621 protein-coding genes were not previously described as male gametophyte expressed (MGE). The mean expression of these 'new' MGE genes was comparable to that of the whole datasets indicating unexplored roles in male gametophyte development. 606 of them were also supported by the proteome data. New MGE genes were annotated with diverse functions, rather than just a few specific groups or families, providing new information about various biological processes. Further, about 27% (1758 genes) of new MGE genes were unclassified, representing a substantial number of candidates with unknown functions.

We identified numerous long non-coding RNAs, highlighting lncRNAs as good candidates for future studies of potential regulators of pollen gene expression. Although, plant lncRNA research is limited and has focused on root development, auxin signalling or fibre development in cotton, 3 lncRNA discovered in *Oryza sativa*, *Brassica campestris* and *Zea mays* caused male sterility if under-expressed (Datta and Paul 2019). lncRNAs are also implicated in rice ovary meiosis (Li et al. 2020). In pollen, cis-acting natural antisense RNAs (cis-NATs) were identified in sperm and vegetative cells of *Arabidopsis* by re-analysis of published RNA-seq and microarray data (Qin et al. 2018). This study reported 1471 potential protein-coding cis-NAT pairs, from which 872 had at least one member expressed in pollen based on an expression threshold  $\geq 1$  RPM. We found



**◀Fig. 9** Differential protein expression during pollen developmental stage transitions. **A** Pearson correlation of biological and technical sample replicates used in this study. **B** Venn diagram of DEPs. Most abundance differences are stage transition-specific; 166 proteins change their expression during each transition and 581 proteins are differentially expressed between BCP-TCP and TCP-MPG. **C** Volcano plots of expressed proteins in pairwise comparisons of UNM-BCP, BCP-TCP and TCP-MPG. The highest number of DEPs occur during BCP-TCP transition. **D** The 10 most enriched GO terms for DEPs during BCP-TCP transition highlights an increase in translation related proteins

an overlap of 834 genes across all pollen stages with our RNA-seq data (expression threshold > 1 TPM in at least one stage). One of the most highly expressed NATs in both datasets was At1g08727.1, which overlaps *TUA1* (At1g64740.1), a pollen-specific *ALPHA-1 TUBULIN* (Carpenter et al. 1992). NAT At1g08727.1 shows the highest expression in late pollen development stages. One lncRNA discovered in our dataset is FLORE (At1g69572), a *cis*-acting natural antisense transcript of *CYCLING DOF FACTOR 5 (CDF5)*, which forms an antagonistic pair with role in circadian regulation and flowering time (Henriques et al. 2017). FLORE may act as a late pollen regulator as it is not expressed in UNM and BCP, but it has peak expression (33 TPM) in TCP. Further examination of genes overlapping expressed natural antisense non-coding RNAs led to the discovery of proton pump interactor 1 (PPI1, At4g27500) and NAT At4g07855, both highly expressed throughout pollen development. PPI1 encodes a 14-3-3 domain protein interacting with H + ATPase. Overall, the dynamic expression of different lncRNA groups suggests functional roles in developing pollen including the modulation of protein-coding RNA expression.

### Transcriptome dynamics during pollen development

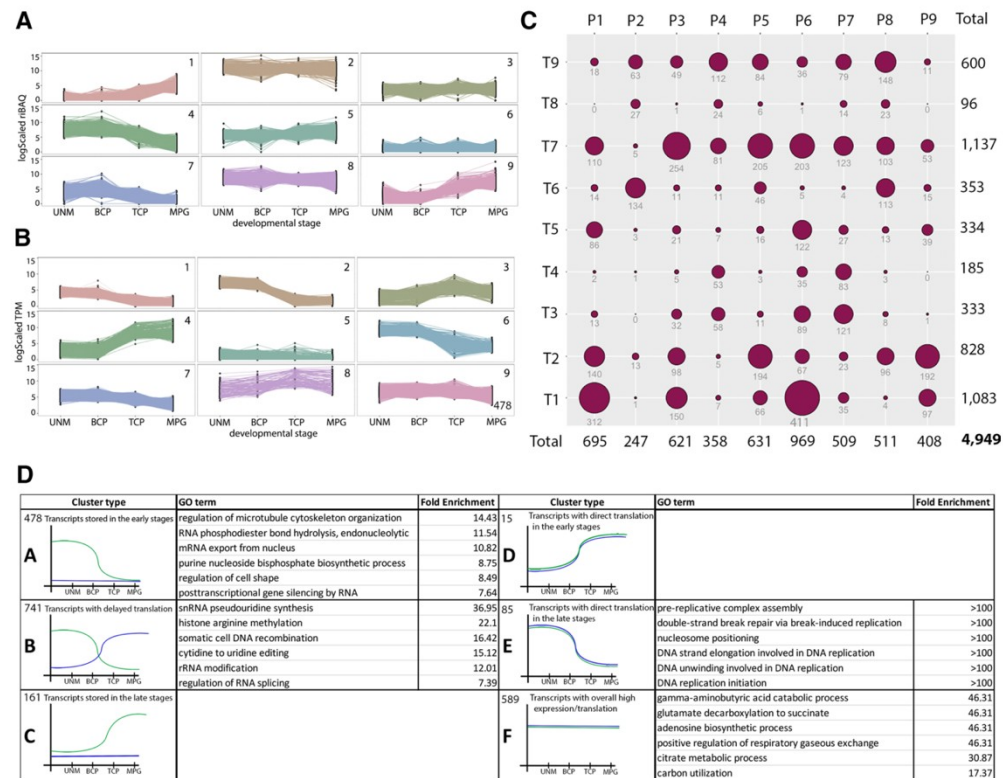
RNA-seq data were analysed and compared across pollen developmental stages as well as between accessions. The differentially expressed genes were categorized by GO enrichment, KEGG pathway and String analyses. The major functional changes are summarized in Fig. 11. Briefly, UNM and BCP stages differ in the number and abundance of mitosis- and cell cycle-related transcripts. In a previous microarray analysis 61 transcripts associated with cell cycle, 45 were detected in developing pollen, with the majority expressed in UNM and BCP (Honys and Twell 2004). For Col-0 RNA-seq data, 60 of these genes were predominantly expressed in early stages, along with a further eight new cell cycle genes (Supplementary File 11). Genes that show increased expression during UNM-BCP transition include early pollen tube and cell wall organization transcripts. Transmembrane transport and glucan metabolism processes were also well represented. In accord with previous data from *Arabidopsis*

(Honys and Twell 2004) and tobacco (Hafidh et al. 2018), translation initiation factors (eIFs) were mostly expressed during early stages.

Although the generative cell lineage identity is established after PMI in BCP the transcriptomes of the two early stages remain similar. This may be explained in part by the limited contribution of the generative cell to the BCP transcriptome relative to the vegetative cell (Honys and Twell 2004). Expression of several pollen tube growth and regulation-associated genes starts in the BCP, suggesting developmental priming for pollen germination. Developmental priming to prepare for either potential heat stress response or stage transition has been described during pollen development in tomato (Chaturvedi et al. 2013; Chaturvedi et al. 2016). The majority of pollen tube growth connected genes upregulated in BCP continue to increase in expression at TCP stage, therefore another explanation could be the cumulative effect of their expression throughout pollen development, as two candidates pectinesterase 5 (PME5, At2g47040) and pectinesterase 1 (PME1, At1g69940) increase nearly tenfold during UNM-BCP and BCP-TCP transitions. On the contrary, myosin 11 C2 (At1g54560) and myosin11 C1 (At1g08730), described to be essential for the high rate of pollen tube growth (Madison et al. 2015), show only a slight increase (At1g08730) or a decrease (At1g54560) in expression after BCP stage, suggesting their early expression as developmental priming.

The current study highlights the major transcriptomic switch between BCP and TCP stages consistent with previous analyses (Honys and Twell 2004; Twell et al. 2006). As a well-known example, protein synthesis genes are transcribed almost exclusively during the early phases of microgametogenesis. We also described a major decline of transcripts encoding ribosomal proteins during BCP-TCP transition which correspond to previous findings (Hafidh et al. 2018; Honys and Twell 2004). The massive synthesis of ribosomal proteins (RP) in tobacco pollen soon after the completion of PMI, is demonstrated by the abundance of RP transcripts (Bokvaj et al. 2015) and by their association with polysomes (Hafidh et al. 2018). This highlights active translation at this stage and persistence of the translation apparatus during late pollen development and the progametic phase (Hafidh et al. 2018; Hafidh and Honys 2021). The profiling and distribution of RP transcripts and proteins identified in this study suggest that the same mechanism of RP translation during early pollen development is active in *Arabidopsis*.

The start of the late developmental phase (TCP) is marked by up-regulation of transcripts involved in vesicular transport, pH regulation and energy metabolism. These transcripts may have roles later in pollen tube growth and regulation, where vesicular transport and formation of a pH gradient are essential for rapid PT growth. In addition, the number of transcripts annotated as ‘pollen developmental’



**Fig. 10** Integration of transcriptomic and proteomic data in developing Col-0 pollen. **A** k-means clustering of expression of 4,949 proteins. **B** k-means clustering of expression for 4,949 transcripts. **C** Matrix indicating numbers of overlapping genes in transcripts and protein clusters. **D** Transcript clusters distinguished according to

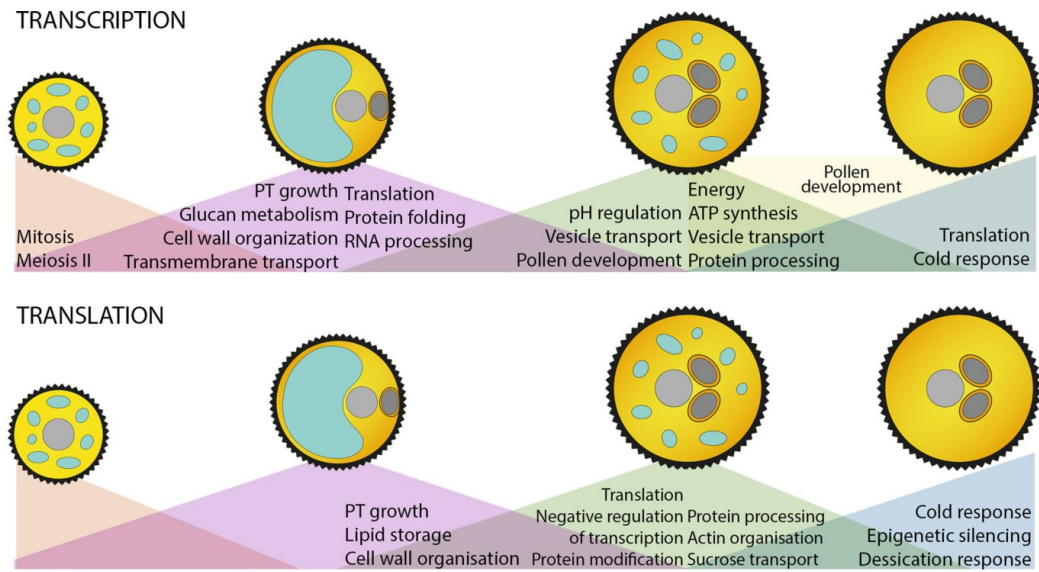
RNA and protein profiles. Plots indicate relative transcript (green) and protein (blue) expression for each cluster. The numbers of transcripts in each cluster and the six most enriched GO terms and their fold enrichment are listed

increases steadily from early to later stages. Some transcripts encoding pollen tube regulatory proteins are predominantly expressed in MPG stage and appear to represent a specific class of late pollen tube mRNAs.

The final developmental stage is distinguished by an increase in the expression of stress response genes including cold and salt stress related proteins, which probably play protective role during the desiccation process of angiosperm pollen which typically undergoes dehydration and developmental arrest prior to dispersal (Franchi et al. 2011). In seeds, the major stress-like condition accompanying dehydration can cause DNA damage including strand breakage or telomeric shortening (Osborne and Boubriak 2002). Among DEGs detected during TCP-MPG transition there were 53 genes involved in DNA repair, including DNA ligase 1

responsible for double-strand breakage repair and the telomerase maintenance genes HOMOLOG OF X-RAY REPAIR CROSS COMPLEMENTING 3 (XRCC3) and RAD50 (Gallego and White 2001; Bleuyard and White 2004).

Protein products of genes upregulated in MPG were distinctively associated with plasma membrane, cell wall and apoplast. These compartments play important roles in pollen germination and in pollen tube guidance, reception, and fertilization, through signalling and crosstalk with female reproductive tissues (Hafidh et al. 2016; Hafidh and Honys 2021; Johnson et al. 2019), cell wall remodelling or stress response (reviewed in Ge et al. 2011). The most prevalent stress-related group, accounting for 60 genes upregulated in MPG, are genes responsive to drought tolerance. Three DELLA repressor genes are upregulated during TCP-MPG



**Fig. 11** Summary of major changes in biological processes in developing *Arabidopsis* pollen. GO terms reflect the analysis of differential gene expression based on transcriptomic and proteomic data at four

developmental stages. Coloured triangles indicate increased expression (left) or decreasing expression (right) of GO processes during stage transitions

transition. In general DELLA proteins work as growth repressors, while DELLA-3 is reported to play role in protein storage in seeds (Hu et al. 2021). Although pollen dehydration is a crucial phase in maturation the underlying mechanisms are not well described, and this study could provide suitable candidate genes for further studies.

Apart from the general trends, we focused on a few selected groups. We categorized transcription factor families and the data are supported by previous findings. For example, 17 MADS-box genes were reported to be expressed in mature pollen (Pina et al. 2005), 16 of which were present in our MPG datasets. Similarly, our results follow the observed divergent expression of TFs, describing early, constitutive, and late TFs (Honys and Twell 2004). The V5 plant transcription factors database (PTFD) (Mitsuda and Ohme-Takagi 2009), which was used for annotation of the RNA-seq data, accounts for 127 bZIP and 225 bHLH TFs, an increase of 1.7-fold in each case. Thus, our data substantially extends knowledge of TF expression. Similarly, ribosomal, stress-related, cell wall-connected and other transcript group were sorted according to temporal expression enabling examination of any of these groups of genes.

When we examined transcriptome differences between Col-0 and Ler-0, although around 3,000 DEGs were found at each stage, no major transcript groups or pathways differed between accessions. The observed differences in expression

of individual genes may therefore be genotype-dependent and influenced by the experimental conditions. Although there is limited information about differences in pollen gene expression between *Arabidopsis* accessions, we observed reduced levels of the *AtSUC1* transcripts in Ler-0 compared with Col-0, consistent with reduced immunodetection of *AtSUC1* in mature pollen of Ler and C24 ecotypes (Feuerstein et al. 2010). In future, studying protein abundance of sucrose transporters between *Arabidopsis* accessions could prove interesting for deciphering ecotype differences in pollen tube germination.

#### Identification of transcript isoform switches during pollen development

The differential usage of transcript isoforms is common in plants, with around 22% of AS events resulting in changes to protein sequences in *Arabidopsis* (Vanechoutte et al. 2017). Among AS events, intron retention is the most frequent, but AS can also result in protein sub-functionalization (Ner-Gaon et al. 2004). Keller et al. 2017 reported 5000 to 8000 genes showing intron retention during pollen development for two tomato cultivars. Intron retention was also reported to be the predominant AS mechanism in developing pollen of *Brassica rapa* (Goliez et al. 2021). Further, intron retention was found to be stage-specific, similar to

our findings in *Arabidopsis*, and IR may thus represent an important mechanism of functional attenuation of related subsets of genes (Golicz et al. 2021). In a previous analysis, isoform usage between mature pollen and seedlings of *Arabidopsis* was reported to be similar, with few pollen-specific changes that were mostly connected to transcripts involved in splicing (Loraine et al. 2013). In our more comprehensive study, numerous transcript isoforms were detected that varied between UNM-BCP and BCP-TCP transitions. Their categorization highlighted processes that were diversified by mRNA processing, including splicing. In all stage comparisons, differential isoform usage was discovered in a variety of genes, including transporters, transcription factors, enzymes, heat stress proteins or ribosomal proteins. The functional consequences of isoform switching included loss or gain of signal peptide, domain presence or change in NMD sensitivity. In UNM-BCP, the switches may be connected to the regulation of PMI. Relevant candidates include kinesin-like proteins KIN12B (At3g23670) and KIN-12A (At4g14150) which link microtubules in the phragmoplast during plant cell division (Vanstraelen et al. 2006). It suggests IR to be an effective way to switch off cytokinetic transcripts after PMI. The highest frequency of isoform switching was detected during BCP-TCP transition when differential expression changes are also maximum. The genes with DIU in this stage transition were enriched for processed connected to mRNA and protein processing. Also, vesicle transport and GTPase activity connected genes were enriched suggesting a regulation of transcripts of proteins required for the rapid pollen tube growth. In the TCP to MPG transition, we observed enrichment for transporters. In summary, pollen development is accompanied by numerous changes in isoform usage specific for each developmental phase.

Although differences in AS events were less frequent between Col-0 and Ler-0, hundreds of switches were present at each stage. Despite originating from the same parental lineage, these accessions show differences in physiological and morphological traits (Passardi et al. 2007). The reported changes in isoform usage could provide leads for understanding the differences in gene regulation between these accessions. For example, a different dominant isoform of UBQ14 and UBQ4 are present in each accession at TCP and MPG stages, and both ubiquitins influence development and environmental response (Sun and Callis 1997). Another candidate with DIU is VOLTAGE DEPENDENT ANION CHANNEL 5 (VDAC5), which lacks a porin domain I in Ler-0 and the dominant isoform is NMD sensitive (Fig. 8E). VDAC5 is reported to carry single nucleotide polymorphisms in geographically restricted populations of the Iberian Peninsula, indicating potential involvement in environmental adaptation (Tabas-Madrid et al. 2018). However, analysis of AS in other plant organs will be needed to provide a perspective of its

potential contribution to the phenotypic differences between these accessions.

### Proteomic dynamics during pollen maturation

Previous proteomic studies have described the major portion of the *Arabidopsis* mature pollen proteome to be involved in protein synthesis, cytoskeleton organization, metabolism, cellular transport and signalling (Ge et al. 2011; Grobei et al. 2009). Transcripts involved in energy metabolism and protein synthesis were reported to be either directly translated at high rate (energy) or transcribed in early stages and translated during late stages (Holmes-Davis et al. 2005). Our developmental proteomic data support and extend these findings by integration with transcriptomic data. Proteome changes during UNM-BCP are the least dramatic and involve cell wall organization, lipid storage and pollen tube developmental (Fig. 11). In accord with transcriptome results, the main shifts in protein synthesis occur during BCP-TCP transition, with TCP as the start point for translation of the protein synthesis machinery and the reduction of transcript abundance. Cytoskeleton organization proteins are also enriched in the later developmental stages.

In the MPG proteome, stress tolerance and epigenetic modification processes are over-represented compared to previous stages. Male gametophyte development involves extensive epigenetic reprogramming (Ashapkin et al. 2019; Borg et al. 2020) and our data show high levels of accumulation of AGO1 (At1g48410) and NRPD1B (At2g40030) in mature pollen. These proteins are active in 21nt siRNA production in the vegetative nucleus. Interestingly, despite the limited contribution of the sperm cell proteome, we also detected increased expression of AGO5 (At2g27880) and AGO9 (At5g21150), a component of the Argonaute complex responsible for de novo methylation via RNA directed DNA methylation and probably siRNA-mediated transposon silencing in sperm cells (Ashapkin et al. 2019). The early peak in transcripts for AGO1, NRPD1B and AGO9 (AGO5 RNA peaks in TCP), highlight these as examples of transcripts with delayed translation during pollen development.

We compared our developmental pollen proteome data with major studies in *Solanum lycopersicum* (Chaturvedi et al. 2013) and *Nicotiana tabacum* (Ischebeck et al. 2014). In tomato 1104 proteins were detected in mature pollen and for proteins well conserved between tomato and *Arabidopsis* (>90% NCBI BLAST similarity scores), we identified 194 homologous *Arabidopsis* genes. Among these, there was an overlap of 70.6% (137 proteins) with our developmental proteome and 130 had riBAQ values higher than 1 in mature pollen, with 255 on average. The tobacco study described 3888 proteins from 8 developmental stages spanning male meiocytes to pollen tubes (Ischebeck et al. 2014). For the 1478 unique tobacco proteins with a homologue in the

*Arabidopsis*, 1348 proteins were identified in our analysis, highlighting the similarity between proteomes of developing pollen of these two species. In total, 111 proteins were shared between *Arabidopsis*, tobacco, and tomato (Supplementary File 12). In summary, *Arabidopsis* pollen developmental proteomes show high similarity with published proteome data for mature pollen of *Arabidopsis* (Grobei et al. 2009), tobacco and tomato. In addition, our proteomic analysis identified 2150 proteins, which have not been previously detected in the male gametophyte (Supplementary File 13).

### Coupling of transcriptome and proteome data reveals mRNA fate in pollen development

Post-transcriptional regulation of mRNAs in developing pollen is a crucial process for understanding male reproductive development. Apart from mRNA storage, there have been several attempts to decipher the relationship between transcription and translation in developing pollen. In previous work (Honys and Twell 2004) and in this article, we showed that most transcripts responsible for PT organization appear at bicellular stage. These transcripts are stored to allow rapid translation during pollen activation and germination (Hafidh et al. 2018). Previous comparisons of *Arabidopsis* MPG proteomic data with transcriptomic data have uncovered inverse relationships between mRNA and protein abundance (Holmes-Davis et al. 2005). These included energy-related genes in contrast to cell wall organization genes where mRNAs were more abundant relative to proteins (Holmes-Davis et al. 2005). They suggested that the energy-related proteins are stored in advance of resumption of metabolic activities upon pollen activation, whereas cell wall proteins are translated to support PT growth. Similarly, transcripts with direct and delayed modes of translation in developing pollen of tomato were suggested based on transcriptomic and proteomic data (Keller et al. 2018).

We addressed mRNA fate and the timing of translation separately by clustering developmental proteome and transcriptome data. Using k-means clustering, we divided the genes among seven groups defined by the relationship between gene expression and translation. The most abundant group consisted of stably transcribed and translated energy and metabolism connected genes. Early-stage synthesized proteins were mainly focused on replication. Genes with delayed translation or with stored transcripts are connected to RNA processing and cytoskeleton organization. The later may account for the pool of stored mRNA to support rapid PT growth upon pollen germination (Hafidh et al. 2016; Honys and Twell 2004; Rutley and Twell 2015). The direct link between transcription and translation during pollen development could be further addressed by studying the translome of active polysomes and monosomes. This approach has been reported for in vitro and in vivo pollen

tubes (Lin et al. 2014) including a recent study of the heat stress response (Poidevin et al. 2020).

### Conclusions

We analysed transcriptome and proteome dynamics accompanying four developmental stages of microgametogenesis in the Col-0 and Ler-0 accessions of *Arabidopsis*. RNA-seq and up to date genome annotation enabled us to extend the coverage and resolution of previous microarray analyses qualitatively and quantitatively. We demonstrated high reproducibility and comparability of both transcriptomic platforms. RNA-seq also allowed us to detect thousands of lncRNAs and their dynamics as potential regulators of pollen development. Similarly, we described numerous alternative splicing events in developing pollen and identified candidate transcripts regulated predominantly by intron retention. To understand mRNA fate and translation dynamics, we compared transcriptomic and proteomic data and proposed transcript groups based on their potential temporal translation. Overall, this work provides an integrated perspective of gene expression dynamics in developing pollen and a foundation for exploration of the role of alternative splicing in the male gametophyte of *Arabidopsis thaliana*.

### Methods

#### Plant cultivation and isolation of microspores and developing pollen

Plants of *Arabidopsis thaliana* accession Columbia-0 (Col-0) and Landsberg erecta (Ler-0) were grown in controlled-environment chambers at 22 °C with a 16-h photoperiod and illumination of 150  $\mu\text{mol}/\text{m}^2/\text{sec}$ . Mature pollen grains (MPG) were collected either with a modified vacuum cleaner using 100, 53 and 5  $\mu\text{m}$  mesh and/or by shaking of cut inflorescences of 5- to 6-week-old plants in 0.1 M mannitol as described (Duplakova et al. 2016). Populations of uninucleate microspores (UNM), bicellular pollen (BCP) and tricellular pollen (TCP) were released from anthers of closed flower buds and separated by Percoll density gradient centrifugation as described (Duplakova et al. 2016; Honys and Twell 2004).

#### RNA extraction, cDNA library preparation and sequencing

Total RNA was isolated from populations of isolated microspores or developing pollen (UNM, BCP, TCP and MPG) using the RNeasy Plant Kit (Qiagen) following the manufacturer's instructions. RNA was Dnase-treated (DNA-free™

Kit Ambion, Life Technologies) according to the manufacturer's protocol. The yield and purity of RNA were determined spectrophotometrically using an Agilent 2100 Bioanalyzer. A slightly modified SmartSeq2 protocol was used to synthesize cDNA from poly(A) + RNA with an oligo(dT)-tailed primer (Picelli et al. 2013, 2014). A low-input Nextera protocol (Baym et al. 2015) was used to prepare the final libraries, which were sequenced on a NextSeq500 instrument with single-end 75 bp read length (SE75).

### RNA-seq data processing, mapping and assembly of reads

The quality of single-end raw reads was revised by FastQC ver. 0.11.8 (Wingett and Andrews 2018), and Cutadapt ver. 1.9.1 (Martin 2011). The quality reads (phred score > 20) were trimmed of technical sequences using the same Cutadapt software and mapped to the *A. thaliana* reference genome (ver. TAIR10) downloaded from Araport (Pasha et al. 2020) with STAR ver. 206.1a (Dobin et al. 2013). The gtf annotation file obtained from Araport was used for STAR index preparation. Gene and isoform counts (including normalized TPM values) were acquired with RSEM (Li and Dewey 2011). The expression threshold for RNA-seq data was set to 3 TPM (Transcripts Per Kilobase Million) for each biological replicate. The data were imported into Rstudio with Tximport (Soneson et al. 2015) and processed further. For differential expression analysis DESeq2 ver. 3.8 (Love et al. 2014) was used with adjusted *p* values < 0.05 and FoldChange  $\geq \pm 2$  as thresholds for establishing differentially expressed genes (DEGs). For analysis of differential exon usage, DexSeq (Anders et al. 2012; Reyes et al. 2013) was used with the STAR transcriptome alignment as input. To obtain differential isoform usage the isoform level quantification output of RSEM was processed by IsoformSwitchAnalyser v. 4.1 (Vitting-Seerup and Sandelin 2017; Vitting-Seerup et al. 2019). The thresholds for both gene and isoform expression were set to 3 (function PreFilter). Adjusted *p* values of < 0.05 and  $\log_2$  fold change > 1 (diF > 0.1 for DIU) were used as statistically significant thresholds in both analyses. Domain switch consequences were analysed according to Pfam v. 33.1 (Mistry et al. 2021) and coding potential was established with CPC2 calculator (Kang et al. 2017). Signal peptide presence was analysed with SignalP-5.0 (Almagro Armenteros et al. 2019).

### Comparison of ATH1 data and RNA-seq data

Affymetrix ATH1 genome array data were MAS5 normalized with the exclusive approach, such that expressed genes have a present detection call in both replicates (Duplakova et al. 2016; Honys and Twell 2004). MAS5 normalized hybridization signals from expressed genes were compared

to normalized TPM values for RNA-seq data. Any ambiguous probe sets representing gene models unique to ATH1 arrays were removed from our analyses (Supplementary File 14).

### Protein extraction and LC–MS/MS analysis

Proteins were isolated from UNM, BCP, TCP and MPG using TRI Reagent solution (Sigma-Aldrich, product No. T9424) following the manufacturer's instructions. Individual protein samples were processed by filter-aided sample preparation (FASP) with modifications. The samples were mixed with 8 M UA buffer (8 M urea in 100 mM Tris–HCl, pH 8.5), loaded onto the Microcon device with MWCO 30 kDa (Merck Millipore) and centrifuged at  $7000 \times g$  for 30 min at 20 °C. The retained proteins were washed (all centrifugation steps after sample loading done at  $14,000 \times g$ ) with 200  $\mu$ L UA buffer. The final protein concentrates kept in the Microcon device were mixed with 100  $\mu$ L of UA buffer containing 50 mM iodoacetamide and incubated in the dark for 20 min. After the next centrifugation step, the samples were washed three times with 100  $\mu$ L UA buffer and three times with 100  $\mu$ L of 50 mM NaHCO<sub>3</sub>. Trypsin (sequencing grade, Promega) was added onto the filter and the mixture was incubated for 18 h at 37 °C (enzyme:protein ratio 1:100). The tryptic peptides were finally eluted by centrifugation followed by two additional elutions with 50  $\mu$ L of 50 mM NaHCO<sub>3</sub>. Directly after FASP, peptides were extracted into LC–MS vials by 2.5% formic acid (FA) in 50% acetonitrile (can) and 100% ACN with addition of polyethylene glycol (20,000; final concentration 0.001%) and concentrated in a SpeedVac concentrator (Thermo Fisher Scientific) prior LC–MS analyses. LC–MS/MS analyses of all peptide mixtures were done using Ultimate 3000 RSLCnano system (SRD-3400, NCS-3500RS CAP, WPS-3000 TPL RS) connected to Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific). Prior to LC separation, tryptic digests were online concentrated and desalted using trapping column (100  $\mu$ m  $\times$  30 mm) filled with 3.5- $\mu$ m X-Bridge BEH 130 C18 sorbent (Waters). After washing of the trapping column with 0.1% FA, peptides were eluted (flow rate -300 nL/min) onto an analytical column (Acclaim Pepmap100 C18, 3  $\mu$ m particles, 75  $\mu$ m  $\times$  500 mm; Thermo Fisher Scientific) with a 100 min nonlinear gradient program (1–56% of mobile phase B; mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 80% ACN). Equilibration of the trapping column and the column was done prior to sample injection to sample loop. The analytical column outlet was directly connected to the Digital PicoView 550 (New Objective) ion source with sheath gas option and SilicaTip emitter (New Objective; FS360-20-15-N-20-C12) utilization. ABIRD (Active Background Ion Reduction Device, ESI Source Solutions) was installed.



MS data were acquired in a data-dependent strategy selecting up to top 10 precursors based on precursor abundance in the survey scan (350–2000  $m/z$ ). The resolution of the survey scan was 60 000 (400  $m/z$ ) with a target value of  $1 \times 10^6$  ions, one microscan and maximum injection time of 200 ms. HCD MS/MS (32% relative fragmentation energy) spectra were acquired with a target value of 50 000 and resolution of 15 000 (at 400  $m/z$ ). The maximum injection time for MS/MS was 500 ms. Dynamic exclusion was enabled for 45 s after one MS/MS spectra acquisition and early expiration was disabled. The isolation window for MS/MS fragmentation was set to 2  $m/z$ .

### Processing of proteomic data

The analysis of the mass spectrometric RAW data files was carried out using MaxQuant software (version 1.6.0.16). MS/MS ion searches were conducted against the modified cRAP database (The common Repository of Adventitious Proteins, cRAP database) containing protein contaminants such as keratin and trypsin, and the UniProtKB protein database for *Arabidopsis thaliana* (see ref. UniprotKB for the ftp server address, the number of protein sequences was 27,567). Default precursor and fragment mass tolerances were used with software MS data recalibration enabled. Oxidation of methionine and proline, deamidation (N, Q) and acetylation (protein N-terminus) as optional modification, carbamidomethylation (C) as fixed modification and one enzyme miss cleavage were set. Peptides and proteins with FDR threshold  $< 0.01$  and proteins having at least one unique or razor peptide were reported by MaxQuant only. Match between runs among all analyses and second peptide identification features of MaxQuant were enabled. Protein abundance was assessed using protein intensities calculated by MaxQuant. Limma R package was used for protein group intensities normalization (loessF) and statistical significance testing of differences between individual stages. Differently expressed proteins (DEPs) were ascertained based on the limma results as follows: (1) protein groups having  $\log_2$  fold change  $> |1|$ , adjusted p value (Benjamini–Hochberg procedure)  $< 0.05$  and quantified in at least 2 replicates were considered as quantitatively changing between the stages compared; (2) qualitative changes were considered separately and contained protein groups being quantified in at least two replicates of one stage and absent in the other one.

### Annotation and enrichment analyses

The lists of DEGs and DEPs were annotated with gene names and symbols derived from ThaleMine v5.0.2 (Pasha et al. 2020). GO Enrichment for biological processes, cellular content and molecular functions was analysed by Panther16.0 (Thomas et al. 2003) with Fisher's Exact test with

False discovery rate (FDR) correction. ReviGo (Supek et al. 2011) was used to summarize GO enrichment analyses and to visualize the top enriched terms. The results were further visualized in MapMan v. 3.5.1R2 (Thimm et al. 2004) Enriched KEGG pathways were calculated and rendered with PathView (Luo et al. 2009, 2017) using default settings. Data processing and sorting was executed in Microsoft Excel and RStudio with ggplot2, venn.diagram and enhanced volcano R packages used for plots rendering.

### k-means clustering

$\log_2$  scaled RiBaq values for proteome analysis and  $\log_2$  scale TPM values for transcriptome analysis were used as input data. Only genes present in both analyses were processed. The number of clusters suitable for the size of the datasets was determined with the Elbow method. This resulted in the division of both datasets into nine clusters according to expression pattern across the four developmental stages. K-means clustering was calculated in R with the tidyverse and maggrire packages.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00497-022-00452-5>.

**Acknowledgements** BK was supported by the Czech Science Foundation Grant Number 19-01723S and Czech Ministry of Education, Youth and Sports grant number LTC20050. LS and DH were supported by the Czech Science Foundation Grant Number 21-15856S and Czech Ministry of Education, Youth and Sports grants number LTC20028 and LTAUSA18115. CM was supported by a European Regional Development Fund project through the Centre for Experimental Plant Biology (CZ.02.1.01/0.0/0.0/16\_019/0000738) and by the Czech Ministry of Education, Youth and Sport Grant Number LTC18034. CHSB, Instruct-CZ Centre of Instruct-ERIC EU consortium, funded by MEYS CR infrastructure project LM2018127, is gratefully acknowledged for support of measurements at the CEITEC Proteomics Core Facility. Computational resources were supplied by the project e-Infrastruktura CZ (e-INFRA LM2018140) within the program Projects of Large Research, Development and Innovations Infrastructures. ZZ was supported by European Regional Development Fund-Project, SINGING PLANT (No. CZ.02.1.01/0.0/0.0/16\_026/00 08446). DHa was supported by ERA-CAPS UK Biotechnology and Biological Research Council grant BB/N005090 awarded to DT. JDB received salary support from Fundação para a Ciência e a Tecnologia (FCT) through an Investigador FCT grant. ACL was supported by FCT grant ERA-CAPS-0001-2014 to JDB. The IGC's Genomics Unit is partially supported by projects LISBOA-01-0246-FEDER-000037 Single cell HUB and LISBOA-01-0145-FEDER-016417 (ONEIDA) co-funded by Programa Operacional Regional Lisboa 2020.

**Author contribution statement** D.H. and D.T. conceived the study. B.K. analysed RNA-seq data, performed the functional characterization of proteome data and k-means clustering analyses. B.K., D.H. and D.T. wrote the manuscript. L.S. and C.M. isolated RNA and proteins from pollen stages of Col-0. DHa prepared Ler-0 RNA samples. Z.Z., D.P. prepared and conducted LC-MS measurements and processed and analysed the MS datasets. J.D.B. and A.C.L. prepared RNA libraries and sequenced the samples on Illumina platform.

**Data availability** The RNA-seq data generated and analysed during the current study are available in ArrayExpress with accession code E-MTAB-9456. The proteome data generated and analysed are available in PRIDE with accession code PXD033305. Affymetrix data analysed during this study are included in Honys and Twell (2004) (<https://doi.org/10.1186/gb-2004-5-11-r85>) Web-based queries of the RNA-seq data can be made at the EVOREPRO database ([www.evorepro.plant.tools](http://www.evorepro.plant.tools)) as described in Julca et al. 2021. (<https://doi.org/10.1038/s41477-021-00958-2>).

## Declarations

**Conflict of interests** The authors declare that they have no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adamczyk BJ, Fernandez DE (2009) MIKC\* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiol* 149(4):1713–1723. <https://doi.org/10.1104/pp.109.135806>
- Aklilu BB, Peurois F, Saintomé C et al (2020) Functional diversification of replication protein a paralogs and telomere length maintenance in Arabidopsis. *Genetics* 215(4):989–1002. <https://doi.org/10.1534/genetics.120.303222>
- Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-Seq Data. *Genome Res* 22(10):2008–2017. <https://doi.org/10.1101/gr.133744.111>
- Armenteros A, Juan J, Tsirigos DK, Sønderby KC et al (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37(4):420–23. <https://doi.org/10.1038/s41587-019-0036-z>
- Ashapkin VV, Kutueva IK, Aleksandrushkina IN, Vanyushin FV (2019) Epigenetic regulation of plant gametophyte development. *Int J Mol Sci*. <https://doi.org/10.3390/ijms20123051>
- Azarov AS, Tokarev BI, Netchepurenko AE (1990) Effect of sodium chloride on pollen germination and pollen tube growth in vitro in *Arabidopsis thaliana* (L.) Heynh. *Arabidopsis Inf Serv* 27:9–12
- Bach-Pages M, Homma F, Kourelis J et al (2020) Discovering the RNA-binding proteome of plant leaves with an improved RNA interactome capture method. *Biomolecules* 10(4):1–21. <https://doi.org/10.3390/biom10040661>
- Baym M, Kryazhimskiy S, Lieberman TD et al (2015) Inexpensive multiplexed library preparation for Megabase-sized genomes. *PLoS ONE* 10(5):1–15. <https://doi.org/10.1371/journal.pone.0128036>
- Becker JD, Boavida LC, Carneiro J, Haury M, Feijó JA (2003) Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiol* 133(2):713–725
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015) The Arabidopsis information resource: making and mining the ‘Gold Standard’ annotated reference plant genome. *Genesis* 53(8):474–485. <https://doi.org/10.1002/dvg.22877>
- Bleuyard JY, White CI (2004) The Arabidopsis homologue of Xrcc3 plays an essential role in meiosis. *EMBO J* 23(2):439–449. <https://doi.org/10.1038/sj.emboj.7600055>
- Boavida LC, McCormick S (2007) Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. *Plant J* 52(3):570–582. <https://doi.org/10.1111/j.1365-3113X.2007.03248.x>
- Bock KW (2006) Integrating membrane transport with male gametophyte development and function through transcriptomics. *Plant Physiol* 140(4):1151–1168. <https://doi.org/10.1104/pp.105.074708>
- Boisson-Dernier A, Roy S, Kritsas K et al (2009) Disruption of the pollen-expressed FERONIA Homologs ANXUR1 and ANXUR2 triggers pollen tube discharge. *Development* 136(19):3279–3288. <https://doi.org/10.1242/dev.040071>
- Bokvaj P, Hafidh S, Honys D (2015) Transcriptome profiling of male gametophyte development in *Nicotiana tabacum*. *Genom Data* 3:106–111. <https://doi.org/10.1016/j.gdata.2014.12.002>
- Borg M, Jacob Y, Susaki D et al (2020) Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nat Cell Biol* 22(6):621–629. <https://doi.org/10.1038/s41556-020-0515-y>
- Cabada Gomez DA, Chavez MI, Cobos AN et al (2020) COPI complex isoforms are required for the early acceptance of compatible pollen grains in *Arabidopsis thaliana*. *Plant Reprod* 33(2):97–110. <https://doi.org/10.1007/s00497-020-00387-9>
- Cao Y, Han Y, Meng D et al (2017) B-BOX Genes: genome-wide identification, evolution and their contribution to pollen growth in pear (*Pyrus bretschneideri* Rehd.). *BMC Plant Biol* 17(1):1–12. <https://doi.org/10.1186/s12870-017-1105-4>
- Carpenter JL, Ploense SE, Snustad DP, Silflow CD (1992) Preferential expression of an  $\alpha$ -Tubulin gene of Arabidopsis in pollen. *Plant Cell* 4(5):557–571. <https://doi.org/10.1105/tpc.4.5.557>
- Chaturvedi P, Ischebeck T, Egelhofer V, Lichtscheid I, Weckwerth W (2013) Cell-specific analysis of the tomato pollen proteome from pollen mother cell to mature pollen provides evidence for developmental priming. *J Proteome Res* 12(11):4892–4903. <https://doi.org/10.1021/pr400197p>
- Chaturvedi P, Ghatak A, Weckwerth W (2016) Pollen proteomics: from stress physiology to developmental priming. *Plant Reprod* 29(1–2):119–132. <https://doi.org/10.1007/s00497-016-0283-9>
- Cheong BE, Beine-Golovchuk O, Gorka M et al (2021) Arabidopsis REI-LIKE proteins activate ribosome biogenesis during cold acclimation. *Sci Rep* 11(1):1–25. <https://doi.org/10.1038/s41598-021-81610-z>
- cRAP database. n.d. atabase. <http://www.Thegpm.Org/Crap/>
- Datta R, Paul S (2019) Long non-coding RNAs: fine-tuning the developmental responses in Plants. *J Biosci* 44(4):1–11. <https://doi.org/10.1007/s12038-019-9910-6>
- Dixit A, Tomar P, Vaine E (2018) A stress-associated protein, AtSAP13, from *Arabidopsis thaliana* provides tolerance to multiple abiotic stresses. *Plant Cell Environ* 41(5):1171–1185. <https://doi.org/10.1111/pce.13103>
- Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics* 29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong X, Yi H, Han CT et al (2016) GDSL Esterase/Lipase genes in *Brassica rapa* L.: genome-wide identification and expression analysis. *Mol Genet Genom* 291(2):531–42. <https://doi.org/10.1007/s00438-015-1123-6>
- Duplakova N, Dobrev PI, Renák D, Honys D (2016) Rapid separation of Arabidopsis male gametophyte developmental stages using

- a percoll gradient. *Nat Protocols* 11(10):1817–32. <https://doi.org/10.1038/nprot.2016.107>
- Feuerstein A, Niedermeier M, Bauer K et al (2010) Expression of the AtSUC1 gene in the female gametophyte, and ecotype-specific expression differences in male reproductive organs. *Plant Biol* 12(SUPPL. 1):105–114. <https://doi.org/10.1111/j.1438-8677.2010.00389.x>
- Fíla J, Závěská Drábková L, Gíbalová A, Honys D (2017) When simple meets complex: pollen and the -omics. In: Feijó J, Obermeyer G (eds) *Pollen tip growth*. Springer, Cham, pp 247–292
- Fíla J, Božena K, David P et al (2020) The beta subunit of nascent polypeptide associated complex plays a role in flowers and siliques development of *Arabidopsis Thaliana*. *Int J Mol Sci* 21(6):1–30. <https://doi.org/10.3390/ijms21062065>
- Fleischer TC, Weaver CM, McAfee KJ, Jennings JL, Link AJ (2006) Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Dev* 20(10):1294–1307. <https://doi.org/10.1101/gad.1422006>
- Franchi GG, Piotto B, Nepi M et al (2011) Pollen and seed desiccation tolerance in relation to degree of developmental arrest, dispersal, and survival. *J Exp Bot* 62(15):5267–5281. <https://doi.org/10.1093/jxb/err154>
- Gallego ME, White CI (2001) RAD50 function is essential for telomere maintenance in arabidopsis. *Proc Natl Acad Sci USA* 98(4):1711–1716. <https://doi.org/10.1073/pnas.98.4.1711>
- Ge W, Song Y, Zhang C et al (2011) Proteomic analyses of Apoplastic proteins from germinating *Arabidopsis Thaliana* pollen. *Biochim Biophys Acta Proteins Proteom* 1814:1964–1973. <https://doi.org/10.1016/j.bbapap.2011.07.013>
- Golicz AA, Allu AD, Li W et al (2021) A dynamic intron retention program regulates the expression of several hundred genes during pollen meiosis. *Plant Reprod*. <https://doi.org/10.1007/s00497-021-00411-6>
- Grobei MA, Qeli E, Brunner E et al (2009) Deterministic protein inference for shotgun proteomics data provides new insights into Arabidopsis pollen development and function. *Genome Res* 19(10):1786–1800. <https://doi.org/10.1101/gr.089060.108>
- Hackenberg D, Twell D (2019) The evolution and patterning of male gametophyte development. *Current topics in developmental biology*. Elsevier Inc, Amsterdam
- Hafidh S, Honys D (2021) Reproduction multitasking: the male gametophyte. *Annu Rev Plant Biol* 72(1):581–614. <https://doi.org/10.1146/annurev-arplant-080620-021907>
- Hafidh S, Potěšil D, Fíla J et al (2016) Quantitative proteomics of the tobacco pollen tube Secretome identifies novel pollen tube guidance proteins important for fertilization. *Genome Biol* 17(1):1–29. <https://doi.org/10.1186/s13059-016-0928-x>
- Hafidh S, Potěšil D, Müller K et al (2018) Dynamics of the pollen Sequestrone defined by subcellular coupled Omics. *Plant Physiol* 178(1):258–282. <https://doi.org/10.1104/pp.18.00648>
- Hamilton ES, Jensen GS, Maksaev G et al (2015) Mechanosensitive channel MSL8 regulates osmotic forces during pollen hydration and germination. *Science* 350(6259):438–441. <https://doi.org/10.1126/science.aac6014>
- Henriques R, Wang H, Liu J et al (2017) The Antiphasic regulatory module comprising CDF5 and Its Antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytol*. <https://doi.org/10.1111/nph.14703>
- Hirsche J, García Fernández JM, Stabenheimer E, Großkinsky DK, Roitsch T (2017) Differential effects of carbohydrates on arabidopsis pollen germination. *Plant Cell Physiol* 58(4):691–701. <https://doi.org/10.1093/pcp/pcx020>
- Holmes-Davis R, Tanaka CK, Vensel WH, Hurkman WJ, McCormick S (2005) Proteome mapping of mature pollen of *Arabidopsis thaliana*. *Proteomics* 5(18):4864–4884. <https://doi.org/10.1002/pmic.200402011>
- Honys D, Twell D (2003) Comparative analysis of the arabidopsis pollen transcriptome. *Plant Physiol* 132(2): 640–52. <https://doi.org/10.1104/pp.103.020925>
- Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in arabidopsis. *Genome Biol* 5(11). <https://doi.org/10.1186/gb-2004-5-11-r854>
- Hu Y, Zhou L, Yang Y et al (2021) The gibberellin signaling negative regulator RGALIKE3 promotes seed storage protein accumulation. *Plant Physiol* 185(4):1697–1707. <https://doi.org/10.1093/plphys/kiaa114>
- Ischebeck T, Valledor L, Lyon D et al (2014) Comprehensive cell-specific protein analysis in early and late pollen development from diploid microsporocytes to pollen tube growth. *Mol Cell Proteom* 13(1):295–310. <https://doi.org/10.1074/mcp.M113.028100>
- Johnson MA, Harper JF, Palanivelu R (2019) A fruitful journey: pollen tube navigation from germination to fertilization. *Annu Rev Plant Biol* 70:809–837. <https://doi.org/10.1146/annurev-arplant-050718-100133>
- Julca I, Ferrari C, Flores-Tornero M et al (2021) Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nat Plants* 7(8):1143–1159. <https://doi.org/10.1038/s41477-021-00958-2>
- Kang YJ, Yang DC, Kong L et al (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* 45(W1):W12–16. <https://doi.org/10.1093/nar/gkx428>
- Keller M, Hu Y, Mesihovic A et al (2017) Alternative splicing in tomato pollen in response to heat stress. *DNA Res* 24(2):205–217. <https://doi.org/10.1093/dnares/dsw051>
- Keller M, Simm S, Bokszczanin KL et al (2018) The coupling of transcriptome and proteome adaptation during development and heat stress response of tomato pollen. *BMC Genom* 19(1):1–20. <https://doi.org/10.1186/s12864-018-4824-5>
- Kim HJ, Ok SH, Bahn SC et al (2011) Endoplasmic reticulum- and Golgi-localized phospholipase A2 plays critical roles in Arabidopsis pollen development and germination. *Plant Cell* 23(1):94–110. <https://doi.org/10.1105/tpc.110.074799>
- Lee JY, Lee DH (2003) Use of serial analysis of gene expression technology to reveal changes in gene expression in Arabidopsis pollen undergoing cold stress. *Plant Physiol* 132(2):517–529. <https://doi.org/10.1104/pp.103.020511>
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* 12:323. <https://doi.org/10.1186/1471-2105-12-323>
- Li X, Shahid MQ, Wen M et al (2020) Global identification and analysis revealed differentially expressed LncRNAs associated with meiosis and low fertility in autotetraploid rice. *BMC Plant Biol* 20(1):1–19. <https://doi.org/10.1186/s12870-020-2290-0>
- Lin SY, Chen PW, Chuang MH et al (2014) Profiling of translatomes of in vivo-grown pollen tubes reveals genes with roles in micropylar guidance during pollination in arabidopsis. *Plant Cell* 26(2):602–618. <https://doi.org/10.1105/tpc.113.121335>
- Liu J, Zhong S, Guo X et al (2013) Membrane-bound RLCKs LIP1 and LIP2 are essential male factors controlling male-female attraction in Arabidopsis. *Curr Biol* 23(11):993–998. <https://doi.org/10.1016/j.cub.2013.04.043>
- Liu F, Hu W, Li F (2020) AUTOPHAGY-RELATED14 and its associated phosphatidylinositol 3-kinase complex promote autophagy in Arabidopsis. *Plant Cell* 32(12):3939–3960. <https://doi.org/10.1105/tpc.20.00285>
- Lorraine AE, McCormick S, Estrada A, Patel K, Qin P (2013) RNA-Seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. *Plant Physiol* 162(2):1092–1109. <https://doi.org/10.1104/pp.112.211441>

- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 15(12):1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinf* 10:1–17. <https://doi.org/10.1186/1471-2105-10-161>
- Luo M, Tai R, Yu CW et al (2015) Regulation of flowering time by the histone deacetylase HDA5 in Arabidopsis. *Plant J* 82(6):925–936. <https://doi.org/10.1111/tpj.12868>
- Luo W, Pant G, Bhavnasi YK, Blanchard SG, Brouwer C (2017) Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res* 45(W1):W501–W508. <https://doi.org/10.1093/nar/gkx372>
- Ma ZX, Leng YJ, Chen GX et al (2015) The THERMOSENSITIVE MALE STERILE 1 Interacts with the BiPs via DnaJ domain and stimulates their ATPase enzyme activities in Arabidopsis. *PLoS ONE* 10(7):1–13. <https://doi.org/10.1371/journal.pone.0132500>
- Madison LS, Buchanan ML, Glass JD et al (2015) Class XI myosins move specific organelles in pollen tubes and are required for normal fertility and pollen tube growth in Arabidopsis. *Plant Physiol* 169(3):1946–1960. <https://doi.org/10.1104/pp.15.01161>
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18(9):1509–1517. <https://doi.org/10.1101/gr.079558.108>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17(1):10–12
- Mascarenhas JP (1990) Gene activity during pollen development. *Annu Rev Plant Physiol Plant Mol Biol* 41(1):317–338. <https://doi.org/10.1146/annurev.pp.41.060190.001533>
- Mistry J, Chuguransky S, Williams L et al (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res* 49(D1):D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- Mitsuda N, Ohme-Takagi M (2009) Functional analysis of transcription factors in Arabidopsis. *Plant Cell Physiol* 50(7):1232–1248. <https://doi.org/10.1093/pcp/pcp075>
- Náprstková A, Malinská K, Závěská Drábková L et al (2021) Characterization of ALBA family expression and localization in Arabidopsis thaliana generative organs. *Int J Mol Sci* 22(4):1–23. <https://doi.org/10.3390/ijms22041652>
- Nelms B, Walbot V (2022) Gametophyte genome activation occurs at pollen mitosis I in maize. *Science* 375(6579):424–429. <https://doi.org/10.1126/science.abl7392>
- Ner-Gaon H, Halachmi R, Savaldi-Goldstein S et al (2004) Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *Plant J* 39(6):877–885. <https://doi.org/10.1111/j.1365-313X.2004.02172.x>
- Osborne DJ, Boubriak I (2002) Telomeres and their relevance to the life and death of seeds. *Crit Rev Plant Sci* 21(2):127–141. <https://doi.org/10.1080/0735-260291044214>
- Pasha A, Shabari S, Cleary A et al (2020) Araport lives: an updated framework for Arabidopsis bioinformatics. *Plant Cell* 32(9):2683–2686. <https://doi.org/10.1105/TPC.20.00358>
- Passardi F, Dobias J, Valério L et al (2007) Morphological and physiological traits of three major Arabidopsis thaliana accessions. *J Plant Physiol* 164(8):980–992. <https://doi.org/10.1016/j.jplph.2006.06.008>
- Picelli S, Björklund ÅK, Faridani OR et al (2013) Smart-Seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10(11):1096–1100. <https://doi.org/10.1038/nmeth.2639>
- Picelli S, Faridani OR, Björklund ÅK et al (2014) Full-length RNA-seq from single cells using Smart-Seq2. *Nat Protoc* 9(1):171–181. <https://doi.org/10.1038/nprot.2014.006>
- Pina C, Pinto F, Feijó JA, Becker JD (2005) Gene family analysis of the arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol* 138(2):744–756. <https://doi.org/10.1104/pp.104.057935>
- Poidevin L, Forment J, Unal D, Ferrando A (2020) Transcriptome and translome changes in germinated pollen under heat stress uncover roles of transporter genes involved in pollen tube growth. *Plant Cell Environ* 44(7):2167–2184. <https://doi.org/10.1111/pce.13972>
- Qin T, Liu X, Li J et al (2014) Arabidopsis microtubule-destabilizing Protein 25 functions in pollen tube growth by severing actin filaments. *Plant Cell* 26(1):325–339. <https://doi.org/10.1105/tpc.113.119768>
- Qin P, Loraine AE, McCormick S (2018) Cell-specific cis-natural antisense transcripts (Cis-NATs) in the sperm and the pollen vegetative cells of Arabidopsis thaliana. *F1000Research* 7(May):1–10. <https://doi.org/10.12688/f1000research.13311.1>
- Raggi S, Demes E, Liu S, Verger S, Robert S (2020) Polar expedition: mechanisms for protein polar localization. *Curr Opin Plant Biol* 53:134–140. <https://doi.org/10.1016/j.pbi.2019.12.001>
- Reyes A, Anders S, Weatheritt RJ et al (2013) Drift and conservation of differential exon usage across tissues in primate species. *Proc Natl Acad Sci USA* 110(38):15377–15382. <https://doi.org/10.1073/pnas.1307202110>
- Robertson WR, Clark K, Young JC, Sussman MR (2004) An Arabidopsis thaliana plasma membrane proton pump is essential for pollen development. *Genetics* 168(3):1677–1687. <https://doi.org/10.1534/genetics.104.032326>
- Röckel N, Wolf S, Kost B, Rausch T, Greiner S (2008) Elaborate Spatial Patterning of cell-wall PME and PME1 at the pollen tube tip involves PME1 endocytosis, and reflects the distribution of esterified and de-esterified pectins. *Plant J* 53(1):133–143. <https://doi.org/10.1111/j.1365-313X.2007.03325.x>
- Rodriguez-Enriquez MJ, Mehdi S, Dickinson HG, Grant-Downton RT (2013) A novel method for efficient in vitro germination and tube growth of Arabidopsis thaliana pollen. *New Phytol* 197(2):668–679. <https://doi.org/10.1111/nph.12037>
- Rutley N, Twell D (2015) A decade of pollen transcriptomics. *Plant Reprod* 28(2):73–89. <https://doi.org/10.1007/s00497-015-0261-7>
- Sauer N, Ludwig A, Knoblauch A et al (2004) AtSUC8 and AtSUC9 encode functional sucrose transporters, but the closely related AtSUC6 and AtSUC7 genes encode aberrant proteins in different Arabidopsis ecotypes. *Plant J* 40(1):120–130. <https://doi.org/10.1111/j.1365-313X.2004.02196.x>
- Schmitz J, Rossoni AW, Maurino VG (2018) Dissecting the physiological function of plant glyoxalase I and glyoxalase I-like proteins. *Front Plant Sci* 871(November):1–7. <https://doi.org/10.3389/fpls.2018.01618>
- Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-Seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4(2):1521. <https://doi.org/10.12688/f1000research.7563.1>
- Sun CW, Callis J (1997) Independent modulation of Arabidopsis thaliana polyubiquitin mRNAs in different organs and in response to environmental changes. *Plant J* 11(5):1017–1027. <https://doi.org/10.1046/j.1365-313X.1997.11051017.x>
- Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0021800>
- Tabas-Madrid D, Méndez-Vigo B, Arteaga N et al (2018) Genome-wide signatures of flowering adaptation to climate temperature: regional analyses in a highly diverse native range of Arabidopsis thaliana. *Plant Cell Environ* 41(8):1806–1820. <https://doi.org/10.1111/pce.13189>
- Thimm O, Bläsing O, Gibon Y et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic

- pathways and other biological processes. *Plant J* 37(6):914–939. <https://doi.org/10.1111/j.1365-3113X.2004.02016.x>
- Thomas PD, Campbell MJ, Kejariwal A et al (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13(9):2129–2141. <https://doi.org/10.1101/gr.772403>
- Twell D, Oh SA, Honys D (2006) Pollen development, a genetic and transcriptomic view. *Plant Cell Monogr* 3:15–45. [https://doi.org/10.1007/7089\\_042](https://doi.org/10.1007/7089_042)
- Ueda M, Matsui A, Tanaka M et al (2017) The distinct roles of class I and II RPD3-like histone deacetylases in salinity stress response. *Plant Physiol* 175(4):1760–1773. <https://doi.org/10.1104/pp.17.01332>
- UniprotKB [ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/reference\\_proteomes/Eukaryota/UP000006548/UP000006548\\_3702.fasta.gz](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/Eukaryota/UP000006548/UP000006548_3702.fasta.gz)
- Updegraff EP, Zhao F, Preuss D (2009) The extracellular lipase EXL4 is required for efficient hydration of Arabidopsis pollen. *Sex Plant Reprod* 22(3):197–204. <https://doi.org/10.1007/s00497-009-0104-5>
- Vanechoutte D, Estrada AR, Lin YC, Loraine AE, Vandepoel K (2017) Genome-wide characterization of differential transcript usage in *Arabidopsis thaliana*. *Plant J* 92(6):1218–1231. <https://doi.org/10.1111/tpj.13746>
- Vanstraelen M, Inzé D, Geelen D (2006) Mitosis-specific kinesins in Arabidopsis. *Trends Plant Sci* 11(4):167–175. <https://doi.org/10.1016/j.tplants.2006.02.004>
- Vitting-Seerup K, Sandelin A (2017) The landscape of isoform switches in human cancers. *Mol Cancer Res* 15(9):1206–1220. <https://doi.org/10.1158/1541-7786.MCR-16-0459>
- Vitting-Seerup K, Sandelin A, Berger B (2019) IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* 35(21):4469–4471. <https://doi.org/10.1093/bioinformatics/btz247>
- Wingett SW, Andrews S (2018) Fastq screen: a tool for multi-genome mapping and quality control. *F1000Research* 7:1–13. <https://doi.org/10.12688/f1000research.15931.1>
- Xia R, Wang J, Liu C et al (2006) ROR1/RPA2A, a putative replication Protein A2, functions in epigenetic gene silencing and in regulation of meristem development in Arabidopsis. *Plant Cell* 18(1):85–103. <https://doi.org/10.1105/tpc.105.037507>
- Xu N, Gao XQ, Zhao XY (2011) Arabidopsis AtVPS15 Is essential for pollen development and germination through modulating phosphatidylinositol 3-phosphate formation. *Plant Mol Biol* 77(3):251–260. <https://doi.org/10.1007/s11103-011-9806-9>
- Yang KZ, Xia C, Liu XL et al (2009) A mutation in THERMOSENSITIVE MALE STERILE 1, encoding a heat shock protein with DnaJ and PDI domains, leads to thermosensitive gametophytic male sterility in Arabidopsis. *Plant J* 57(5):870–882. <https://doi.org/10.1111/j.1365-3113X.2008.03732.x>
- Yang H, You C, Yang S et al (2021) The role of calcium/calcium-dependent protein kinases signal pathway in pollen tube growth. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2021.633293>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## **GOLEM: distribution of Gene regulatOry eLEMents within the plant promoters**

**My contribution:** In the development of the motif analysis tool GOLEM, I acquired and analysed all the RNA-seq data utilized for the program. I was actively involved in discussions about the program functionality. My contributions included conducting background literature research to select relevant organisms and motifs of interest and determine the programs criteria. Additionally, I participated in the critical review of the manuscript text and figures.

Current status: ready for submission.

Beta version running on: <https://golem.ncbr.muni.cz/>



## GOLEM: distribution of gene regulatory elements within the plant promoters

Lukáš Nevošád,<sup>1</sup> Božena Klodová,<sup>2,3</sup> David Honys,<sup>2,3</sup> Radka Svobodová,<sup>4,5</sup> Tomáš Raček<sup>4,5</sup> and Petra Procházková Schrupfová<sup>4,5,\*</sup>

<sup>1</sup>Tripomatic s.r.o., Za Parkem 14, 60200, Brno, Czech Republic, <sup>2</sup>Institute of Experimental Botany of the Czech Academy of Sciences, Rozvojová 263, 165 02, Prague, Czech Republic, <sup>3</sup>Department of Experimental Plant Biology, Faculty of Science, Charles University, Viničná 5, 128 00, Praha, Czech Republic, <sup>4</sup>National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00, Brno, Czech Republic and <sup>5</sup>CEITEC – Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00, Brno, Czech Republic

\*Corresponding author. [schpetra@sci.muni.cz](mailto:schpetra@sci.muni.cz)

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

### Abstract

**Motivation:** The regulation of gene expression during tissue development is extremely complex. One of the key regulatory mechanisms of gene expression involves the recognition of regulatory motifs by various proteins in the promoter regions of many genes. Localisation of these motifs in proximity to the transcription start site (TSS) or translation start site (ATG) is critical for regulating the initiation and rate of transcription. The levels of transcription of individual genes, regulated by these motifs, can vary significantly in different tissues and developmental stages, especially during tightly regulated processes such as sexual reproduction. However, the precise localisation and visualisation of the regulatory motifs within gene promoters with respect to gene transcription in specific tissues, can be challenging.

**Results:** Here, we introduce a program called GOLEM (Gene regulatOry eLEMents) which enables users to precisely locate any motif of interest with respect to TSS or ATG within the relevant plant genomes across the plant Tree of Life (*Marchantia*, *Physcomitrium*, *Amborella*, *Oryza*, *Zea*, *Solanum* and *Arabidopsis*). The visualisation of the motifs is performed with respect to the transcript levels of particular genes in leaves and male tissues, and can be compared with genome-wide distribution regardless of the transcription level.

**Availability and implementation:** GOLEM is freely available at <https://golem.ncbr.muni.cz>.

### 1. Introduction

The regulation of gene expression is a dynamic process that requires a tightly orchestrated control mechanism. The precise timing and magnitude of gene transcription, especially during plant development, can be regulated by several DNA sequence motifs. These regulatory motifs can be found in the vicinity of the transcription start sites (TSS) - upstream or even downstream, including the five-prime untranslated region (5' UTR) - and serve as binding sites for various proteins. These proteins act as molecular switches that can activate or repress gene expression (Galli et al., 2020; Schmitz et al., 2022).

Dynamic changes in gene transcription are essential, especially during the development of plant reproductive tissues. Several short DNA motifs involved in the regulation of key genes required for the differentiation of male germ line cells (sperm cells/pollen) have already been identified: LAT52 pollen-specific motif in tomato

(AGAAA) (Bate and Twell, 1998); LAT52/59 core motif (GTGA) (Twell et al., 1991); DOF core motif (AAAG) (Yanagisawa, 2002; Li et al., 2014); ARR1AT motif (NGATT) (Sharma et al., 2011); pollen Q-element (AGGTCA) (Hamilton et al., 1998), and many others (Sharma et al., 2011; Peters et al., 2017; Hoffmann et al., 2017; Li et al., 2014). Furthermore, even common motifs like the TATA box (TATAWA), which are involved in initiating transcription, exhibit multiple variants that may contribute to the establishment of unique expression patterns (Bernard et al., 2010).

Transcriptomics has gained significant popularity in recent years, as it has the ability to provide detailed insights into gene expression dynamics across different tissues, developmental stages, or experimental conditions (Tyagi et al., 2022). Transcriptome sequencing (RNA-Seq) is an important method for investigating gene regulation. While most transcriptome analyses have traditionally focused on easily accessible plant materials such as

leaves or seedlings, there is a growing trend towards exploring transcriptomes from intricate and inaccessible tissues, including sperm cells and various pollen developmental stages (Julca et al., 2021; Klodová et al., 2022).

Achieving precise localisation and visualisation of regulatory motifs in genes that exhibit high transcription levels in specific tissues requires a multi-step procedure, which may be limited by the user's familiarity with various bioinformatics software. We present a user-friendly online software GOLEM (Gene regulatOry eLEMents), which allows browsing various tissues such as sporophyte (leaves) or male gametophyte developmental tissues (antheridia, pollen stages, sperm cells) across the relevant plant genomes within the plant Tree of Life. Our software enables us to investigate the precise localisation and distribution of motifs of interest in gene promoters, which can be specified by the level of gene expression in specific tissues. Furthermore, tracking of the genome-wide distribution across mosses, monocots and dicots, regardless of the transcription level, may aid to track the evolution of regulatory motifs. Additionally, the set of genes with specific motifs at defined positions can be exported for further analysis.

## 2. Materials and methods/Algorithm

The GOLEM program is divided into two main phases: data processing pipeline and data visualisation.

### 2.1. Data processing

The reference genomes and annotations files from *Marchantia polymorpha*, *Physcomitrium patens*, *Amborella trichopoda*, *Oryza sativa*, *Zea mays*, *Solanum lycopersicum*, and *Arabidopsis thaliana* were downloaded in the FASTA format and general feature (GFF3) format, respectively (Supplementary Table 1). The data processing pipeline first parses genomic location data of single genes provided in GFF3 files. The GFF3 file is used to identify the position of transcription initiation (five\_prime\_UTR) and translation initiation (start\_codon), i.e. determines the positions of TSS (transcription start site) and ATG (first translated codon) in the reference genome. A fixed number of base pairs before and after the TSS or ATG are included in the dataset.

Next, the pipeline matches selected genes against a Transcript Per Million (TPM) table from various tissues and developmental stages. The TPM values express normalized transcription rates of each gene obtained from RNA-seq data sets. The TPM values for *M. polymorpha*, *P. patens*, *A. trichopoda*, *O. sativa*, and *S. lycopersicum* tissue were acquired from Conekt database (<https://conekt.sbs.ntu.edu.sg>) (Julca et al., 2021). The TPM values for pollen developmental stages of *A. thaliana* Columbia (Col-0) and Landsberg erecta (Ler) were extracted from (Klodová et al., 2022). The TPM values for *Z. mays* and some of the *A. thaliana* developmental stages (e.g., leaves, seedlings, tapetum) were established from the FASTQ files using the pipeline described in Supplementary Data 1 of (Klodová et al., 2022). The normalised TPM values used for the data processing pipeline are listed in Supplementary Table 2.

The output of the data processing pipeline is a separate FASTA-compatible file that contains each valid gene from the original input, along with information about the position of TSS and ATG, and transcription rates in each stage added as comments. The pipeline also generates a validation log that provides information about genes that were excluded, i.e.

non-protein coding genes (noStartCodonFound), pseudogenes (noFivePrimeUtrFound, noTpmDataFound), genes without TSS (noFivePrimeUtrFound, if relevant for certain organism).

Motif search uses regular expressions to search the input string of base pairs. For each motif, the reverse complement was calculated and then translated both into regular expressions. When the regular expression is run against the input data, we record all results and calculate their relative positions (adjusted to the middle of the motif) relative to TSS and ATG.

### 2.2. Data visualisation

The data visualisation phase consists of five steps, as depicted in Fig. 1. In the selected genome (Fig. 1a), the user can then choose the genomic interval to be searched, effectively specifying the window of the sequence where the search is performed, and focus on a defined region in the vicinity of TSS or ATG (Fig. 1b). The user can also select motifs for the search, including custom motifs and preset motifs resulting from previous research. Motifs are searched in both forward and reverse forms, and the reverse form is calculated automatically (Fig. 1c).

Before the entire analysis, the user confirms the tissue and developmental stages to be searched and selects the method for selecting genes for analysis (Fig. 1d). Gene selection uses a given percentile (default is 90th) to select the genes that are the most/least transcribed in tissues or developmental stages of interest to exclude the genes with low or even negligible transcription. The number of selected genes within the given percentile can be tracked during the proceeding steps. Besides the given percentile, the number of genes included in the analysis may be determined by the user. Regardless of the transcription level, the motif distribution can also be included in the analysis (stage Genome).

The goal of the analysis is to visualise the distribution of motifs of interest in the vicinity of the TSS or ATG of all protein-coding genes, or exclusively in selected genes that exhibit high/low transcription levels in particular tissues and developmental stages (Fig. 1e). Results are presented graphically, with each stage colour-coded, and can be displayed as simple counts or percentages. The user can also choose to display either the number of motifs found or aggregate the motifs by genes.

Additionally, the set of genes with specific motifs at defined positions can be exported for further analysis (Fig. 1f). Within the analysis, the application allows the user to export each data series or export the aggregated data for all data series in XLSX format. The user can also see the distribution of individual motifs and drill down through them.

The data visualisation phase involves an application written in Flutter/Dart (Meiller, 2021), which can be run as a standalone application or compiled into JavaScript and hosted on the web as a single-page web application (<https://golem.ncbr.muni.cz>).

### 2.3. Limitations

The entire processing happens on the client within the application or web browser, with input files loaded into memory. The program's ability to work with large datasets may be constrained by the available memory and the web browser's local client settings. We found that even in the web application, where performance is limited due to the inefficiencies of JavaScript compared to platform native code, performance is satisfactory



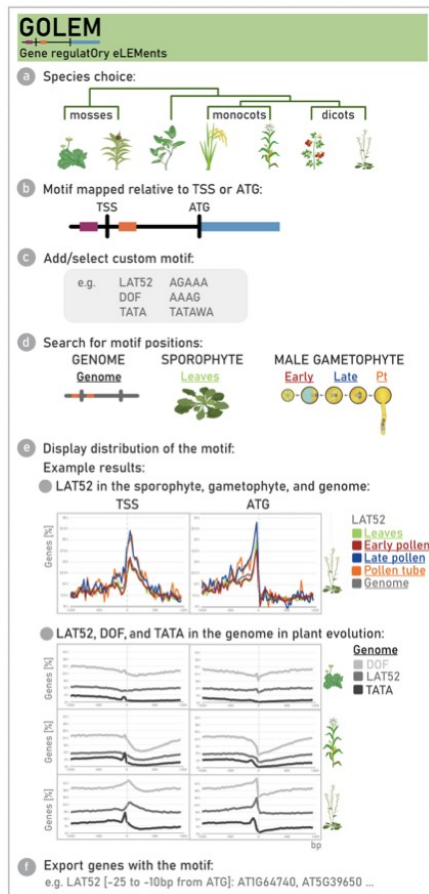


Fig. 1: An illustrative workflow of the GOLEM software. (a) Genome data for chosen species is downloaded on the web browser (*Marchantia*, *Physcomitrium*, *Amborella*, *Oryza*, *Zea*, *Solanum* and *Arabidopsis*). (b) Region in the vicinity of TSS or ATG is specified. (c) Motifs of interest are defined. (d) Leaves or male tissues and the transcriptional levels of the genes included in the analysis are defined. Additionally, the genome-wide distribution, regardless of the transcription level, can be analysed. (e) The distribution of the exemplified LATS2 motif shows that it is located downstream of the TSS but upstream of ATG, with a higher occurrence in the promoters of genes transcribed during Late pollen development. The genome-wide distribution of DOF, LATS2, and TATA-box motifs enables their tracking across the evolutionary tree. (f) The accession numbers of genes with certain motif within a defined region and developmental stages may be exported in XLSX format tables.

on modern computers without the need for significant code optimisation.

### 3. Results

Gene expression is primarily controlled through the specific binding of various proteins to diverse DNA sequence motifs. Our software GOLEM is a convenient tool for visualising the distribution of motifs of interest in the vicinity of TSS or ATG and identifying genes expressed at a certain level within selected tissues.

The distribution of one core promoter motif TATA (TATA-box) (Bernard et al., 2010), together with two pollen-associated motifs: LATS2 (LAT52 pollen-specific motif from tomato, POLLEN1LeLAT52) (Bate and Twell, 1998) and DOF core motif (DOFCORE) (Yanagisawa, 2002; Li et al., 2014) – are exemplified in Fig. 1e. Our analysis revealed that the preferential position of the LATS2 (the top of the peak) is located downstream of the TSS and upstream to ATG, i.e. in the five-prime untranslated region (5' UTR). Moreover, the number of the genes containing the LATS2 in *A. thaliana* is higher within the 5' UTR region of genes exhibiting elevated expression (90th percentile) in Late pollen development and in Pollen tubes, as opposed to genes expressed during Early pollen development or in Leaves.

Additionally, our analysis has shown that the TATA-box resides upstream of the TSS in all three exemplified genomes across the evolutionary tree (moss *P. patens*, monocot *Z. mays* and dicot *A. thaliana*). On the other hand, the pollen-associated motifs DOF and LATS2 show in *A. thaliana* a sharp peak centred downstream from the TSS.

From the genes with LATS2 motif expressed in the Late pollen stage of *A. thaliana*, the XLSX format table was exported. From this table two pollen-associated genes (ATG164740, AT5G39650), with LATS2 motif between -25 to -10 base pair (bp) area, are exemplified in Fig. 1f. Generally, the accession numbers of the genes exported from GOLEM can be analysed by various bioinformatical approaches: e.g., Gene description search; Gene ontology (GO) enrichment analysis; Differential expression analysis; Protein-protein interaction network analysis.

### 4. Conclusion

To address the challenge of precise localisation and visualisation of regulatory motifs near transcription and translation start sites, which play a critical role in gene regulation in specific tissues, we introduce the GOLEM software. Our software offers a user-friendly platform to investigate the distribution of any motif of interest in gene promoters across diverse plant genomes and developmental tissues.

### 5. Acknowledgement

Biological Data Management and Analysis Core Facility of CEITEC Masaryk University, funded by ELIXIR CZ research infrastructure (MEYS Grant No: LM2023055), is gratefully acknowledged for supporting the research presented in this paper.

### 6. Author contribution

P.P.S. and D.H. conceived the study. B.K. analysed RNA-seq data and calculated the TPM. L.N. imposed computation analysis and

visualisation tool. T.R. and R.S. implemented the website and supported program accessibility. P.P.S. wrote the paper with the help of all co-authors.

## 7. Declaration of Competing Interest

The authors report no declarations of interest.

## 8. Funding

This work was supported for P.P.S. and D.H. by the Czech Science Foundation [21-15841S]; and collaboration with Tripomatic s.r.o. was supported by the European Regional Development Fund Project 'SINGING PLANT' [CZ.02.01/0.0/0.0/16.026/0008446].

## References

- N. Bate and D. Twell. Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Molecular Biology*, 37:859–869, 1998.
- V. Bernard, V. Brunaud, and A. Lechary. TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics*, 11(1):1–15, 2010.
- M. Galli, F. Feng, and A. Gallavotti. Mapping regulatory determinants in plants. *Frontiers in Genetics*, 11:591194, 2020.
- D. A. Hamilton, Y. H. Schwarz, and J. P. Mascarenhas. A monocot pollen-specific promoter contains separable pollen-specific and quantitative elements. *Plant Molecular Biology*, 38:663–669, 1998.
- R. D. Hoffmann, L. I. Olsen, J. O. Husum, J. S. Nicolet, J. F. Thøfner, A. P. Wätjen, C. V. Ezike, and M. Palmgren. A cis-Regulatory Sequence Acts as a Repressor in the Arabidopsis thaliana Sporophyte but as an Activator in Pollen. *Molecular Plant*, 10(5):775–778, 2017.
- I. Julca, C. Ferrari, M. Flores-Tornero, S. Proost, A.-C. Lindner, D. Hackenberg, L. Steinbachová, C. Michaelidis, S. Gomes Pereira, C. S. Misra, et al. Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nature Plants*, 7(8):1143–1159, 2021.
- B. Klodová, D. Potěšil, L. Steinbachová, C. Michaelidis, A.-C. Lindner, D. Hackenberg, J. D. Becker, Z. Zdráhal, D. Twell, and D. Honys. Regulatory dynamics of gene expression in the developing male gametophyte of Arabidopsis. *Plant Reproduction*, pages 1–29, 2022.
- J. Li, J. Yuan, M. Li, et al. Characterization of Putative cis-Regulatory Elements in Genes Preferentially Expressed in Arabidopsis Male Meicytes. *BioMed Research International*, 2014, 2014.
- D. Meiller. *Modern App Development with Dart and Flutter 2: A Comprehensive Introduction to Flutter*. Walter de Gruyter GmbH & Co KG, 2021.
- B. Peters, J. Casey, J. Aidley, S. Zohrab, M. Borg, D. Twell, and L. Brownfield. A Conserved cis-Regulatory Module Determines Germline Fate through Activation of the Transcription Factor DUO1 Promoter. *Plant Physiology*, 173(1):280–293, 2017.
- R. J. Schmitz, E. Grotewold, and M. Stam. Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *The Plant Cell*, 34(2):718–741, 2022.
- N. Sharma, S. D. Russell, P. L. Bhalla, and M. B. Singh. Putative cis-regulatory elements in genes highly expressed in rice sperm cells. *BMC Research Notes*, 4(1):1–10, 2011.
- D. Twell, J. Yamaguchi, R. A. Wing, J. Ushiba, and S. McCormick. Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. *Genes & Development*, 5(3):496–507, 1991.
- P. Tyagi, D. Singh, S. Mathur, A. Singh, and R. Ranjan. Upcoming progress of transcriptomics studies on plants: An overview. *Frontiers in Plant Science*, 13:1030890, 2022.
- S. Yanagisawa. The Dof family of plant transcription factors. *Trends in Plant Science*, 7(12):555–560, 2002.

## **A decade of pollen phosphoproteomics**

**My contribution:** I conducted analyses of published pollen phosphoproteome studies to facilitate comparisons across different angiosperms and between roots and pollen. This was done to identify common regulatory pathways and shared gene groups. I was responsible for creating the figures, discussing the results, and co-authoring the manuscript.

Current status: published in International Journal of Molecular Sciences. IF<sub>2022</sub>: 5.6, cited 2x (NCBI)



Review

# A Decade of Pollen Phosphoproteomics

Božena Klodová<sup>1,2</sup> and Jan Fila<sup>1,\*</sup>

<sup>1</sup> Laboratory of Pollen Biology, Institute of Experimental Botany of the Czech Academy of Sciences, Rozvojová 263, 165 02 Prague, Czech Republic; klodova@ueb.cas.cz

<sup>2</sup> Department of Experimental Plant Biology, Faculty of Science, Charles University, Viničná 5, 128 00 Prague, Czech Republic

\* Correspondence: fila@ueb.cas.cz; Tel.: +420-225-106-452

**Abstract:** Angiosperm mature pollen represents a quiescent stage with a desiccated cytoplasm surrounded by a tough cell wall, which is resistant to the suboptimal environmental conditions and carries the genetic information in an intact stage to the female gametophyte. Post pollination, pollen grains are rehydrated, activated, and a rapid pollen tube growth starts, which is accompanied by a notable metabolic activity, synthesis of novel proteins, and a mutual communication with female reproductive tissues. Several angiosperm species (*Arabidopsis thaliana*, tobacco, maize, and kiwifruit) were subjected to phosphoproteomic studies of their male gametophyte developmental stages, mostly mature pollen grains. The aim of this review is to compare the available phosphoproteomic studies and to highlight the common phosphoproteins and regulatory trends in the studied species. Moreover, the pollen phosphoproteome was compared with root hair phosphoproteome to pinpoint the common proteins taking part in their tip growth, which share the same cellular mechanisms.

**Keywords:** phosphoproteomics; pollen tube; male gametophyte; root hair; signal transduction; kinase motif



**Citation:** Klodová, B.; Fila, J. A Decade of Pollen Phosphoproteomics. *Int. J. Mol. Sci.* **2021**, *22*, 12212. <https://doi.org/10.3390/ijms222212212>

Academic Editors: Sixue Chen and Setsuko Komatsu

Received: 30 September 2021  
Accepted: 8 November 2021  
Published: 11 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Species' existence on Earth is maintained by reproduction. The angiosperm (Angiospermae) life cycle consists of two altering generations—a diploid sporophyte and a haploid gametophyte [1]. The adult plants form the sporophyte, in the flowers of which, the spores of two distinct sexes (female and male) and sizes are produced by meiosis. These heterospores undergo mitotic divisions, by which multicellular gametophytes are formed. The female gametophyte develops within the ovary, where it is protected from any damage and in most species, it is composed of seven cells with eight nuclei [2]. On the other hand, a mature male gametophyte is formed by 2 or 3 cells [3]. The microspores undergo the asymmetrical pollen mitosis I, which gives rise to two distinct cells. The smaller generative cell (composed mainly of a nucleus) is engulfed by the bigger vegetative cell. Mature pollen grains are shed from anthers either in such a bi-cellular stage or alternatively undergo pollen mitosis II that forms two sperm cells out of one generative cell prior to pollen grain shedding, meaning they will be in a mature state tri-cellular [4,5]. Mature pollen aims at delivering the genetic information in an intact state to the pistil and to fulfil this task, it represents a resistant, metabolically quiescent stage with a dehydrated cytoplasm surrounded by a tough cell wall. Upon pollination, the cytoplasm of pollen grains re-hydrates [6] and it becomes metabolically active and later, the rapid pollen tube growth starts. The pollen tube growth through transmitting tissues of a pistil is accompanied by intensive communication between these structures [7]. Pollen mitosis II, that forms two sperm cells out of one generative cell, takes place in the bicellular pollen after pollination, for instance, in the case of tobacco after 10–12 h of pollen tube growth [8]. Finally, the pollen tube delivers two sperm cells, the male gametes, to the mature embryo sac. Both carried sperm cells take part in fertilization. One sperm cell fertilizes the egg cell (representing female gamete) to form the zygote and later the embryo, whereas the second sperm cell fuses with

the central nucleus of the embryo sac to form endosperm. Such a phenomenon is called double fertilization and is typical of angiosperms [9].

The change from metabolically quiescent, resistant mature pollen to a metabolically active, rapidly growing pollen tube is precisely regulated both at the level of protein synthesis and posttranslational modifications. The former regulation is mediated by the synthesis of mRNAs for storage in translationally inactive EDTA/puromycine-resistant particles (EPPs [10,11]), later described as monosomes [12], since, for instance, tobacco (*Nicotiana tabacum*) pollen tube growth was reported to be highly dependent on translation, but nearly independent of transcription [13]. The stored transcripts are de-repressed once the rapid pollen tube growth starts. Then, the post-translational modifications during pollen tube growth are most importantly represented by phosphorylation, which represents one of the most dynamic posttranslational modifications that mediates the regulation of numerous cellular processes. A similar re-hydration-related phosphorylation was described in xerophyte *Craterostigma plantagineum* [14,15]. The other post-translational modifications (namely glycosylation, methylation, myristoylation or acetylation) were also reported to play an important role in male gametophyte development [16]. Glycoproteins are on the one hand an important structural part of pollen tube cell walls and on the other hand play their roles in pollen tube perception [17].

This review aims at an analysis of the known phosphoproteomic datasets acquired on male gametophyte stages and compares them with the root hair phosphoproteome, since these structures share the same type of tip growth that relies on common cellular mechanisms [18–20].

## 2. Male Gametophyte Phosphoproteomic Studies

Various enrichment protocols were applied [21,22] to study protein phosphorylation on a large scale by phosphoproteomic techniques. The enrichment techniques are inevitable since (1) only several percent of cellular proteome are phosphorylated in a cell at a given time; (2) both phosphorylated and native isoforms of the same protein co-exist in the cell [23], sometimes even with a much higher concentration of their non-phosphorylated forms; (3) phosphorylated peptides are hardly detected in a positive ion scan mode during mass spectrometry if they are mixed with their non-phosphorylated counterparts [24].

The first phosphoproteomic dataset acquired from any angiosperm male gametophyte stage was represented by *Arabidopsis* (*Arabidopsis thaliana*) mature pollen, which was published nearly 10 years ago by Mayank and colleagues [25]. The first phosphoproteomic study relied on three phosphopeptide-enriching methods (immobilized metal affinity chromatography—IMAC, metal oxide affinity chromatography—MOAC, and sequential elution from IMAC—SIMAC) and collectively identified 962 phosphopeptides carrying 609 phosphorylation sites, which belonged to 598 phosphoproteins (Table 1). The total number of identified phosphopeptides could be higher than the number of identified phosphorylation sites. This is caused by the fact that the same phosphorylation site is carried by more than one phosphopeptide. Alternatively, some authors also calculate phosphopeptides, which lack the conclusively positioned phosphorylation site due to the insufficient support from the MS/MS spectra. In *Arabidopsis* pollen phosphoproteome, there prevailed proteins annotated by TopGO [26] which were involved in the regulation of metabolism and protein function, metabolism, protein fate, protein with a binding function, signal transduction mechanisms, and cellular transport. It is worth mentioning that various protein kinases (including AGC protein kinases, calcium-dependent protein kinases, and sucrose non-fermenting protein kinases 1) were amongst the identified phosphopeptides. Two over-represented phosphorylation motifs in the *Arabidopsis* pollen phosphoproteome were identified—a prolyl-directed motif xxxxxxS\*Pxxxxx, and a basic motif xxxRxxS\*xxxxxx (the phosphorylation site here and onwards is represented by an asterisk).

**Table 1.** Summary of the publications that presented angiosperm male gametophyte phosphoproteomes.

Species	Citation	Enrichment Technique	Studied Stages		Number of Identified Phosphoproteins	Number of Identified Phosphopeptides	Number of Identified Phosphorylation Sites	pSer:pThr:pTyr Ratio	Phosphorylation Motifs
			Mature Pollen	Activated Pollen					
<i>Arabidopsis thaliana</i>	Mayank et al. 2012 [25]	IMAC, TiO <sub>2</sub> -MOAC, SIMAC	×		598	962	609	86:14:0.16	1 prolyl-directed (xxxxxxS*Pxxxxx) 1 basic (xxxRxxS*xxxxxx)
<i>Nicotiana tabacum</i>	Fila et al. 2012 [27]	Al(OH) <sub>3</sub> -MOAC, TiO <sub>2</sub> -MOAC of the already identified peptides	×	×	139	52	52	67.3:32.7:0	not identified, too small data set
<i>Nicotiana tabacum</i>	Fila et al. 2016 [28]	TiO <sub>2</sub> -MOAC	×	×	301	471	432	86.4:13.4:0.2	2 prolyl-directed (xxxxxxS*Pxxxxx; xxxxxxT*Pxxxxx) 2 basic (xxxRxxS*xxxxxx; xxxKxxS*xxxxxx) 2 acidic (xxxxxxS*Dxxxxx; xxxxxxS*xDDxxx) 8 prolyl-directed 5 basic 4 acidic 10 other
<i>Zea mays</i>	Chao et al. 2016 [29]	IMAC	×		2257	4638	5292	81.5:14.5:4	6 prolyl-directed 5 basic 8 acidic 20 other
<i>Actinidia deliciosa</i>	Vannini et al. 2019 [30]	MOAC phosphoprotein enrichment + IMAC-Ti phosphopeptide enrichment		×	711	1299	1572	90.3:9:0.7	

The second angiosperm species that was subjected to male gametophyte phosphoproteomic studies was tobacco (*Nicotiana tabacum*) [27,28]. Tobacco became the first species in which the activated pollen grains were taken into consideration, since it identified phosphoproteins from mature pollen, 30-min activated pollen [27,28], and in the more recent study also from 5-min activated pollen [28]. The former study relied on phosphoprotein enrichment by aluminium hydroxide matrix, the eluate of which was separated both by a conventional two-dimensional gel electrophoresis (2D-GE), and by nano liquid chromatography (nLC) [27]. Although 139 phosphoprotein candidates were identified, the number of exactly matched phosphorylation sites was lower, since it identified only 52 phosphorylation sites (Table 1). The number of phosphorylation sites identified in the tobacco male gametophyte was notably broadened in the second study that applied phosphopeptide enrichment by titanium dioxide to identify phosphopeptides from mature pollen, 5-min activated pollen, and 30-min activated pollen [28]. The study described 301 phosphoproteins, which contained 471 phosphopeptides that carried 432 exactly matched phosphorylation sites (Table 1). Furthermore, several regulated phosphopeptides that changed their abundance between the studied stages were identified. There were seven such categories, including phosphopeptides present exclusively in either studied stage. Like in *Arabidopsis*, the most abundant functional categories were represented by protein synthesis, together with protein destination and storage, transcription, and signal transduction. The motif search in the second phosphoproteomic study revealed five motifs with a central phosphoserine and one motif with a central phosphothreonine. There were prolyl-directed phosphorylations on both serine and threonine (xxxxxxS\*Pxxxxx, and xxxxxxT\*Pxxxxx), two basic motifs (xxxRxxS\*xxxxxx, and xxxKxxS\*xxxxxx), and two acidic motifs (xxxxxxS\*Dxxxxx, and xxxxxxS\*xDDxxx).

In 2016, the first monocot, represented by maize (*Zea mays*) mature pollen [29], was subjected to phosphoproteomic studies, but no activated stage of male gametophyte was studied. This study relied solely on gel-free techniques combined with IMAC phosphopeptide enrichment. It led to the identification of 4638 phosphopeptides in 2257 proteins that carried 5292 phosphorylation sites (Table 1). The number of phosphorylation sites identified is roughly 10 times higher, whereas the number of phosphopeptides is approximately 5–10 times higher than in *Arabidopsis* or tobacco pollen phosphoproteomes. The increase could be caused (1) in case of *A. thaliana* by technical improvements after a few years (*Arabidopsis* phosphoproteome was published 4 years before), and (2) compared to tobacco, maize represents a sequenced plant with an annotated genome [31,32]. It is likely that several tobacco MS spectra were not coupled with any sequence from the available databases since the tobacco genome was not fully annotated when the analyses were performed [33], and although the annotations were improved then, they are still far from completion [34]. Chao et al. (2016) were notably more successful in identifying the phosphorylation motifs over-represented in the presented phosphoproteome—the

dataset comprised of 23 phosphoserine motifs and 4 phosphothreonine motifs, representing a total of 27 motifs. There were 8 prolyl-directed motifs, 5 basic motifs, and 4 acidic motifs, which usually represented more specified versions of the above tobacco and Arabidopsis motifs. However, there appeared also a variety of 10 newly discovered motifs. The phosphoprotein categories in maize phosphoproteome were represented by DNA synthesis/chromatin structure, transcription regulation, protein modification, cell organization, signal transduction, cell cycle, vesicular transport, transport of ions and various metabolic pathways. It is worth mentioning that Chao et al. found 430 protein kinases and 105 phosphatases. Some kinases represented the families, the phosphorylation motifs of which were up-regulated in the present phosphoproteome—for instance, calcium-dependent protein kinases (CDPK), leucine rich repeat kinases (LRRK), SNF1-related protein kinases (SnRK), and mitogen-activated protein kinases (MAPK). Finally, Chao et al. (2016) clearly demonstrated that the enrichment techniques are inevitable for studying protein phosphorylation by high-throughput methods. There were 5146 total proteins without phosphorylation in the maize pollen proteome, 1604 proteins in both datasets (total proteome and phosphoproteome), and then an additional 653 phosphoproteins were identified exclusively upon phosphopeptide enrichment. It is obvious that quite a big part of the phosphoproteome would remain undetectable in case the enrichment was not carried out at all.

The last large-scale phosphoproteomic dataset published was that of kiwifruit (*Actinidia deliciosa*) [30]. However, this study did not aim at the identification of developmentally related phosphopeptides under normal conditions, but rather at the identification of phosphorylation regulation upon inhibition by MG132. The peptide aldehyde MG132, also named *N*-Benzylloxycarbonyl-L-leucyl-L-leucyl-L-leucinal, represents a proteasome inhibitor [35]. Nevertheless, the crosstalk between protein phosphorylation and degradation in the male gametophyte was described by high-throughput methods the first time. Collectively, 1299 unique phosphopeptides from 711 phosphoproteins were identified, which carried 1572 phosphorylation sites (Table 1). They took part in protein metabolism, RNA and DNA processing, signalling and development. Moreover, many of these phosphoproteins had their homologues in *A. thaliana* and many of them were either annotated in the phosphoproteomic databases or were homologous to Mayank's *A. thaliana* pollen phosphoproteome [25]. However, several candidates were identified in pollen grains newly, a role which might be related to the proteasome inhibition. In general, MG132 treatment caused notable changes in protein phosphorylation, but not in overall protein expression, by which it pinpointed the importance of post-translational modifications for the regulation rather than the synthesis of novel proteins.

This review article has mainly focused on angiosperms. However, Chen et al. (2012) conducted a study investigating the pollen proteome of *Picea wilsonii*, the first gymnosperm to be analysed by a phosphoproteomic approach [36]. Like kiwifruit pollen phosphoproteome, it did not aim at developmental phosphoproteomics since it studied phosphoproteins related to pollen tube growth on media with low sucrose and calcium ion concentration and as such represented the study of phosphorylation upon various stresses.

### 3. Common Phosphoproteins in Angiosperm Male Gametophyte Phosphoproteomes

We compared angiosperm mature pollen phosphoproteomes together (Arabidopsis [25], tobacco [28], and maize [29]) to find the common regulatory trends in male gametophytes of these species (Supplementary Table S1). We did not include kiwifruit pollen in these analyses since it represented a different dataset—activated pollen that was influenced by the addition of dimethyl sulfoxide (DMSO) in case of the negative control or even by MG132-mediated proteasome inhibition [30].

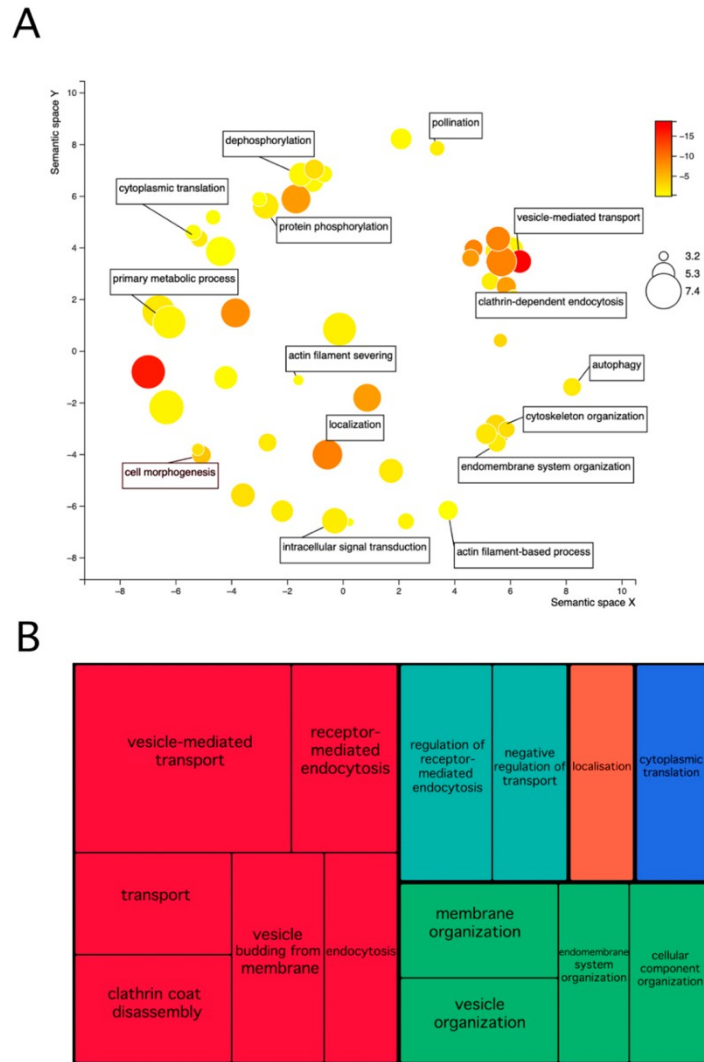
First, we compared Arabidopsis and maize pollen phosphoproteomes. As was mentioned above, the Arabidopsis mature pollen phosphoproteome presented 598 phosphoproteins, whereas in maize pollen, there were identified 2257 phosphoproteins. The maize GRMZM identifiers of genome assembly R73\_RefGEN\_v3 were converted to Zm identifiers with MaizeMine v 1.3, and the homologue search between maize and Arabidopsis

were executed by the engines on the same webpage. The comparison of maize AGI homologues with Arabidopsis pollen phosphoproteome resulted in 323 unique identifiers (527 in total, Table S1). To unravel the biological significance of these phosphoproteins, we carried out enrichment analyses for gene ontology terms and KEGG pathway (Figure 1A). The list included 54 transport proteins, mainly taking part in vesicular transport of all three types—COP1, COP2, and clathrin-coated vesicles. Besides these, there appeared also proteins playing their roles in endocytosis and vesicle movement on actin filaments. Then, in connection with pollen desiccation, 13 genes related to salt stress were present in both datasets. Both phosphoproteomes also shared proteins responsible for pollen tube growth and 14 proteins seem to possess a double function, since they were annotated with functions in root development. It is likely that these candidates are common to root hairs and pollen tubes since these tissues share the same mechanisms of tip growth. Moreover, 26 proteins responsible for protein phosphorylation were present. Amongst them, there were 3 mitogen-activated protein kinases—MAPK (At1g18150, At1g73670, and At3g07980), and a cyclin-dependent protein kinase CDK (At4g28980)—that recognize the prolyl-directed phosphorylation motifs (xxxxxxS\*Pxxxxx, and xxxxxxT\*Pxxxxx) [37]. Then there were 2 casein kinases—CK (At4g26100, and At5g57015) that target the acidic motifs xxxxxxS\*DxExxx, and xxxxxxS\*xDDxxx [37]. Finally, the basic motifs (xxxRxxS\*xxxxxx, and xxxKxxS\*xxxxxx) [37] were recognized by SNF1-related protein kinases—SnRK (At1g09020, At3g01090, and At3g29160), and Ca<sup>2+</sup>-dependent protein kinases—CDPK (At1g35670, and At4g09570).

All mentioned kinase families were reported to play important roles during pollen tube growth [38]. The CDKs appear in the phosphoproteomic datasets since they are required for cell divisions that are part of male gametophyte development [39] and for their activity, they require to be phosphorylated by CDK-activating kinases [40]. Then, they regulate pre-mRNA splicing of callose synthase in pollen tubes to control the formation of a cell wall [41]. SnRKs were already reported to play a key role in pollen germination, where its mutation resulted in the compromised pollen hydration on the stigma [42]. Moreover, the SnRK-mediated phosphorylation is involved in communication by reactive oxygen species [43]. Then, CPK11 and CPK24 were involved in Ca<sup>2+</sup>-dependent regulation of the K<sup>+</sup> channels [44] and CPK6 was reported to phosphorylate actin depolymerizing factor 1, by which the dynamics of actin filaments are regulated [45].

All motifs recognized by the mentioned kinase families usually appeared as over-represented in pollen phosphoproteomes. To test whether the phosphorylation sites in kinases are conserved between Arabidopsis and maize pollen phosphoproteomes, we compared the exact positions of phosphorylation sites in these datasets together. There was one common phosphorylation site, particularly VSFNDTPSAIFWT\*DYVATR in mitogen-activated protein kinase 8 (At1g18150, and its maize homologue GRMZM2G062761). Then, several other phosphopeptides carry most likely the conserved phosphorylation site, but the phosphorylation position in the Arabidopsis dataset was unfortunately not identified conclusively. However, these proteins share at least the peptide sequence with maize pollen phosphoproteome: serine/threonine-protein kinase SRK2A (At1g10940), serine/threonine-protein kinase SRK2G (At5g08590), serine/threonine-protein kinase SRK2H (At5g63650), SNF1-related protein kinase catalytic subunit  $\alpha$  KIN10 (At3g01090), SNF1-related protein kinase catalytic subunit  $\alpha$  KIN11 (At3g29160), and Shaggy-related protein kinase iota (At1g06390). Collectively, most kinases with conserved phosphopeptides between maize and Arabidopsis pollen phosphoproteomes were represented by the kinases, with known phosphorylation motifs in the phosphoproteomic datasets.





**Figure 1.** Comparison of pollen phosphoproteomes. **(A)**—GO biological processes enrichment analysis of phosphoproteins common to Arabidopsis and maize. The colours represent the false discovery rate of the enriched term, and the size of the circle represents the relative size of the GO term. **(B)**—A Treemap of enriched GO biological processes among phosphoproteins present in all three pollen samples (Arabidopsis, maize, and tobacco). The plots in **(A,B)** were rendered by Revigo [46].

Among the enriched molecular processes, the phosphoproteins shared between Arabidopsis and maize were divided into the following groups: 141 proteins had a binding capacity and were further distinguished as RNA binding, which included, for example, 9 translation initiation factors, cytoskeletal protein binding, phosphatidylinositol binding or AMP binding. The second distinct group consisted of protein kinases and kinase activators. Finally, three proteins were annotated to localize into the polarized growth, namely *KINKY POLLEN* (At5g49680), putative clathrin assembly protein (At1g03050), and receptor-like kinase lost in pollen tube guidance—LIP1 (At5g16500). These regulatory proteins represent conserved candidates with an important role for pollen tube growth and guidance. The protein *KINKY POLLEN* was reported to play its role in vesicular transport both in pollen tubes and in root hairs, so its mutations led to an aberrant pollen tube [47]. Then, the LIP1 receptor-like kinase was important for pollen tube guidance [48].

The homologues to *Nicotiana tabacum* sequences were retrieved with an NCBI command line blastn tool, with dc-megablast task [49]. The top hits were used for further analysis resulting in a list of 170 unique AGI identifiers. Of these, 55 phosphoproteins were shared exclusively with maize, 21 were common exclusively with Arabidopsis, and 31 were shared with both these datasets, leaving aside 63 unique unshared phosphopeptides (Table S1). The enrichment of biological functions and molecular processes in GO term analysis was similar to the other studied species. The phosphoproteins were involved in endocytosis and vesicular transport. Furthermore, translation and mRNA processing represented the enriched processes, which may correspond to the activated state of pollen and preparation for pollen tube burst. As for the molecular function, 58 proteins were involved in protein binding, from which 30 candidates were annotated as RNA binding. Finally, three proteins, with their functions in chromatin structure, were present.

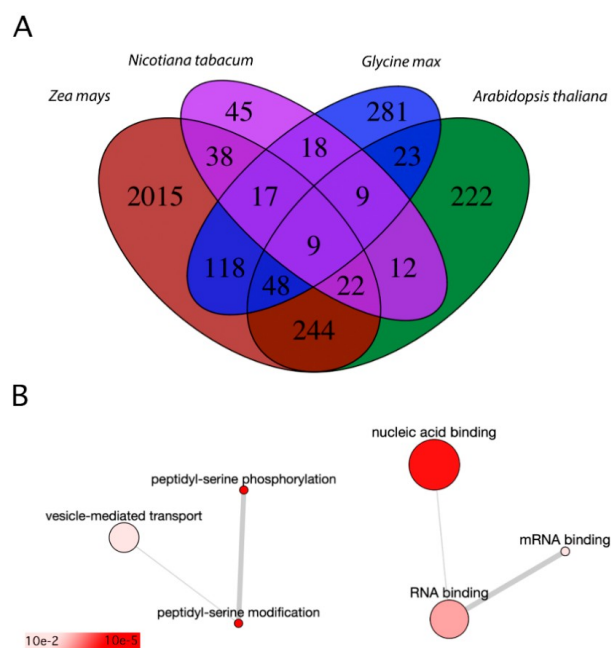
Considering the overlaps in the pollen datasets, maize and Arabidopsis pollen phosphoproteomes showed a much higher similarity to each other than to the tobacco dataset. This could have been caused by (1) lower total number of phosphoproteins in the tobacco dataset; (2) the type of tissue, since tobacco studies also included 5-min activated and 30-min activated pollen; and (3) pollen type, since both maize and Arabidopsis share tri-cellular pollen, whereas tobacco sheds bi-cellular pollen [5]. Nevertheless, 31 phosphoprotein homologues were present in all 3 datasets. These common proteins include mostly candidates taking part in vesicular transport, suggesting that they represent the basic conserved mechanisms which are important for pollen development (Figure 1B).

To conclude, the phosphoproteins shared at least by some species belonged to similar functional categories. There was usually at least some of the categories typical for tip growth—small GTPase signalling, ion gradient formation, cytoskeleton organization together with vesicular transport [18–20]. Then, the genes, which take part in regulatory mechanisms in protein synthesis, were also amongst the abundant functional categories. Overall, the phosphorylation of specific protein involved mainly in pollen tube growth seems to be conserved in the plant's evolution.

#### 4. Common Trends for Male Gametophyte and Root Hairs

After comparing the three male gametophyte phosphoproteomes (Arabidopsis, tobacco, and maize) together and pinpointing common phosphoproteins, the same male gametophyte datasets were compared with root hair phosphoproteome (Figure 2A, Supplementary Table S2). Root hairs and pollen tubes share the same type of growth—tip growth—and due to this, they rely on the common regulatory mechanisms, such as small GTPase signalling, ion gradient formation, cytoskeleton organization together with regulations of vesicular transport, reactive oxygen species (ROS) signalling and a massive decrease in pH [18–20,50]. The only available root hairs phosphoproteome belongs to soybean (*Glycine max*), the roots of which were studied with respect to the nodule formation that accommodate the symbiotic nitrogen-fixing bacteria, typical for leguminous plants [51]. They presented both root hair phosphoproteome and the phosphoproteome of the corresponding shaved roots (i.e., roots with removed root hairs). Collectively,

the study led to the identification of 1625 phosphopeptides carrying 1659 phosphorylation sites, which belonged to 1126 phosphoproteins. These phosphoproteins were assigned to the following functional categories: DNA/RNA-related proteins, signal transduction, miscellaneous group (proteins with multiple functions), and protein trafficking.



**Figure 2.** Comparison of pollen phosphoproteomes with root hair phosphoproteome. (A)—Venn diagram shows the overlap of several Arabidopsis homologue phosphoproteins discovered in pollen samples of Arabidopsis (*Arabidopsis thaliana*), maize (*Zea mays*), and tobacco (*Nicotiana tabacum*) together with soybean (*Glycine max*) root hair phosphoproteome. (B)—The biological processes (left) and molecular function (right) GO terms of phosphoproteins shared between the Arabidopsis pollen phosphoproteome and soybean root hair phosphoproteome. The colour represents false discovery rate of the enriched term. The plots were rendered by Revigo [46].

To establish the shared regulatory pathways between polarized tip growth of pollen tubes and root hairs, we compared Arabidopsis mature pollen phosphoproteome (since its genome is from the studied species annotated best [52]) with soybean root hair phosphoproteome. Only phosphopeptides that were present in the root hairs were used for further comparative analyses since phosphopeptides identified solely in the shaved roots were removed for being irrelevant to root hairs. There were 825 annotated phosphopeptides present in the root hairs, and the Arabidopsis homologues were retrieved from the Phytome database [53] using the Wm82.a4.v1 as a reference genome [54,55]. In total, 254 proteins (represented by 89 unique AGI identifiers) were shared between Arabidopsis pollen phosphoproteome and soybean root hair phosphoproteome (Table S2). These included proteins taking part in peptidyl-serine phosphorylation (5 proteins) and vesicle-mediated transport (11 proteins). These 5 kinases were represented by calcium-dependent protein kinase 4 (At4g09570), calcium-dependent protein kinase 11 (At1g35670), 3-phosphoinositide-dependent protein kinase 1 (At5g04510), 3-phosphoinositide-dependent protein kinase 2 (At3g10540), and casein kinase 1-like protein 1 (At4g26100). As mentioned above, CPK11 was involved in the  $\text{Ca}^{2+}$ -dependent regulation of the  $\text{K}^+$  channels [44], whereas it in-

hibited (together with CPK4) the root growth by phosphorylation of 1-aminocyclopropane-1-carboxylate synthase, by which its activity during ethylene synthesis was increased [56]. The 3-phosphoinositide-dependent protein kinase 1 was important in several physiological processes where cell proliferation and growth are of key importance [57], and it was proven to be the regulator of AGC1 kinases [58]. The casein kinase 1-like protein 1 regulated cell division by the phosphorylation of Kip-related protein 6 [59] and their other activities throughout plant development were reviewed recently [60]. Then, if biological function was considered, 36 proteins were reported to show the binding capacity (Figure 2B).

Most of these shared proteins were common also to maize pollen phosphoproteome (57 out of 89) or tobacco phosphoproteome (18 out of 89). Collectively, there were 9 common phosphoproteins for all compared datasets (including tobacco, maize and *Arabidopsis* pollen phosphoproteomes, and soybean root hairs phosphoproteome). These proteins had the following AGI identifiers: At1g11360, At1g20760, At1g21630, At1g59610, At5g57870, At1g20110, At5g41950, At5g52200, and At4g35890. Six of these proteins were functionally annotated; there were candidates working in RNA metabolism (La-related protein 1), translation initiation factor 4G-1, protein phosphatase inhibitor 2, dynamin 2B, and proteins FREE1, and HLB1. In summary, the phosphorylation of tip growth regulators seems to be partially conserved between pollen tip growth and root hair tip growth. However, these modifications can probably maintain a different role in each tissue. In future contexts, it may prove interesting to repeat this comparison within one species.

### 5. Beyond Pollen Phosphoproteomics

In the previous sections, we considered the published male gametophyte phosphoproteomes. However, it should be mentioned that also whole anthers were subjected to phosphoproteomic studies. Although mature pollen grains are part of anther samples, the surrounding sporophyte tissues usually dominate. Ye and colleagues identified the proteome and phosphoproteome of *Arabidopsis thaliana* anthers [61]. In total, they identified 3908 phosphorylation sites on 1637 phosphoproteins. Amongst these 1637 phosphoproteins, there appeared 493 newly identified ones, whereas the others were already deposited to the public phosphoproteomic database and/or were identified in Mayank's mature pollen phosphoproteome [25]. The other species with known anther phosphoproteome were represented by kenaf [?] and rice [61]. Unfortunately, there are not any pollen phosphoproteomic datasets for these species, so a direct comparison of mature pollen and anther phosphoproteomes is not currently possible.

### 6. Conclusions

The studies of protein phosphorylation in angiosperm male gametophyte initiated in 2012 by *Arabidopsis thaliana* mature pollen phosphoproteome. After almost 10 years, there appeared more studies, namely on tobacco, maize, and kiwifruit. However, the kiwifruit study was performed with respect to proteasome inhibition by MG132, but not to pollen development under normal conditions. The only activated pollen phosphoproteome is represented so far by the dataset from tobacco. For the future, the activated pollen of more species should be studied and compared to mature pollen since the phosphorylation dynamics is the most interesting aspect of their post-translational modifications.

The comparison of mature pollen phosphoproteomes between different angiosperm species revealed that the common phosphoproteins played their role in the vitally important processes for pollen tube growth—vesicular transport, metabolism, protein phosphorylation, and cytoskeleton dynamics. It seems that the basic cellular processes are conserved even between monocots and dicots, but the number of available datasets remains limited. For the future, the data acquired on more species should enable the comparison of mature pollen from both monocots and dicots with both bicellular and tricellular pollen (recently reviewed in [? ]). Such a comparison will most likely highlight the pollen mitosis II-related kinases and other regulatory proteins.





After the decade of pollen phosphoproteomics, the research is surely not finished and deserves our future interest, especially on the emphasis of activated pollen.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms222212212/s1>. Supplementary Table S1—The comparison of pollen phosphoproteomes from maize (*Zea mays*) [29], tobacco (*Nicotiana tabacum*) [28], and *Arabidopsis thaliana* [25]. Supplementary Table S2—The comparison of pollen phosphoproteomes (from maize [29], tobacco [28], and *Arabidopsis* [25]) with soybean root hair phosphoproteome [51].

**Author Contributions:** Conceptualization, J.F.; data analysis, B.K.; writing—original draft preparation, B.K. and J.F.; writing—review and editing, J.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Czech Ministry of Education, Youth and Sports [LTC20050] and the Czech Science Foundation [19-01723S].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We acknowledge David Honys for his valuable comments during revision of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

2D-GE	two-dimensional gel electrophoresis
AMP	adenosine monophosphate
CDK	cyclin-dependent protein kinase
CDPK	calcium-dependent protein kinase
CK	casein kinase
DMSO	dimethyl sulfoxide
EDTA	2,2',2'',2'''-(Ethane-1,2-diyl)dinitrilo)tetraacetic acid
EPP	EDTA/puromycine-resistant particles
GO	gene ontology
IMAC	immobilized metal affinity chromatography
LRRK	leucine rich repeat kinase
MAPK	mitogen-activated protein kinase
MOAC	metal oxide affinity chromatography
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NCBI	National Center for Biotechnology Information
nLC	nano liquid chromatography
ROS	reactive oxygen species
SIMAC	sequential elution from IMAC
SnRK	SNF1-related protein kinase

## References

1. Friedman, W.E.; Floyd, S.K. Perspective: The origin of flowering plants and their reproductive biology—A tale of two phylogenies. *Evolution* **2001**, *55*, 217–231. [[PubMed](#)]
2. Christensen, C.A.; Subramanian, S.; Drews, G.N. Identification of gametophytic mutations affecting female gametophyte development in *Arabidopsis*. *Dev. Biol.* **1998**, *202*, 136–151. [[CrossRef](#)] [[PubMed](#)]
3. Borg, M.; Twell, D. Life after meiosis: Patterning the angiosperm male gametophyte. *Biochem. Soc. T.* **2010**, *38*, 577–582. [[CrossRef](#)] [[PubMed](#)]
4. Williams, J.H.; Taylor, M.L.; O'Meara, B.C. Repeated evolution of tricellular (and bicellular) pollen. *Am. J. Bot.* **2014**, *101*, 559–571. [[CrossRef](#)]





34. Edwards, K.D.; Fernandez-Pozo, N.; Drake-Stowe, K.; Humphry, M.; Evans, A.D.; Bombarely, A.; Allen, F.; Hurst, R.; White, B.; Kernodle, S.P.; et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genom.* **2017**, *18*, 448. [[CrossRef](#)]
35. Lee, D.H.; Goldberg, A.L. Proteasome inhibitors: Valuable new tools for cell biologists. *Trends Cell Biol.* **1998**, *8*, 397–403. [[CrossRef](#)]
36. Chen, Y.; Liu, P.; Hoehenwarter, W.; Lin, J. Proteomic and phosphoproteomic analysis of *Picea wilsonii* pollen development under nutrient limitation. *J. Proteome Res.* **2012**, *11*, 4180–4190. [[CrossRef](#)]
37. Lee, T.Y.; Lin, Z.Q.; Hsieh, S.J.; Bretana, N.A.; Lu, C.T. Exploiting maximal dependence decomposition to identify conserved motifs from a group of aligned signal sequences. *Bioinformatics* **2011**, *27*, 1780–1787. [[CrossRef](#)]
38. Heberle-Bors, E.; Voronin, V.; Touraev, A.; Testillano, P.S.; Risueno, M.C.; Wilson, C. MAP kinase signaling during pollen development. *Sex. Plant Reprod.* **2001**, *14*, 15–19. [[CrossRef](#)]
39. Takatsuka, H.; Umeda-Hara, C.; Umeda, M. Cyclin-dependent kinase-activating kinases CDKD1 and CDKD3 are essential for preserving mitotic activity in *Arabidopsis thaliana*. *Plant J.* **2015**, *82*, 1004–1017. [[CrossRef](#)]
40. Shimotohno, A.; Matsubayashi, S.; Yamaguchi, M.; Uchimiya, H.; Umeda, M. Differential phosphorylation activities of CDK-activating kinases in *Arabidopsis thaliana*. *FEBS Lett.* **2003**, *534*, 69–74. [[CrossRef](#)]
41. Huang, X.-Y.; Niu, J.; Sun, M.-X.; Zhu, J.; Gao, J.-F.; Yang, J.; Zhou, Q.; Yang, Z.-N. CYCLIN-DEPENDENT KINASE G1 is associated with the spliceosome to regulate CALLOSE SYNTHASE 5 splicing and pollen wall formation in *Arabidopsis*. *Plant Cell* **2013**, *25*, 637–648. [[CrossRef](#)] [[PubMed](#)]
42. Li, D.D.; Guan, H.; Li, F.; Liu, C.Z.; Dong, Y.X.; Zhang, X.S.; Gao, X.Q. Arabidopsis shaker pollen inward K<sup>+</sup> channel SPIK functions in SnRK1 complex-regulated pollen hydration on the stigma. *J. Integr. Plant Biol.* **2017**, *59*, 604–611. [[CrossRef](#)] [[PubMed](#)]
43. Gao, X.Q.; Liu, C.Z.; Li, D.D.; Zhao, T.T.; Li, F.; Jia, X.N.; Zhao, X.Y.; Zhang, X.S. The Arabidopsis KIN beta gamma subunit of the SnRK1 complex regulates pollen hydration on the stigma by mediating the level of reactive oxygen species in pollen. *PLoS Genet.* **2016**, *12*, e1006228. [[CrossRef](#)] [[PubMed](#)]
44. Zhao, L.-N.; Shen, L.-K.; Zhang, W.-Z.; Zhang, W.; Wang, Y.; Wu, W.-H. Ca<sup>2+</sup>-dependent protein kinase 11 and 24 modulate the activity of the inward rectifying K<sup>+</sup> channels in Arabidopsis pollen tubes. *Plant Cell* **2013**, *25*, 649–661. [[CrossRef](#)]
45. Dong, C.H.; Hong, Y. Arabidopsis CDPK6 phosphorylates ADF1 at N-terminal serine 6 predominantly. *Plant Cell Rep.* **2013**, *32*, 1715–1728. [[CrossRef](#)] [[PubMed](#)]
46. Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [[CrossRef](#)]
47. Prociassi, A.; Guyon, A.; Pierson, E.S.; Giritch, A.; Knuiman, B.; Grandjean, O.; Tonelli, C.; Derksen, J.; Pelletier, G.; Bonhomme, S. KINKY POLLEN encodes a SABRE-like protein required for tip growth in Arabidopsis and conserved among eukaryotes. *Plant J.* **2003**, *36*, 894–904. [[CrossRef](#)]
48. Liu, J.J.; Zhong, S.; Guo, X.Y.; Hao, L.H.; Wei, X.L.; Huang, Q.P.; Hou, Y.N.; Shi, J.; Wang, C.Y.; Gu, H.Y.; et al. Membrane-bound RLCKs LIP1 and LIP2 are essential male factors controlling male–female attraction in Arabidopsis. *Curr. Biol.* **2013**, *23*, 993–998. [[CrossRef](#)]
49. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST plus: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
50. Zhang, M.J.; Zhang, X.S.; Gao, X.Q. ROS in the male–female interactions during pollination: Function and regulation. *Front. Plant Sci.* **2020**, *11*, 177. [[CrossRef](#)]
51. Tran, H.N.N.; Brechenmacher, L.; Aldrich, J.T.; Clauss, T.R.; Gritsenko, M.A.; Hixson, K.K.; Libault, M.; Tanaka, K.; Yang, F.; Yao, Q.M.; et al. Quantitative phosphoproteomic analysis of soybean root hairs inoculated with *Bradyrhizobium japonicum*. *Mol. Cell. Proteom.* **2012**, *11*, 1140–1155.
52. Cheng, C.Y.; Krishnakumar, V.; Chan, A.P.; Thibaud-Nissen, F.; Schobel, S.; Town, C.D. Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **2017**, *89*, 789–804. [[CrossRef](#)] [[PubMed](#)]
53. Goodstein, D.M.; Shu, S.Q.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [[CrossRef](#)] [[PubMed](#)]
54. Valliyodan, B.; Cannon, S.B.; Bayer, P.E.; Shu, S.Q.; Brown, A.V.; Ren, L.H.; Jenkins, J.; Chung, C.Y.L.; Chan, T.F.; Daum, C.G.; et al. Construction and comparison of three reference-quality genome assemblies for soybean. *Plant J.* **2019**, *100*, 1066–1082. [[CrossRef](#)] [[PubMed](#)]
55. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.X.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.J.; Thelen, J.J.; Cheng, J.L.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463*, 178–183. [[CrossRef](#)]
56. Luo, X.; Chen, Z.; Gao, J.; Gong, Z. Abscisic acid inhibits root growth in Arabidopsis through ethylene biosynthesis. *Plant J.* **2014**, *79*, 44–55. [[CrossRef](#)] [[PubMed](#)]
57. Otterhag, L.; Gustavsson, N.; Alsterfjord, M.; Pical, C.; Lehrach, H.; Gobom, J.; Sommarin, M. Arabidopsis PDK1: Identification of sites important for activity and downstream phosphorylation of S6 kinase. *Biochimie* **2006**, *88*, 11–21. [[CrossRef](#)] [[PubMed](#)]
58. Xiao, Y.; Offringa, R. PDK1 regulates auxin transport and Arabidopsis vascular development through AGC1 kinase PAX. *Nat. Plants* **2020**, *6*, 544–555. [[CrossRef](#)]
59. Qu, L.; Wei, Z.; Chen, H.H.; Liu, T.; Liao, K.; Xue, H.W. Plant casein kinases phosphorylate and destabilize a cyclin-dependent kinase inhibitor to promote cell division. *Plant Physiol.* **2021**, *187*, 917–930. [[CrossRef](#)]

60. Kang, J.M.; Wang, Z. Mut9p-LIKE KINASE family members: New roles of the plant-specific casein kinase I in plant growth and development. *Int. J. Mol. Sci.* **2020**, *21*, 1562. [[CrossRef](#)]
61. Ye, J.Y.; Zhang, Z.B.; Long, H.F.; Zhang, Z.M.; Hong, Y.; Zhang, X.M.; You, C.J.; Liang, W.Q.; Ma, H.; Lu, P.L. Proteomic and phosphoproteomic analyses reveal extensive phosphorylation of regulatory proteins in developing rice anthers. *Plant J.* **2015**, *84*, 527–544. [[CrossRef](#)] [[PubMed](#)]

**The beta subunit of nascent polypeptide associated complex plays a role in flowers and siliques development of *Arabidopsis thaliana***


My contribution: In this article, I was responsible for the sample preparation for RNA-seq and proteomic experiments. I conducted the bioinformatic analysis of the transcriptomic and the functional analysis of proteomic data. Additionally, I contributed to various experiments, including cloning, genotyping, microscopy, and qPCR measurements and analysis. I also co-authored the manuscript and created two figures.

Current status: Published in International Journal of Molecular Sciences. IF<sub>2022</sub>: 5.6, cited 3x (ResearchGate)



Article

# The Beta Subunit of Nascent Polypeptide Associated Complex Plays A Role in Flowers and Siliques Development of *Arabidopsis thaliana*

Jan Fila <sup>1,\*</sup>,<sup>†</sup> , Božena Klodová <sup>1,2,†</sup>, David Potěšil <sup>3</sup>, Miloslav Juříček <sup>4</sup>, Petr Šesták <sup>1,2</sup>, Zbyněk Zdráhal <sup>3,5</sup> and David Honys <sup>1,2</sup>

<sup>1</sup> Laboratory of Pollen Biology, Institute of Experimental Botany of the Czech Academy of Sciences, 16502 Praha 6, Czech Republic; klodova.bozena@gmail.com (B.K.); petr-sestak@seznam.cz (P.S.); david@ueb.cas.cz (D.H.)

<sup>2</sup> Department of Experimental Plant Biology, Faculty of Science, Charles University, 12800 Praha 2, Czech Republic

<sup>3</sup> Mendel Centre for Plant Genomics and Proteomics, Central European Institute of Technology, Masaryk University, 62500 Brno, Czech Republic; david.potesil@ceitec.muni.cz (D.P.); zdrahal@sci.muni.cz (Z.Z.)

<sup>4</sup> Station of Apple Breeding for Disease Resistance, Institute of Experimental Botany of the Czech Academy of Sciences, 16502 Praha 6, Czech Republic; juricek@ueb.cas.cz

<sup>5</sup> Laboratory of Functional Genomics and Proteomics, National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 62500 Brno, Czech Republic

\* Correspondence: fila@ueb.cas.cz; Tel.: +420-225-106-452

† These authors contributed equally to this work.

Received: 14 February 2020; Accepted: 14 March 2020; Published: 17 March 2020



**Abstract:** The nascent polypeptide-associated (NAC) complex was described in yeast as a heterodimer composed of two subunits,  $\alpha$  and  $\beta$ , and was shown to bind to the nascent polypeptides newly emerging from the ribosomes. NAC function was widely described in yeast and several information are also available about its role in plants. The knock down of individual NAC subunit(s) led usually to a higher sensitivity to stress. In *Arabidopsis thaliana* genome, there are five genes encoding NAC $\alpha$  subunit, and two genes encoding NAC $\beta$ . Double homozygous mutant in both genes coding for NAC $\beta$  was acquired, which showed a delayed development compared to the wild type, had abnormal number of flower organs, shorter siliques and greatly reduced seed set. Both NAC $\beta$  genes were characterized in more detail—the phenotype of the double homozygous mutant was complemented by a functional NAC $\beta$  copy. Then, both NAC $\beta$  genes were localized to nuclei and cytoplasm and their promoters were active in many organs (leaves, cauline leaves, flowers, pollen grains, and siliques together with seeds). Since flowers were the most affected organs by *nac $\beta$*  mutation, the flower buds' transcriptome was identified by RNA sequencing, and their proteome by gel-free approach. The differential expression analyses of transcriptomic and proteomic datasets suggest the involvement of NAC $\beta$  subunits in stress responses, male gametophyte development, and photosynthesis.

**Keywords:** *Arabidopsis thaliana*; chaperone; flower bud proteome; flower bud transcriptome; male gametophyte; nascent polypeptide-associated complex

## 1. Introduction

The nascent polypeptide-associated complex (NAC) is present in archaea and eukaryotes, and on the contrary absent from bacteria [1]. The abbreviation NAC should not be confused with the family of transcription factors, called also NAC. Throughout this manuscript, the mentioned abbreviation NAC will be used in the meaning of nascent polypeptide-associated complex. In eukaryotes, the NAC

complex was described as a heterodimer composed of an  $\alpha$ - and a  $\beta$ -subunit [1] whereas archaeal NAC was represented by an  $\alpha$ -homodimer [2]. By now, the roles of NAC complex were mainly studied in yeast and animals [3]. Yeast genome encodes for two *NAC $\beta$*  homologues (named  $\beta$ 1, and  $\beta$ 3), and a single *NAC $\alpha$*  gene. The yeast *NAC $\beta$*  proteins differed in their affinity to the  $\alpha$ -subunit since the  $\alpha/\beta$ 1 complex was significantly more abundant than the alternative dimer  $\alpha/\beta$ 3 [4]. Beside this, both isoforms of the NAC complex play likely distinct roles since the function of Ssb chaperone complex was complemented by both individual isoforms of NAC complex with a different efficiency [5]. NAC complex formed by  $\alpha/\beta$  heterodimer usually binds to the newly synthesized nascent polypeptides emerging from the ribosome, and thus functions as a chaperone [6,7]. It is able to bind proteins in various conformations, such as unfolded, folded, or intrinsically disordered ones [8], and likely interacts with the ribosome via the C-terminal domain of the  $\beta$ -subunit [9]. Moreover, NAC complex has likely more functions on the cellular level, for instance in protein sorting to endoplasmic reticulum [7,10] or to mitochondrion [11]. Nevertheless, the role of NAC complex in protein sorting to mitochondria was observed exclusively in vitro but not in vivo. Thus, the role of NAC complex in protein sorting in yeast is not yet confirmed on the molecular level. However, the role of NAC complex in protein sorting was proven in *Caenorhabditis elegans* where it blocked the ribosome binding site on the endoplasmic reticulum in case the protein should reside in the cytoplasm [12]. Besides its function as a chaperone and in protein sorting, NAC was also hypothesized to play a role in ribosome biogenesis [13]. In higher eukaryotes, the NAC complex was essential for embryo development since its knock down caused embryonic lethality in mouse, nematode and drosophila [14–16]. Until now, the function of NAC subunits forming the heterodimeric complex was considered but they can both play their roles also individually as single subunits. The *NAC $\beta$*  subunit is also known as basic transcription factor 3 (BTF3) [15–17], whereas the sole  $\alpha$ -subunit works as a transcriptional co-activator [18,19].

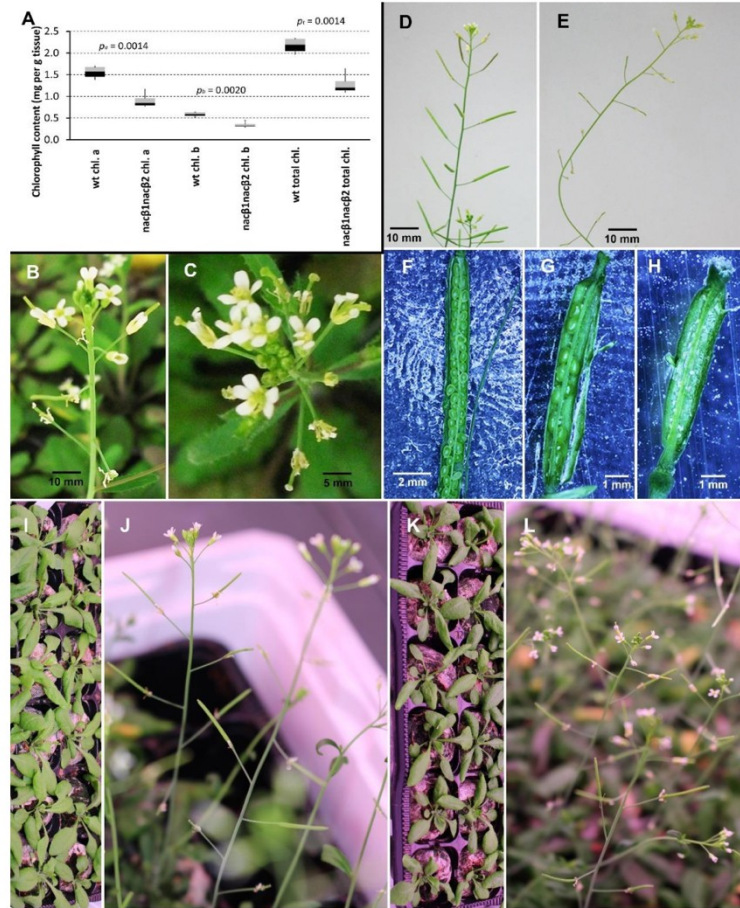
Several information about the role of the NAC complex are also known in plants. The complexity of the formed heterodimer variants could be even higher in *Arabidopsis thaliana* since its genome encodes for five *NAC $\alpha$*  genes (*NAC $\alpha$ 1*—At3g12390, *NAC $\alpha$ 2*—At3g49470, *NAC $\alpha$ 3*—At5g13850, *NAC $\alpha$ 4*—At4g10480, *NAC $\alpha$ 5*—At1g33040) and two *NAC $\beta$*  homologues (*NAC $\beta$ 1*—At1g73230, and *NAC $\beta$ 2*—At1g17880) according to the TAIR database [20]. *Arabidopsis NAC $\beta$ 1*, and *NAC $\beta$ 2* (named in the study as BTF3, and BTF3L) were reported to play their roles in cold stress, during which were phosphorylated by OST1 [21]. *NAC $\alpha$ 2* was reported to interact with phytochromes in the cytoplasm [22], and *NAC $\beta$*  was shown to interact with eIF(iso)4E [23]. The other plant studies regarding NAC complex or its subunits were carried out on different species than *Arabidopsis thaliana*. Firstly, the NAC role in plant development was reported. *NAC $\beta$*  in *Nicotiana benthamiana* was shown to play its role during leaf development since its leaves showed an abnormal morphology with reduced chloroplast size and chlorophyll content after *NAC $\beta$*  silencing [24]. A similar *NAC $\beta$*  function was reported in wheat (*Triticum aestivum*) [25] and a role in seed germination and seedling growth was shown in rice (*Oryza sativa*) [26]. Secondly, NAC complex plays a key role under stress conditions, particularly during drought or salt stress in barley (*Hordeum marinum*) [27] and in transgenic *A. thaliana* overexpressing *NAC* genes from *Spartina alterniflora* [28]. *NAC $\beta$*  by itself was also shown to be responsible for tolerance to freeze and drought stress in wheat (*Triticum aestivum*) [29], and for flooding tolerance in soybean (*Glycine max*) [30]. Not only abiotic stress but also *NAC $\beta$*  role during hypersensitive response to pathogens in hot pepper (*Capsicum annuum*) [31] was reported. Last but not least, in *Nicotiana benthamiana*, the role of *NAC $\alpha$*  in cell-to-cell movement of Brome mosaic virus (BMV) was presented [32].

Herein, we investigated the role of *NAC $\beta$*  subunits in the flowers and siliques development of *Arabidopsis thaliana*; we characterized both *NAC $\beta$*  genes (complementation analysis, overexpression, subcellular localization, promoter activity) and identified and characterized flower bud transcriptome and proteome of double homozygous mutant plants in both *NAC $\beta$*  genes, suggesting the likely role of *NAC $\beta$*  proteins in stress responses, male gametophyte development, and photosynthesis.

## 2. Results

### 2.1. *nacβ1nacβ2* Double Homozygous Plants Have Abnormal Flowers and Reduced Seed Set

The seeds of *Arabidopsis thaliana* T-DNA insertion lines bearing the insertion in each of the two genes coding for NACβ (SALK\_043673 in *NACβ1—nacβ1* and GK-368\_H02 in *NACβ2—nacβ2*; the lower-case letters in italics represent a homozygous mutant) were sown, and subsequently conventionally genotyped by polymerase chain reaction (PCR) combining gene-specific and insert-specific primer pairs. None of the studied homozygous mutants showed any significant changes in their phenotypic traits (such as whole plant habitus, and various organ traits, including mature pollen) when compared to the Columbia-0 wild type (Col-0 wt) plants. The two mentioned T-DNA insertion lines were crossed with each other (Figure S1D,E). According to quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) data, the SALK\_043673 line represented a knockout, whereas GK-368\_H02 carried knocked down target gene (Figure S1F). The first generation after the cross consisted of heterozygous plants in both studied genes (*nacβ1/+;nacβ2/+*), and by a series of self-crosses, the double homozygous mutant in both NACβ genes (*nacβ1nacβ2*) was acquired. The genotype of the crossed plants and their offspring was determined by a set of conventional PCR reactions combining gene-specific and insert-specific primer pairs. Exclusively, the *nacβ1nacβ2* double homozygous plants showed phenotypic defects compared to the Col-0 wt plants as follows (Figure 1; Figure S1A–C; Figure S2). (i) The *nacβ1nacβ2* plants were delayed in their development for 10–14 days compared to the control Col-0 wt plants. Although the *nacβ1nacβ2* germinated only 1–3 days later or nearly at the same time as the Col-0 wt plants did, they lagged longer (10–14 days) during stem formation, setting the first flowers, setting the first siliques, and finally starting seed maturation (Figure S1A–C). (ii) The average chlorophyll content in the *nacβ1nacβ2* leaves (1.28 mg total chlorophyll per g tissue) represented only three fifths of the Col-0 wt leaves (2.17 mg total chlorophyll per g tissue) (Figure 1A). However, the ratio of chlorophyll a to chlorophyll b remained practically the same (2.62 or 2.65, respectively). (iii) Around 90% flowers of the *nacβ1nacβ2* showed abnormal number of flower organs, mostly 5–6 sepals, 5–6 petals, and 7 anthers but there was not only one distinct phenotype category but several of them (Figure 1B,C). The most abundant category (28% total flowers) was represented by flowers with 5 sepals, 5 petals, and 6 anthers followed by 12% flowers with 5 sepals, 5 petals, and 7 anthers (Figure S2A). Furthermore, the gynoecia in notable part of flowers were not grown together. (iv) The *nacβ1nacβ2* siliques were significantly shorter than the Col-0 wt ones; they reached in average only 5 mm (Figure 1E, Figure S2B) compared to the Col-0 wt average of 14 mm (Figure 1D, Figure S2B). (v) The phenotypic characteristics were obvious also inside the siliques: the *nacβ1nacβ2* plants produced significantly less seeds (median of total seeds per silique reached 15, whereas the Col-0 wt had the median of 53 seeds per silique, Figure S2C). Furthermore, the proportion of the normal-sized and -coloured seeds in the *nacβ1nacβ2* siliques reached 14–23% (Figure 1G,H, Figure S2D) since most of the seeds were represented by the aborted embryo sacks that did not undergo fertilization. On the other hand, the Col-0 wt siliques were filled with 98% viable seeds (Figure 1F, Figure S2D).

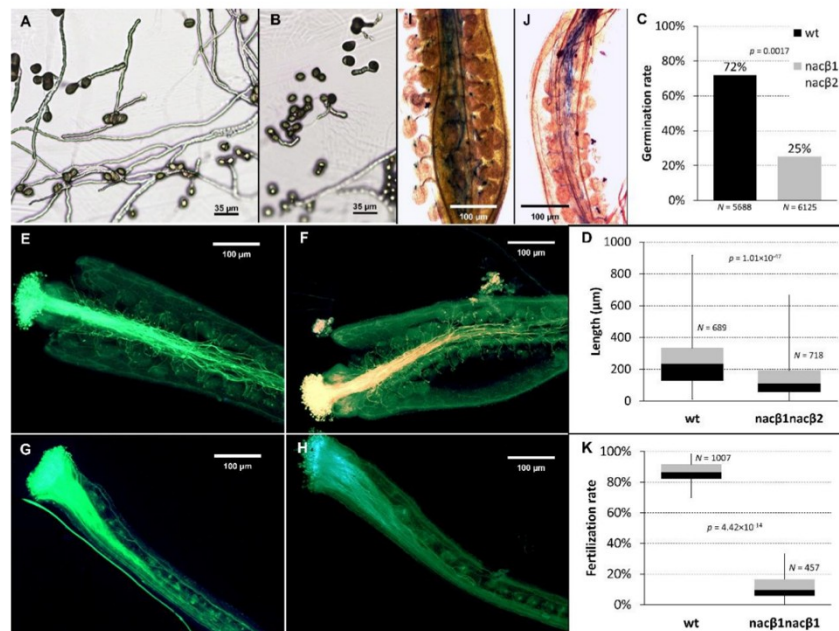


**Figure 1.** Phenotype defects of the *nacβ1nacβ2* plants and their complementation. (A)—Box-and-whisker plot showing the median and quartiles from four independent experiments revealing the chlorophyll a, chlorophyll b, and total chlorophyll content in the *nacβ1nacβ2*, and Col-0 wild type (wt) leaves. The obtained values of chlorophyll content in *nacβ1nacβ2*, and Col-0 were compared by Student's t-test. The calculated *p*-values of three comparisons are given:  $p_a$ —chlorophyll a;  $p_b$ —chlorophyll b;  $p_t$ —total chlorophyll. (B)—Col-0 wt inflorescence with normally looking flowers. Scale bar represents 10 mm. (C)—*nacβ1nacβ2* inflorescence with flowers containing different number of floral organs. Scale bar represents 5 mm. (D)—Col-0 wt siliques of a normal size with a normal seed set. Scale bar represents 10 mm. (E)—*nacβ1nacβ2* siliques, which were significantly shorter. Scale bar represents 10 mm. (F)—Col-0 wt seeds inside unripen siliques, most of the seeds are normal-sized and green. Scale bar represents 2 mm. (G)—*nacβ1nacβ2* seeds inside the siliques of less severe phenotype, several normal-sized and green seeds are combined with aborted seeds. Scale bar represents 1 mm. (H)—*nacβ1nacβ2* seeds inside the siliques of more severe phenotype where most seeds are aborted. Scale bar represents 1 mm. (I, J)—*nacβ1nacβ2* plants complemented with proNACβ1::NACβ1-GFP construct. I—Plants with leaf rosettes, 27 day after seed sowing. J—Flowering plants 36 days after seed sowing. (K, L)—*nacβ1nacβ2* plants complemented with proNACβ2::NACβ2-GFP construct. K—Plants with leaf rosettes, 27 day after seed sowing. L—Flowering plants 36 days after seed sowing.

## 2.2. *nacβ1nacβ2* Double Mutation Compromises Progamic Phase and Ovule Targeting

Since the *nacβ1nacβ2* mature pollen grains did not show any developmental defects compared to the Col-0 wt as described in Reňák et al. (2011) [33], the processes in male gametophyte development and in pollen tube growth and targeting that underly the formation of *nacβ1nacβ2* phenotype were tested. (i) *In vitro* pollen tube cultivation. Under the *in vitro* conditions [34], *nacβ1nacβ2* pollen grains germinated less efficiently compared to Col-0 wt since only, on average, 25% *nacβ1nacβ2* pollen grains germinated (Figure 2B,C). On the other hand, the Col-0 wt pollen tubes showed average 72% germination (Figure 2A,C). Moreover, *nacβ1nacβ2* pollen grains showed not only a reduced germination rate but also pollen tubes after 8 h cultivation were shorter than the Col-0 wt ones. The *nacβ1nacβ2* pollen tubes reached average 142μm, whereas the control Col-0 wt pollen tubes reached almost double average length of 252μm (Figure 2D). These data suggested that pollen germination together with pollen tube growth were affected in the *nacβ1nacβ2* plants. However, the pollen germination *in vitro* can be different from *in vivo* since the cultivation *in vitro* lacks female gametophyte cues, which strongly influence the behavior of male gametophyte [35]. (ii) *Aniline blue staining of the in vivo-grown pollen tubes*. To address this limitation of the *in vitro* pollen tube cultivation, the pollen grains were germinated again, this time *in vivo*. The flower buds to be pollinated were emasculated the day before, and the pollen grains were let to germinate on their pistils. After the overnight growth, the callose in the cell wall of the grown pollen tubes was stained by aniline blue [36]. The *nacβ1nacβ2* embryo sacs were less efficient in attracting the Col-0 wt pollen tubes (Figure 2F) compared to the Col-0 wt embryo sacs (Figure 2E). Similarly, the *nacβ1nacβ2* pollen tubes grew more slowly and were less efficiently attracted by the Col-0 wt embryo sacs (Figure 2G). In case of *nacβ1nacβ2* pistils pollinated with *nacβ1nacβ2* pollen grains, the targeting efficiency of the pollen tubes was dramatically reduced (Figure 2H). The *in vivo* pollen tube growth verified that *nacβ1nacβ2* pollen tube growth was negatively influenced and suggested that *nacβ1nacβ2* showed also defects in female gametophyte. (iii) *Blue dot assay*. Aniline blue staining of the *in vivo* grown pollen tubes provides with information about ovule targeting but does not enable quantification of the targeted versus untargeted ovules. To score the fertilization events, blue dot assay was performed [37]. By calculating the proportion of targeted embryo sacs by wt pollen tubes (carrying the proLAT52–GUS construct), it was possible to decipher the targeting efficiency, which reached in case of *nacβ1nacβ2* pistils 11 % (Figure 2J,K) whilst in case of Col-0 wt pistils was much higher, 82% (Figure 2L,K).





**Figure 2.** Functional tests revealing the causes of the *nacβ1nacβ2* phenotype. (A)—Col-0 wt pollen tubes cultivated 8 hours in vitro. Scale bar represents 35 μm. (B)—*nacβ1nacβ2* pollen tubes cultivated 8 hours in vitro. Scale bar represents 35 μm. (C)—Column chart showing average germination efficiency in three independent experiments of the Col-0 wt pollen and *nacβ1nacβ2* pollen under in vitro conditions. Both datasets were statistically compared by Student’s t-test, p-value of which is given. (D)—Box-and-whisker plot showing the median and quartiles of pollen tube length from three independent experiments germinating Col-0 wt pollen and *nacβ1nacβ2* pollen in vitro. Both datasets were statistically compared by Student’s t-test, p-value of which is given. (E)—Aniline blue staining of Col-0 wt pistils in vivo pollinated with Col-0 wt pollen. Scale bar represents 100 μm. (F)—Aniline blue staining of Col-0 wt pistils in vivo pollinated with *nacβ1nacβ2* pollen. Scale bar represents 100 μm. (G)—Aniline blue staining of *nacβ1nacβ2* pistils in vivo pollinated with Col-0 wt pollen. Scale bar represents 100 μm. (H)—Aniline blue staining of *nacβ1nacβ2* pistils in vivo pollinated with *nacβ1nacβ2* pollen. Scale bar represents 100 μm. (I)—The blue dot assay showing the fertilization efficiency of Col-0 wt pistil by labelling the pollen tubes with proLAT52–GUS. Every fertilization event is represented by a blue dot. Scale bar represents 100 μm. (J)—The blue dot assay showing the fertilization efficiency of *nacβ1nacβ2* pistil by labelling the pollen tubes with proLAT52–GUS. Scale bar represents 100 μm. (K)—Box-and-whisker plot showing the median and quartiles of fertilization rate of *nacβ1nacβ2* and Col-0 wt pistils in three independent experiments. The embryo sacs marked in blue were considered as fertilized whereas the unstained ones were regarded as unfertilized. Both datasets were statistically compared by Student’s t-test, p-value of which is given.

### 2.3. *nacβ1* and *nacβ2* Alleles Show Partly Reduced Transmission Efficiency

So far, the phenotype of the *nacβ1nacβ2* plants was described together with the nature of phenotypic traits in male gametophyte development, pollen tube growth and targeting efficiency. However, it was also necessary to test the transmission efficiency of the *nacβ* mutant alleles to the next generation and compare it with the wild type *NACβ* alleles. To achieve this, a set of two self-crosses, and eight out-crosses was performed. Since there are two *NACβ* genes in *Arabidopsis thaliana* genome, the

genotypic background of the second gene was considered. At first, plants heterozygous in one *NACβ* gene in the mutant background of the other one (*nacβ1/+;nacβ2*, and *nacβ1;nacβ2/+*, respectively) were let to self-pollinate. Their offspring was cultivated and conventionally genotyped by PCR with a set of insert-specific and gene-specific primer pairs (Table 1). The number of heterozygous plants (*nacβ1/+;nacβ2* or *nacβ1;nacβ2/+*, respectively) was nearly the same in both self-crosses as expected by the Mendelian laws of inheritance (13% more plants in *NACβ1*, and 10% more in *NACβ2*). On the contrary, the number of homozygous wild type plants (*NACβ1nacβ2* or *nacβ1NACβ2*, respectively) was in both cases higher than expected (11% more in *NACβ1*, and 29% more in *NACβ2*), whereas the number of homozygous mutants (*nacβ1nacβ2* or *nacβ1nacβ2*, respectively) was lower than the Mendelian one (46% less in *NACβ1*, and 48% less in *NACβ2*). Based on these self-crosses, it was deduced that the mutant allele of both studied genes showed a reduced transmission to the next generation compared to the wild type allele.

**Table 1.** The transmission efficiency of the *nacβ1* and *nacβ2* alleles in self-crosses of the heterozygous plants on the mutant background of the other *NACβ* allele (*nacβ1/+;nacβ2* or *nacβ1;nacβ2/+*, respectively).  $\chi^2$  column shows the p-value of the  $\chi^2$  test.

Parent	Selfing Offspring	Expected Mendelian Ratio	Observed Ratio	Change in Transmission Efficiency	$\chi^2$
<i>nacβ1/+;nacβ2</i> <i>NACβ1</i> transmission (N = 119)	<i>NACβ1nacβ2</i>	29.75	33	+11%	$7.5 \times 10^{-2}$
	<i>nacβ1/+;nacβ2</i>	59.50	67	+13%	
	<i>nacβ1nacβ2</i>	29.75	19	-46%	
<i>nacβ1;nacβ2/+</i> <i>NACβ2</i> transmission (N = 115)	<i>nacβ1NACβ2</i>	28.75	37	+29%	$8.8 \times 10^{-3}$
	<i>nacβ1;nacβ2/+</i>	57.50	63	+10%	
	<i>nacβ1nacβ2</i>	28.75	15	-48%	

To further elucidate the transmission efficiency of the mutant alleles via male and female gametophyte, two rounds of reciprocal out-crosses were carried out, each of which was represented by four crosses, i.e., eight crosses in total. The first round studied the transmission efficiency of one *NACβ* gene on the wild type background of the other *NACβ* gene, whereas the second round considered the transmission efficiency of one *NACβ* gene on the mutant background of the other *NACβ* gene. (i) In the first round, the heterozygous plants in one *NACβ* gene on the mutant background of the other one (*nacβ1/+;nacβ2*, and *nacβ1;nacβ2/+*, respectively) were crossed with Col-0 wt plants (genotype *NACβ1NACβ2* in both cases). These crosses were performed in two ways, the first time the pistils of *nacβ1/+;nacβ2*, or *nacβ1;nacβ2/+*, respectively, were pollinated with wild type pollen, whereas the other time their pollen was used for the pollination of wild type pistils. Such setup of crosses enabled to test the transmission efficiency of the mutant allele via both female and male gametophytes (Table 2). The transmission efficiency of the mutant and the wild type allele was calculated from the occurrence of homozygous mutant and heterozygous plants in the offspring of these crosses. Theoretically, they should occur with the same frequency. The *nacβ1* mutant allele showed higher transmission efficiency compared to the wild type allele via both female (+20%), and male gametophyte (+52%). On the contrary, the *nacβ2* mutant allele was transmitted with a slightly reduced efficiency than wild type alleles were, again both via female (-15%) and male gametophyte (-22%). (ii) The second round of out-crosses was performed on the mutant background of the non-tested *NACβ* gene. The tested gene was heterozygous in one crossed plant (*nacβ1/+;nacβ2*, and *nacβ1;nacβ2/+*, respectively) whereas in the second plant was represented by two wild type alleles (*NACβ1nacβ2*, and *nacβ1NACβ2*, respectively). Again, two crosses were performed for each *NACβ* locus, once looking for transmission efficiency via male and the second time via female gametophyte (Table 3). The mutant allele of *NACβ1* showed a comparable transmission efficiency as the wild type allele did both via female (-5%) and male gametophyte (-1%). Similarly, the mutant allele of *NACβ2* was transmitted with a slightly worse efficiency than the wild type alleles were both via female and male gametophyte (both times 17% less).

To sum up, the *nacβ1* mutant allele showed a comparable male and female transmission efficiency on the mutant background of the other *NACβ* gene, and even better male and female transmission efficiency on the wild type background. On the contrary, the *nacβ2* mutant alleles showed slightly less efficient transmission by both male and female gametophyte on both backgrounds of the other *NACβ* gene (mutant and wild type).

**Table 2.** The transmission efficiency of the *nacβ1* and *nacβ2* alleles in out-crosses of the heterozygous plants (*nacβ1/+;nacβ2* or *nacβ1;nacβ2/+*, respectively) with Col-0 (wt) via both female and male gametophyte.  $\chi^2$  column shows the p-value of the  $\chi^2$  test. ♀ means female, whereas ♂ means male.

Cross	Mendelian Rate of Mutant Allele	Observed Rate of Mutant Allele	Mutant Allele Transmission Efficiency—Wild Type Background	$\chi^2$
<i>nacβ1/+;nacβ2</i> ♀ × Col-0 (wt) ♂ (N = 88)	44	53	<i>nacβ1</i> ♀ +20%	$5.5 \times 10^{-2}$
<i>nacβ1;nacβ2/+</i> ♀ × Col-0 (wt) ♂ (N = 118)	59	50	<i>nacβ2</i> ♀ -15%	$9.8 \times 10^{-2}$
Col-0 (wt) ♀ × <i>nacβ1/+;nacβ2</i> ♂ (N = 128)	64	97	<i>nacβ1</i> ♂ +52%	$5.4 \times 10^{-9}$
Col-0 (wt) ♀ × <i>nacβ1;nacβ2/+</i> ♂ (N = 133)	66.5	52	<i>nacβ2</i> ♂ -22%	$1.2 \times 10^{-2}$

**Table 3.** The transmission efficiency of the *nacβ1* and *nacβ2* alleles in out-crosses of the heterozygous plants (*nacβ1/+;nacβ2* or *nacβ1;nacβ2/+*, respectively) in the mutant background of the other *NACβ* gene via both female and male gametophyte.  $\chi^2$  column shows the p-value of the  $\chi^2$  test. ♀ means female, whereas ♂ means male.

Cross	Mendelian Rate of Mutant Allele	Observed Rate of Mutant Allele	Mutant allele Transmission Efficiency—Mutant Background	$\chi^2$
<i>nacβ1/+;nacβ2</i> ♀ × <i>NACβ1nacβ2</i> ♂ (N = 114)	57	54	<i>nacβ1</i> ♀ -5%	$5.7 \times 10^{-1}$
<i>nacβ1;nacβ2/+</i> ♀ × <i>nacβ1NACβ2</i> ♂ (N = 143)	71.5	59	<i>nacβ2</i> ♀ -17%	$3.7 \times 10^{-2}$
<i>NACβ1nacβ2</i> ♀ × <i>nacβ1/+;nacβ2</i> ♂ (N = 142)	71	70	<i>nacβ1</i> ♂ -1%	$8.7 \times 10^{-1}$
<i>nacβ1NACβ2</i> ♀ × <i>nacβ1;nacβ2/+</i> ♂ (N = 140)	70	58	<i>nacβ2</i> ♂ -17%	$4.3 \times 10^{-2}$

Taken together, the self-crosses and out-crosses showed a reduced transmission efficiency of *nacβ2* mutant allele compared to the wild type regardless on the background of the *NACβ1*. On the other hand, the *nacβ1* mutant allele showed a reduced transmission efficiency in the self-cross, a comparable transmission efficiency in the outcrosses on the mutant background of the *NACβ2* gene whereas its transmission efficiency was even higher in the outcrosses on the wild type background of *NACβ2*.

#### 2.4. The *nacβ1nacβ2* Mutant Phenotype Was Complemented by any Functional *NACβ* Allele

After testing the transmission efficiency of the mutant alleles, it was necessary to prove that the *nacβ1nacβ2* phenotype was caused by the insertions in both *NACβ*-coding genes. The first experiment tested whether the mutant phenotype would be reverted after a new functional copy of a single *NACβ* gene (containing native promoter and coding genomic sequence tagged with GFP) was inserted to the *nacβ1nacβ2* plants. However, due to the reduced seed set of the *nacβ1nacβ2* plants, we were not able to recover enough transformants, so the experiment was repeated by transforming plants having one original functional copy of one *NACβ* gene, i.e., *nacβ1/+;nacβ2*, and *nacβ1;nacβ2/+*. Every time, the construct carried the same *NACβ* gene, alleles of which were represented as homozygous mutant in the transformed plants, so *nacβ1;nacβ2/+* plants were transformed with proNACβ1::NACβ1-GFP

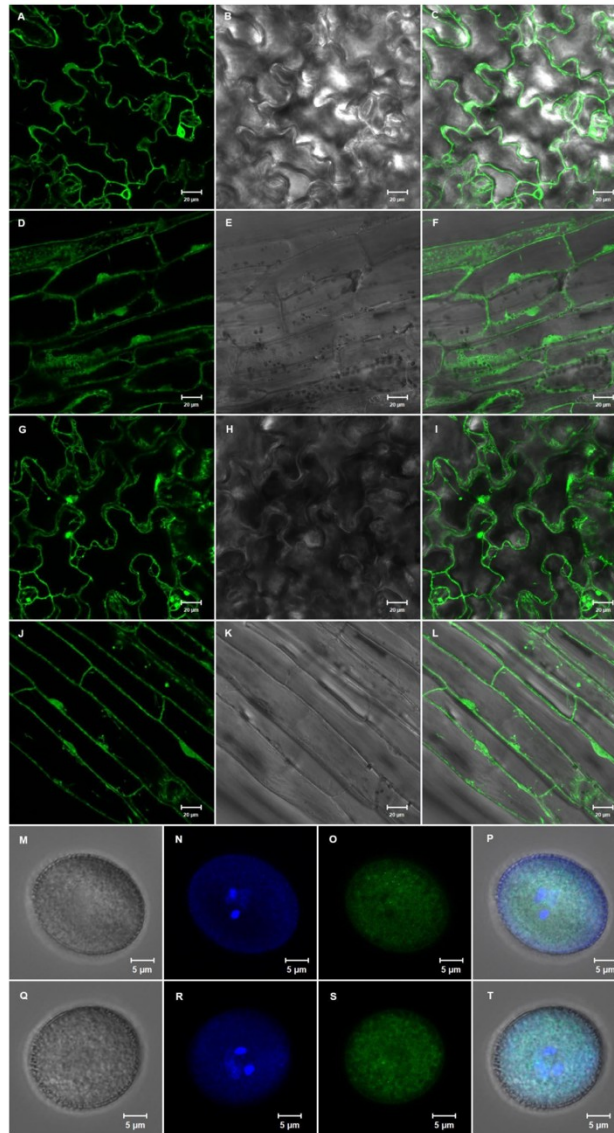
construct, whereas the *nacβ1/+;nacβ2* plants were transformed with proNACβ2::NACβ2–GFP construct. Several *nacβ1nacβ2* plants carrying the construct (either with NACβ1, or NACβ2) segregated in the next generation, and one novel functional allele of any NACβ gene rescued the phenotypic defects of the *nacβ1nacβ2* mutants, which became undistinguishable from the Col-0 wt plants (Figure 1I–L).

In parallel, *nacβ1nacβ2* plants were crossed with Col-0 wt, again reciprocally. The resulting heterozygous plants in both NACβ loci (*nacβ1/+;nacβ2/+*) did not show the mutant phenotype regardless of the cross direction and looked like the Col-0 wt plants. The complementation analysis together with the backcross verified that the observed *nacβ1nacβ2* phenotype was caused by the insertions in both NACβ loci and that its phenotype can be rescued by presence of any single functional NACβ allele.

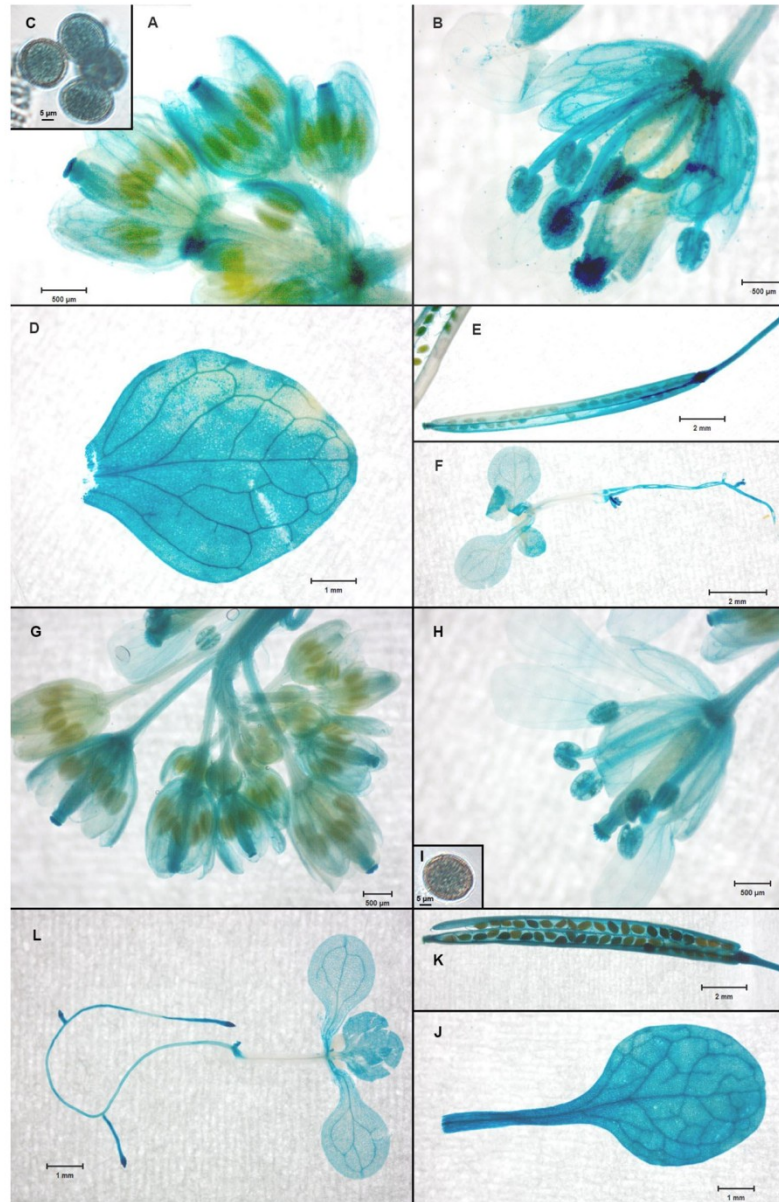
Then, the opposite effect of the NACβ expression was studied, the effect of overexpression. Each NACβ gene represented by its genomic coding sequence was expressed under the 35S cauliflower mosaic virus (CaMV) promoter, which shows a strong expression in most *A. thaliana* sporophyte tissues [38]. The 35S-driven overexpression of each NACβ separately as well as both at the same time did not lead to any phenotypic changes and the transformed plants looked like the conventional Col-0 wt. Then the overexpression was also performed in male gametophyte (under the LAT52 promoter [39]) and female gametophyte (under DD33 promoter [40]). Again, no phenotypic defects were observed.

#### 2.5. The NACβ Proteins Localized into Cytoplasm and Nuclei and Their Promoters were Active in Multiple Organs

To determine the subcellular localization, the same constructs were used as for the complementation (proNACβ1::NACβ1–GFP, and proNACβ2::NACβ2–GFP). The 10-day *A. thaliana* seedlings, which were transformed by *Agrobacterium tumefaciens*, were selected on MS/2 plates with proper antibiotics. The acquired selectants were observed under the confocal microscope. The genes coding for both NACβ genes (NACβ1, and NACβ2) showed the same subcellular localization, since they both were localized inside the cells of the hypocotyls and the leaves to cytoplasm, and the nuclei (Figure 3A–L). In mature pollen, NACβ1 and NACβ2 proteins were observed inside granular structures in the vegetative cell cytoplasm but not inside the vegetative nuclei (Figure 3M–T). Then, the promoter analysis was performed by glucuronidase marker gene driven directly by each NACβ promoter (proNACβ1–GUS, and proNACβ2–GUS). The 10-day old seedlings and several organs (leaves, cauline leaves, flowers, pollen grains, and siliques together with seeds) from older plants were stained for the presence of glucuronidase activity and observed under dissection microscope. Both NACβ genes showed very similar pattern of promoter activity (Figure 4), so none of them showed any organ specificity compared to the other gene, and so was in accordance with the published transcriptomics data [41,42]. Both genes were active in the whole seedlings except for their hypocotyls, and then in leaves, cauline leaves, flower buds, flowers, pollen grains, siliques, and seeds. Inside flowers, their activity was most prevalent in pistils, and anthers. Taken together, the genes showed a house-keeping nature of their promoters since they were present nearly in the whole plant.



**Figure 3.** The subcellular localization of NAC $\beta$ 1 and NAC $\beta$ 2 expressed in ten-day seedlings under their own promoters, and the expression of NAC $\beta$ 1 and NAC $\beta$ 2 under the LAT52 promoter in vegetative cells of the pollen grains observed under the confocal microscope. (A–F)—proNAC $\beta$ 1::NAC $\beta$ 1-GFP, scale bars represent 20  $\mu$ m. A–C—Leaves. D–F—Hypocotyl. (G–L)—proNAC $\beta$ 2::NAC $\beta$ 2-GFP, scale bars represent 20  $\mu$ m. G–I—Leaves. J–L—Hypocotyl. (M–P)—proLAT52::NAC $\beta$ 1-GFP in vegetative cells of the pollen grains, scale bars represent 5  $\mu$ m. (Q–T)—proLAT52::NAC $\beta$ 2-GFP in vegetative cells of the pollen grains, scale bars represent 5  $\mu$ m. (A,D,G,J,O,S)—GFP fluorescence. (B,E,H,K,M,Q)—Bright field. (N,R)—DAPI staining. (C,F,I,L,P,T)—All signals merged together.



**Figure 4.** The promoter analysis of NAC $\beta$ 1 and NAC $\beta$ 2 fused with glucuronidase, observed under the dissection microscope (all images except for (C), and (I)) and optical microscope (C,I). The construct carried NAC $\beta$ 1-GUS (A–F) or NAC $\beta$ 2-GUS (G–L). (A,G)—Inflorescence. (B,H)—Flower. (C,I)—Pollen grain. (D,J)—Leaf. (E,K)—Silique. (F,L)—Ten-day old seedling. Scale bars represent 5  $\mu$ m (C,I), 500  $\mu$ m (A,B,G,H), 1000  $\mu$ m (D,J,L), or 2000  $\mu$ m (E,F,K), respectively.

## 2.6. RNA Sequencing of the *nacβ1nacβ2* Flower Bud Transcriptome Revealed Genes Important for Stress Responses and Male Gametophyte Development

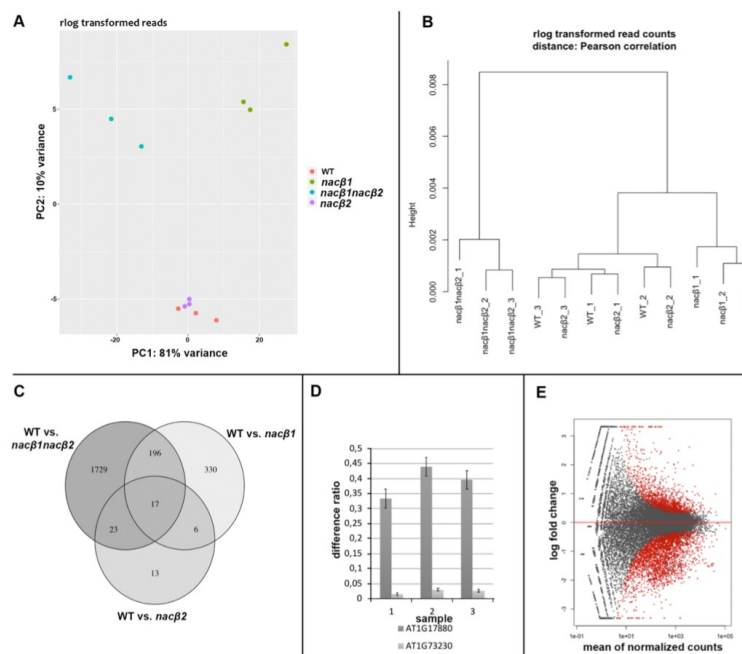
Since the phenotypic defects of the *nacβ1nacβ2* plants were most prevalent in the flowers and siliques, and the data from pollen cultivation *in vitro*, pollen cultivation *in vivo* and blue dot assay showed that the defects are likely in both gametophytes and/or sporophyte, flower buds (including the stages up to the stage 12 [43]) were selected as the studied organ to search for genes, transcription of which was influenced in the *nacβ1nacβ2* plants. To accomplish this task, two approaches were chosen, transcriptomics and proteomics. Moreover, to see whether the single phenotype-less *nacβ* mutants (*nacβ1* or *nacβ2*) bear any changes on the transcriptomic level, they were also included in the analysis. RNA genome-wide transcriptional analyses were performed on Illumina platform and yielded approximately 40.6 million paired-end 101 base pair (bp) reads per sample after quality (>Q20) control and technical sequences trimming. On average 94.3% of reads were uniquely mapped to *A. thaliana* reference genome (TAIR10).

Of these, differentially expressed genes (DEGs) with adjusted p-value < 0,05, and FoldChange  $\geq \pm 2$  were selected. Principal component analyses (PCA) with PC1 explaining 81% variance and PC2 10% as well as in Pearson correlation hierarchical clustering analysis, three clusters are revealed: *nacβ1nacβ2*, *nacβ1*, and *nacβ2* together with Col-0 wt (Figure 5A,B). The clustering corresponded to the observed phenotype and to the identified DEGs. There were identified in total 1965 DEGs in *nacβ1nacβ2* (Table S3), 59 DEGs in *nacβ2*, and 549 DEGs in *nacβ1* (Figure 5C). Moreover, only *NACβ1* was differentially expressed in the *nacβ1nacβ2* mutant leading to nearly total silencing but on the contrary, the expression of *NACβ2* was only lowered (Figure 5D). Out of the total 1965 DEGs in the *nacβ1nacβ2*, 363 were up-regulated and 1602 down-regulated (Figure 5E). To further analyze functional potential of the DEG list, the transcripts were classified according to biological processes, molecular function and cellular component using gene ontology (GO) enrichment analysis. There were successfully recognized 1919 DEGs by the software, which were used for the enrichment analyses. Of these, 91 (87 downregulated) DE male germ-line connected genes were included in five enriched biological processes GO terms, namely “pollen development”, “pollen sperm cell differentiation”, “pollen tube development”, “pollen tube growth”, and “regulation of pollen tube growth” (Table S3). 19 of these DEGs with 14 others were also included in the enriched “pollen tube” GO cellular component category. “Pectin catabolic process” and “cell wall modification” enriched GO biological process GO terms included 32 DEGs with 8 pectinesterases, 10 pectinesterase inhibitors and 6 pectine lyases. “Cell wall” was enriched in cellular component and “pectinase activity” in molecular function. “Protein complex oligomerization” enriched GO term included 12 small heat shock family genes. In the DEGs list, there were upregulated 18 genes encoding other heat shock proteins and 2 genes encoding heat shock factors. Among other enriched biological process GO terms, there were “regulation of pH” with 10 DE Cation/H(+) antiporters, “calcium-mediated signaling”, “terpene metabolism”, “response to hydrogen peroxide”, and “response to insect”. Visualization of GO enrichment analysis is showed in Figure 6A,B. The transcriptomics data were validated by measuring expression of 12 candidate genes by quantitative qRT-PCR. The quantities of these genes according to qRT-PCR corresponded to expression values obtained from next generation sequencing with correlation coefficient  $R^2 = 0,7299$  (Figure S3).

## 2.7. Proteomic Analysis of the *nacβ1nacβ2* Flower Buds Revealed Proteins with Role in Photosynthesis and Stress Responses

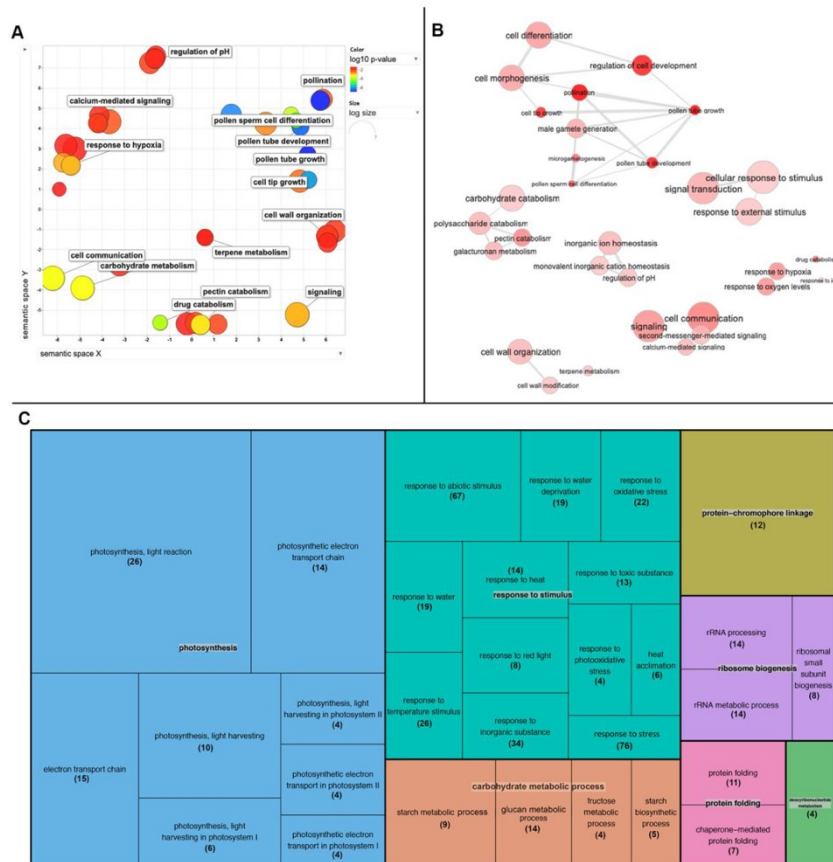
To acquire a closer insight into the regulatory networks influenced by the knock-down of both *NACβ* subunits, the proteomic analysis of the *nacβ1nacβ2* flower buds was performed and compared to the Col-0 wt plants. In total, there were 5499 protein groups reported that were identified according to at least two peptides in at least two out of three biological replicates in at least one of the samples and have measured intensity in at least two replicates in at least one sample. For the differentially expressed proteins, the following criteria were set: the fold change between the *nacβ1nacβ2* and the wt had to be  $\geq 2$  and the adjusted p-value < 0.05. From the total 5499 analyzed protein groups, 460 were differentially expressed in *nacβ1nacβ2*, with 290 downregulated and 170 upregulated ones (Table

S4). All proteins were mapped with GO enrichment analyses. 38 differentially expressed proteins (DEPs) connected to photosynthesis were downregulated in the *nacβ1nacβ2*. They were present in the enriched GO terms “photosynthesis”, “response to red light”, “photosystem electron transport chain in photosystem I”, “protein-chromophore linkage”, “photosynthesis, light harvesting in photosystem I”, and “photosynthesis, light harvesting in photosystem II” of biological processes. In cellular component category, chloroplast localization was represented by “chloroplast envelope”, “chloroplast thylakoid lumen”, “plastoglobuli”, and “photosystem I reaction center”. Biological process-enriched GO terms connected to stress response were represented by “water deficient”, “response to toxic substances”, “response to oxidative stress”, and “response to heat”; they included 44 primary downregulated proteins. “Response to heat” was represented by 7 upregulated chaperones which were also present in “protein-folding” enriched GO term together with 5 other proteins. 6 of these chaperones were connected directly to de-novo folding. Other interesting GO terms were represented by “rRNA processing”, “ribosome biogenesis”, and “ribosome small subunit assembly” and metabolism terms “starch metabolism”, and hexose metabolism”. Visualization of GO enrichment analysis is showed in Figure 6C.



**Figure 5.** Analyses of flower bud transcriptome. (A)—PCA based clustering of regularized log transformed read counts of four analyzed genotypes *nacβ1nacβ2*, *nacβ1*, *nacβ2*, and Col-0 wt explaining 81% of PC1 variance and 10% of PCA2 variance shows three distinctive clusters. (B)—Hierarchical clustering of regulated log transformed read counts of four genotypes *nacβ1nacβ2*, *nacβ1*, *nacβ2* and Col-0 wt based on Pearson correlation distance shows three distinctive clusters. (C)—Venn diagram showing comparison of DEG numbers in *nacβ1nacβ2*, *nacβ1*, *nacβ2* and Col-0 wt. (D)—Relative expression of *NACβ1* (At1g73230) and *NACβ2* (At1g17880) in *nacβ1nacβ2* when compared to expression in Col-0 wt showing knock-out of *NACβ1* and knock-down of *NACβ2*. (E)—MA plot of gene expression in *nacβ1nacβ2* compared to Col-0 wt. Genes with adjusted *p*-value < 0,05 are shown in red.





**Figure 6.** Functional analysis of DEGs and DEPs in *nacβ1nacβ2*. (A)—Enrichment analyses of gene ontology (GO) biological process scatter plot for 1909 DEGs. Names of representative GO terms that were present in the clusters are shown. The bubbles without any name represent other relative GO terms. The various colors indicate  $\log_{10}$  FDR value of GO enrichment analysis for respective GO term, whereas the distance indicates semantic similarity of the categories. Bigger size of bubbles indicates more parent GO terms. (B)—Interactive graph of GO biological process terms of 1909 DEGs. GO terms are connected according to their relationship. Bigger size of bubbles indicates more parent GO terms. Brighter color indicates lower  $\log_{10}$  FDR value of GO enrichment analysis for respective GO term. (C)—Enrichment analyses GO biological process tree map constructed with REVIGO for 399 differentially expressed proteins (DEPs). Each square represents a GO term. The same color connects GO terms that belong to the same parent category. The name of parent category is highlighted in bold. The size of squares is proportional to the FDR value for that term enrichment analysis. One gene can be present in more categories. The maps were constructed with ReviGO [44].

### 2.8. The Germination Efficiency of *nacβ1nacβ2* Seeds Was Lower Under the Salt and Osmotic Stresses

Since the transcriptomic and proteomic analyses revealed several regulated genes connected with stress responses (e.g., chaperones), sensitivity of *nacβ1nacβ2* seed germination to salt and osmotic stress compared to the Col-0 wt was tested. The sodium chloride (NaCl) delayed the germination process

and reduced the number of germinated *nacβ1nacβ2* mutant seeds already at 50 mM concentration (Figure S4A). The *nacβ1nacβ2* germination was more influenced by 100 mM and 125 mM NaCl since the total number of germinated seeds gradually decreased with an increasing NaCl concentration. These concentrations (50 mM, 100 mM, and 125 mM) affected the Col-0 wt seeds rather by delaying germination since their germination rate remained nearly unaffected after the last scoring in 13 days (Figure S4C,E). On the contrary, 150 mM NaCl significantly reduced the total number of germinated seeds of both *nacβ1nacβ2* and wt (Figure S4D), representing probably a too high concentration of NaCl for the seeds to survive.

The influence of mannitol on the germination of *nacβ1nacβ2* seeds was similar to the influence of NaCl. Under 250 mM and 300 mM concentrations, the *nacβ1nacβ2* germination was delayed and the number of germinated seeds was slightly reduced, whereas Col-0 wt seeds were rather affected by a slightly delayed germination (Figure S4B). On the other hand, the influence of 350 mM mannitol was comparable to the effect of 150 mM NaCl since the germination of both *nacβ1nacβ2* and wt was significantly slowed down and the number of germinated seeds reached only around 60% of the control if observed 13 days after seed sowing.

### 3. Discussion

#### 3.1. Phenotype Analysis of *nacβ1nacβ2*

The *nacβ1nacβ2* mutant showed a notable phenotype: delayed development, lower amount of chlorophyll in the leaves, unusual number of flower organs, and shorter siliques with a reduced seed set. Similarly, in *Nicotiana benthamiana* [24] and wheat (*Triticum aestivum* [25]), the NACβ downregulation caused an abnormal leaf morphology together with a lower chlorophyll content (and thus leaf yellowing) whereas the size of the whole plant remained the same. Our experiments testing the behavior of the male and female gametophytes revealed that the *nacβ1nacβ2* plants showed likely defects in sporophyte tissues of the flowers and/or defects in both gametophytes as the germination rate of pollen tubes was reduced together with the velocity of the growth since they reached shorter distance in the same time as the wild type did. Furthermore, the targeting efficiency was diminished also on the female side since wild type pollen tubes were less efficiently attracted by the *nacβ1nacβ2* ovules. The targeting efficiency was the worst in case the *nacβ1nacβ2* ovules were targeted by *nacβ1nacβ2* pollen tubes.

The NACβ genes are both quite strongly expressed according to the transcriptomic datasets presented in eFP browser [45], and according to the strong GUS expression driven by NACβ promoters in nearly all organs throughout plant development (Figure 4). Although the observed phenotype was quite clear, one would probably expect a more severe phenotype if the mutants represent a knock-out/down of such strongly expressed genes with a “housekeeping” function in translation. Moreover, in several model animals (for instance mouse, nematodes and drosophila), the NACβ knock out caused embryonic lethality [14–16]. Thus, it is likely that the NACβ function is mimicked in case of its knock-out to some extent by other chaperones since their transcripts were more abundant in the *nacβ1nacβ2* transcriptome compared to the wild type (Table S3). This hypothesis was further supported by yeast experiments, in which the loss of the whole NAC complex did not affect the yeast growth under the optimal conditions [4] whereas the yeast cells with a knocked out NAC complex together with two heat shock proteins 70 (Hsp70) chaperones SSB (stress 70 B) were still viable but their growth was notably slowed down [13]. Thus, Hsp70 chaperones in yeast can to some extent complement NAC’s function. It should be also mentioned that the survival and limited fertility of the studied *nacβ1nacβ2* plants were likely enabled by the fact that one NACβ gene (*NACβ2*) was knocked down rather than knocked out, and the remnants of its transcript could carry some basic function in the *nacβ1nacβ2* plants. On the contrary, the second NACβ gene (*NACβ1*) was proven to be knocked out completely. Last but not least, it is worth mentioning that the knock out of some house-keeping genes showed rather a tissue-specific phenotype than a severe phenotype affecting dramatically the

whole plant [46]. The different nature of both alleles could also explain the different transmission efficiency of the *NACβ1* and *NACβ2* mutant alleles since the knocked-down allele likely still carries out its regulatory function. Thus, the knock-down allele of *NACβ2* seemed to be more harmful than the knock-out allele of *NACβ1*. The higher transmission efficiency of the *NACβ1* mutant allele than the one of the wild type allele was likely caused by the wild type background of the *NACβ2* gene, which could carry out the *NACβ1* function. Moreover, it is likely that both *NACβ* genes are likely regulated together.

The flower phenotype tempted us to speculate whether meristem organizing, or flower development genes had a changed abundance in the *nacβ1nacβ2* transcriptome compared to the Col-0 wt plants. Meristem development disruption can be exhibited in several aspects, one is organ identity aberration, which was disrupted for example in *apetala 3 (ap3)*, *pistillata (pi)*, or *agamous (ag)* mutants [47], whereas the *clavata3 (clv3)*, *clavata1 (clv1)*, and *wiggle (wig)* mutants possessed a changed meristem size and, consequently, aberrations in organ numbers [48]. Another option can be connected to the duration of meristematic maturation. Prolonged meristematic activity under certain conditions led to larger meristem and then also to changes in organ number. This phenomenon was recently observed in close *A. thaliana* relative *Cardamine hirsuta* [49]. However, despite the observed phenotype, the abundance of the meristem organizing, or flower development genes was not significantly changed in the *nacβ1nacβ2* flower bud transcriptome. Thus, we speculate that the stress-like conditions that were shown by the transcriptomics data, where many heat shock proteins were upregulated in the *nacβ1nacβ2*, could lead to a prolonged meristematic activity and thus to disrupted number of floral organs and in general could alter protein homeostasis.

### 3.2. *NACβ* Paralogues

Interestingly, the above-mentioned phenotype was observed exclusively in the *nacβ1nacβ2* plants since the presence of any single functional *NACβ* allele (regardless whether *NACβ1* or *NACβ2*) suppressed the mutant phenotype. Such an observation was done independently three times, first when crossing the homozygous plants of two T-DNA insertion lines (SALK\_043673 bearing insert in *NACβ1*, and GK368-H02 with insert in *NACβ2*), the phenotype did not appear until double homozygous mutants (*nacβ1nacβ2*) were acquired, so plants with one functional allele (either *nacβ1/+;nacβ2*, or *nacβ1;nacβ2/+*) did not show any phenotypic changes. Second, the *nacβ1nacβ2* plants (with the above phenotype) were backcrossed with Col-0 wt (genotype *NACβ1NACβ2*), giving rise to double heterozygous plants (*nacβ1/+;nacβ2/+*), again with no phenotypic defects. Third, the *nacβ1nacβ2* phenotype was reverted to the wild type phenotype by creating the transgenic plants carrying one functional copy of any *NACβ* gene.

Similar behavior to the alleles of the *NACβ* genes was observed for instance in case of *SEPALLATA (SEP)* since only its triple mutant *sep1/sep2/sep3* showed a severe flower phenotype as its flowers carried sepal-like structures instead of the proper flower organs, whereas in contrast, any of its single mutants showed only a subtle phenotype [50]. Such a behavior was explained by the redundancy of these three *SEPALLATA* paralogues. In case of *NACβ*, the phenotype was suppressed, and the function was restored by single functional copy of any *NACβ* genes, indicating that both genes were functionally redundant and very similar to each other. Due to the high level of sequence similarity between these two proteins (reaching 89% similarity on the amino acid level where most differences were in the C-terminal part of the protein), it is likely that both *NACβ* genes show a similar function. Furthermore, both *NACβ* genes showed a similar subcellular localization together with promoter activity and none showed any spatial or temporal specificity. Unlike two yeast *NACβ* genes, which showed a distinct function and a different efficiency for creating the dimer complexes with the single yeast *NACα* gene [3], *A. thaliana* *NACβ* genes seem to be much closer to each other and rather share the same function. However, in the future on-going experiments, it remains to be tested which of the *NACα*, and *NACβ* homologues appear in the NAC dimers and if there are any binding preferences between them at all.

The  $\beta$ -subunit was first reported as a part of NAC heterodimer together with  $\alpha$ -subunit [7] but both subunits were also reported to play different roles outside the heterodimer by influencing gene

expression [18,51,52]. The mutual regulation of  $\alpha$ - and  $\beta$ -subunits was also proposed, claiming that NAC $\alpha$  transcriptional coactivator function is inhibited by the presence of NAC $\beta$  to promote the heterodimer formation [53]. Nevertheless, the RNAseq data from *nac $\beta$ 1nac $\beta$ 2* flower buds did not show any changes in the abundance of NAC $\alpha$  subunits in the *nac $\beta$ 1nac $\beta$ 2* plants compared to the wt, therefore it can be concluded that on the transcriptomic level in the flower buds, NAC $\beta$  did not influence the expression of NAC $\alpha$ . On the other hand, on the proteomic level, there was shown the downregulation of NAC $\alpha$ 3 (At5g13850), NAC $\alpha$ 4 (At4g10480), and NAC $\alpha$ 5 (At1g33040) in *nac $\beta$ 1nac $\beta$ 2* making likely the expression regulation on the level of translation.

### 3.3. Subcellular Localization of NAC $\beta$ Proteins

Both paralogues encoding NAC $\beta$  proteins were localized into both cytoplasm and nucleus in *Arabidopsis thaliana* seedlings (Figure 3A–L) regardless of whether N-terminal or C-terminal GFP fusion was employed. There were not any differences in the localization of both homologues, thus the localization also supported the likelihood of functional similarity of these two NAC $\beta$  genes. Our localization data represented by in frame fusion of NAC $\beta$  coding sequences with green fluorescent protein, were consistent with the subcellular localization of both *Arabidopsis thaliana* NAC $\beta$ s [21], and wheat (*Triticum aestivum*) NAC $\beta$  localization in *A. thaliana* protoplasts [25]. The protein localization is also in accordance with several software predictions, where the NAC $\beta$  proteins were predicted to be localized either to nucleus or cytoplasm. For instance, NucPred [54] predicted NLS in both NAC $\beta$ s, represented in both proteins as amino acids 22–25 (RRKK) with the score 0.40 in NAC $\beta$ 1 or 0.44 in NAC $\beta$ 2, respectively. However, these NLS predictions were not strong, which agreed with the dual targeting (nucleus, and cytoplasm). This is also in accordance with the cytoplasmic localization of both NAC $\beta$ s in the vegetative cell of the pollen grain (Figure 3M–T), the NLS being likely too weak for these tissues. The combined localization also accords with the NAC function, either as a heterodimeric complex binding the nascent polypeptide or each subunit individually as a transcription factor or activator, the former localized likely to cytoplasm, whereas the latter function requires nuclear localization, reviewed by Kogan and Gvozdev (2014), and Rospert et al. (2002) [1,3]. In the cytoplasm, the NAC $\beta$ 1 (At1g73230) was reported to bind to the ribosomes, which was mediated by phosphorylation carried out by OST1 protein [21].

### 3.4. Flower Bud Transcriptome and Proteome

The presented transcriptomic and proteomic analyses showed the genes that were influenced in their abundance in the floral buds of *nac $\beta$ 1nac $\beta$ 2* mutants compared to the Col-0 (wt) plants. Next generation sequencing of the flower bud transcriptome followed by differential expression (DE) analysis showed 1965 differentially expressed genes (DEGs) between the *nac $\beta$ 1nac $\beta$ 2* mutant and Col-0 wt plants, whereas there were only 59 and 549 DEGs in single NAC $\beta$  mutants *nac $\beta$ 2* and *nac $\beta$ 1*, respectively. These results were in accordance with the observed phenotype since *nac $\beta$ 1nac $\beta$ 2* mutants showed a notable phenotype (delayed development, lower amount of chlorophyll in leaves, flowers with a different number of flower organs, shorter siliques, reduced seed set) whereas both single mutants were undistinguishable from the Col-0 wt plants (see in more detail above). Thus, it was unlikely that plants of the same look as the wild types will have many transcripts of a different abundance.

From GO term enrichment analysis, three groups of genes emerged. Firstly, there were 87 downregulated genes connected to male germ-line development from sperm-cell differentiation to pollen tube development and growth. Lower germination rate, slower growth and lower fertilization rate were also observed in pollen tubes of the *nac $\beta$ 1nac $\beta$ 2* plants. The presence of transcripts from different pollen developmental stages was probably caused by the nature of collected flower bud tissue which included flowers up to stage 12 [43]. From the stage 10 do 12, pollen mitosis I and II take place, so there were likely present both bicellular and tricellular pollen in the acquired samples. Also, as described in other plant species, mRNAs in developing pollen are stored in translationally inactive complexes [55,56]. Pollen activation is then followed by massive translation of these pre-prepared

mRNAs, enabling, among others, rapid growth of pollen tube [57]. This could explain high amount of pollen tube regulatory genes in the analyzed RNA. The list consisted of genes involved in various molecular processes, many of which were connected to reproduction defects upon silencing. Among these playing role in pollen tube development and growth, there were heterodimeric MADS box agamous-like proteins AGL66 (At1g77980) and AGL104 (At1g22130). Double mutant *agl66 agl104* exhibits delayed pollen grain germination, aberrant pollen tube growth, reduced pollen viability and fertility [58]. Also, RALF4 (At1g28270) and RALF19 (At2g3375), peptide ligands responsible for pollen tube integrity, were present [59]. Four pectin catabolism genes, two pectin esterase inhibitors PME11 (At1g48020) and PME12 (At3g17220) and two pectinesterases PME1 (At1g69940) and PME5 (At1g10770) were downregulated. Cell wall modification proteins are especially important for pollen tube cell wall rigidity and dynamics [60]. In sperm-cell development category, sperm cell specific MYB transcription factor DUO1 (At3g60460) and 13 genes encoding DUO1 targets were downregulated. DUO1 is responsible for division of germ cell into two sperms cells via regulation of mitotic cyclin CYCB1 accumulation during G2 phase [61]. Among its targets, there was downregulated DMP9/DAU2 (At5g39650), which is responsible for double fertilization and its silencing leads to seed abortion [62]. Last but not least, three zinc finger transcription factors DAZ1 (At2g17180), DAZ2 (At4g35280), DAZ3 (At4g35700) or vacuolar aquaporin TIP5;1 (At3g35700) important for pollen tube growth were included. In summary, these 87 downregulated genes could be collectively responsible for the observed pollen phenotype.

The second group consisted of 18 upregulated heat shock proteins. There were three HSP70 family proteins HSP70-4 (AT3G12580), BiP3 (HSP70-8, AT2G32120), and HSP70-T (AT3G32120) together with HSP90.1 (At5g52640), and HSP101 (At1g74310). HSP70 forms a heterodimer with HSP40 to solubilize protein aggregates and has its role in stress tolerance. BiP3 together with BiP1 and BiP2 were expressed exclusively under stress conditions. Decrease in activity of these chaperones led to a worse pollen tube growth [63]. There were also 3 small heat shock proteins (HSP18.2 AT5G59720, HSP17.4 AT3G46230, and HSP17.6A AT5G12030) and 9 HSP20-like family proteins (AT1G53540, ATHSP22.0 AT4G10250, AT2G29500, AT5G51440, AT1G07400, AT1G59860, AT5G37670, AT1G54050, and AT4G21870), which play their role in both solubilizing protein aggregates, and in plant heat stress and immunity. They are usually less abundant and with a spatially- or temporally-specific expression throughout plant development, such as embryonic development and seed maturation [64,65]. In the DEG list, there were also 18 downregulated LEA proteins, which act as chaperones during water deficiency and as such promote dehydration stress tolerance.

In the GO enrichment analyses of the proteome, 38 photosynthetic downregulated proteins were identified in the *nacβ1nacβ2* flower buds. These proteins were mainly connected to light harvesting on photosystem I or II, response to red light or electron transport chain. There were 6 chlorophyll a-b binding proteins (LHCB2.2 AT2G05070, LHCB3 AT5G54270, LHCB4.1 AT5G01530, LHCB4.2 AT3G08940, and *lhb1B2* AT2G34420). They represent light harvesting complexes in the thylakoid membrane. Silencing of LHCB led to the modulation of ROS levels and decrease in sensibility to ABA during stomatal opening, which consequently led to lower stress tolerance [66]. Moreover, ABA1 (AT5G67030), zeaxanthin epoxidase responsible for the first step of ABA biosynthesis, was downregulated. Seven other proteins were parts of photosystem I (LHCA6 AT1G19150, *psaC* ATCG01060, *psaA* ATCG00350, *psaD1* AT4G02770, *psaE1* AT4G28750, *psaE2* AT2G20260, and *psaN* AT5G64040). *psaE1* mutants exhibit pale green leaves [67]. Also, photosystem repair and stress tolerance proteins were present as well as protein D1 (ATCG00020) and protein D2 (ATCG00270) of photosystem II. Taken together, downregulation of photosynthetic proteins supports lower chlorophyll content phenotype observed in the mutant. It may also lead to the delayed development of *nacβ1nacβ2* plants. Besides, it may suggest that beside ER and potential mitochondria targeting of nascent polypeptide chains, NAC or its downstream effectors, are also responsible for effective chloroplast targeting.

Other GO terms enriched in the *nacβ1nacβ2* proteome were connected to stress response. As was previously reported in plants, NACβ was observed to be important for stress tolerance, usually leading

to lower tolerance upon silencing and enhanced tolerance if overexpressed [28,29]. Furthermore, the germination rate of the *nacβ1nacβ2* seeds was slowed down under the lower concentrations of NaCl or mannitol compared to the Col-0 wt (Figure S4). As observed in the flower bud transcriptome, there was also upregulation of chaperones in the *nacβ1nacβ2* proteome. Six of the chaperones were related to de novo protein folding including two chaperonins connected with chloroplast import—CPN60B3 (AT5G56500), and CPN10-1 (AT3G60210)—and three HSP70 family chaperones HSP70-2 (AT5G02490), HSP70-4, and HSP70-8. HSP70-8 and HSP70-4 were also upregulated as transcripts. In mammals, HSP70L1 was shown to be NAC partner in protein folding [68], whereas in yeast, the mutation of HSP70-type chaperones *ssb1* and *ssb2* was complemented by NAC [5]. Also, HSP70/40 complexes were important for cytoplasmatic de novo protein folding [69]. HSP40 family *dnaJ2* (AT5G22060) was also upregulated in the mutant. However, heat stress proteins were downregulated—ZEP (At5g67030) important for ABA synthesis, xanthophyll cycle and photooxidative stress, and Y3IP1 (AT5G44650) responsible for photosystem I assembly and photooxidative stress response. Other stress-related proteins were connected to water deficit, response to toxic substances and oxidative stress, majority of which were downregulated in the mutant. For instance, 6 LEA proteins were among the downregulated in both analyses.

To summarize transcriptomics and proteomics data together, there were shared 15 upregulated genes/proteins and 98 downregulated ones (Table S5). Despite of the low overlap in the GO terms by themselves, there were observed several similar trends between the transcriptome and the proteome, for instance upregulation of chaperones or down-regulation of stress-related proteins. The absence of DE proteins connected with pollen development (unlike in the transcriptome analyses where there were detected several of them) could be explained by the mRNA storage during pollen maturation and simply absence of yet not-translated proteins, since only 13 proteins encoded by the DEGs connected to pollen development were discovered in the proteome analysis. Seven of these proteins passed the statistical thresholds and were also downregulated in the mutant. The analysis also proposes the role of NACs in regulation of photosynthetic apparatus.

#### 4. Materials and Methods

##### 4.1. Plant Cultivation

The seeds of Columbia-0 (wild type), and T-DNA insertion lines SALK\_043673 (insert in At1g73230, *NACβ1*), and GK-368H02 (insert in At1g17880, *NACβ2*) were purchased from the Nottingham Arabidopsis Stock Centre (NASC; <http://arabidopsis.info>, Nottingham, United Kingdom). The double homozygous mutants (*nacβ1nacβ2*) were acquired by a conventional cross of these two T-DNA insertion lines to acquire double heterozygous plants, and then they were self-pollinated to get the double homozygous mutant. The seeds were first sterilized briefly in ethanol, and then 5 min in 0.45% (w/v) sodium hypochlorite with Tween 20. The solution was exchanged for distilled water, and the seeds were washed five times in distilled water to remove residual sodium hypochlorite. Finally, the seeds were placed into 70% ethanol, from which were put onto sterilized filtration paper and were let dry. They were sown either onto plates with MS medium (1% w/v sucrose, 1% w/v agar, 2.3 g MS basal salts from Sigma-Aldrich, M5524 to 1 L media [70]) or directly onto Jiffy tablets. The plant pots/plates were stratified for 3 days in the cold room (4 °C) and then transferred to the cultivation room under long day (16h light/8h dark) conditions at 21 °C. 14 days after sowing, the plants (both from Jiffy tablets, and MS plates) were pricked to the new pots with Jiffy tablets.

##### 4.2. Insertion Lines and Genotyping

The plants were genotyped by PCR. First, the genomic DNA was extracted from leaves of individual plants by a modified CTAB (cetyltrimethylammonium bromide) protocol [71]. This extracted genomic DNA served as a template for two PCR reactions, the first one revealing the wild type allele by the forward and reverse gene-specific primers (Table S1), and the second one detecting

insert by a combination of either forward or reverse gene-specific and insert-specific primer (Table S1). The used insert-specific primers were LbB1.3 in the SALK line (SALK\_043673), and o8409 in the GK line (GK-368H02). The PCR reactions contained 0.4  $\mu$ M each primer (Sigma-Aldrich, Haverhill, United Kingdom), 0.2 mM each dNTP (Promega, Madison, USA), 1  $\mu$ L DNA template (the extracted genomic DNA from leaves), 1U *Taq* DNA polymerase (Merck, Brno, Czech Republic) and 1 $\times$  reaction buffer according to manufacturer's instructions. The PCR cycle was carried out as follows: 1) 94  $^{\circ}$ C, 2 min; 40 cycles: 2) 94  $^{\circ}$ C, 30 s, 3) 55  $^{\circ}$ C, 30 s, 4) 72  $^{\circ}$ C, 1.5 min; 5) 72  $^{\circ}$ C, 5 min. Post PCR, the reaction was blended with 6 $\times$  Loading Dye (Thermo Fisher Scientific, Waltham, USA) and its aliquot was electrophoresed in 1.2% (w/v) agarose gel with ethidium bromide in 1 $\times$  TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA [72]). GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) was used as a molecular marker. The gels were imaged by Gel Documentation System G-Box (Syngene, Cambridge, United Kingdom).

#### 4.3. Phenotype Characterization

Several phenotypic traits of the *nac $\beta$ 1nac $\beta$ 2* plants were compared to Col-0 wild type plants. These observations were performed in three independent sample batches considered as biological replicates. (i) The chlorophyll content in the leaves was determined according to Witham et al. (1971) [73] with slight modifications. Briefly, 0.1 g fresh weight of leaves were grinded by a pestle in a mortar with 5 mL 80% (v/v) acetone. The extract was filtered, and the sample was topped up to 10 mL with 80% (v/v) acetone. The optical density of the extracted samples was measured by a spectrophotometer (BioMate 3, Thermo Fisher Scientific, Waltham, USA) blanked with 80% acetone in 10 mm thin cuvettes at 645 nm, 652 nm, and 663 nm. Finally, chlorophyll content was calculated according to the formulas given in Witham et al. (1971) [73]. (ii) Flower buds, which were maximum one day before anthesis, were dissected and observed under the dissection microscope and the numbers of flower organs (sepals, petals, and anthers) were calculated and categorized. (iii) The length of green immature siliques on the main stem was measured by a paper ruler. (iv) The number of seeds inside the siliques was determined as follows. The green immature siliques were cut from the plant, stuck to the glass slide by a double-sided tape, and opened by an injection needle. The number of developed green seeds, pale normal-sized seeds, and aborted seeds was calculated under the dissection microscope (Zeiss Stemi 508, Oberkochen, Germany). The graphs were constructed in MS Excel (Microsoft, Redmond, USA).

#### 4.4. Transmission Analysis

The heterozygous plants in one *NAC $\beta$*  gene on the mutant background of the other one (*nac $\beta$ 1/+nac $\beta$ 2*, and *nac $\beta$ 1nac $\beta$ 2/+*; + symbolizes heterozygous gene) were let self-pollinate. Their seeds were cultivated as above and genotyped by PCR under the conditions as above. From the offspring genotype, the transmission coefficients were calculated. Next, several crosses were performed. Firstly, the heterozygous plants in one *NAC $\beta$*  gene on the mutant background of the other one (*nac $\beta$ 1/+nac $\beta$ 2*, and *nac $\beta$ 1nac $\beta$ 2/+*, respectively) were crossed with Col-0 wild type plants in both cases (carrying two wild type alleles of each *NAC $\beta$*  gene). Secondly, the heterozygous plants in one *NAC $\beta$*  gene on the mutant background of the other one (*nac $\beta$ 1/+nac $\beta$ 2*, *nac $\beta$ 1nac $\beta$ 2/+*, respectively) were crossed with plants representing the same gene as a homozygous mutant (*nac $\beta$ 2*, and *nac $\beta$ 1*, respectively). All the mentioned crosses were carried reciprocally, plants of each genotype serving once as the donor of the pistils and the second time donating pollen grains. The crosses were performed as follows; the flower buds before anther dehiscence of the pistil donors were emasculated and the pistils were let to mature for 1–2 days. After their papillary cells were developed, they were pollinated with the desired pollen. The acquired seeds were sown and cultivated as above, and the plants were conventionally genotyped by PCR using the primers as above. From the offspring genotype, the transmission coefficients were calculated, and  $\chi^2$ -test was performed.

#### 4.5. Pollen Tube Cultivation *in vitro*

Pollen tube cultivation *in vitro* was carried out according to Boavida and McCormick (2007) [34]. Briefly, the cultivation medium (0.01% (w/v)  $H_3BO_3$ , 1 mM  $CaCl_2$ , 1 mM KCl, 5% (w/v) sucrose, pH 7.5, 1% (w/v) low-melting agarose) was dropped onto the glass slide and let solidify. The tested pollen grains were shed from the open flowers onto the medium and the slides were incubated inside the small boxes in the cultivation room. After 8-hours growth, the pollen tubes were observed under the inverted epifluorescent microscope (Nikon TE2000E, Tokyo, Japan), and the images analyzed by NIS elements (Nikon, Tokyo, Japan).

#### 4.6. Pollen Tube Cultivation *in vivo* and Aniline Blue Staining

The young flower buds were emasculated before anther dehiscence. The pistils were let to develop papillary cells for 1–2 days, and then were pollinated. The pollen tubes were let grow overnight. The pistils were stained by aniline blue according to Mori et al. (2006) [74]. Briefly, the harvested pistils were fixed for 2 hours in ethanol with acetic acid (3:1), and subsequently hydrated in ethanol series (70%, 50%, and 30% ethanol). After an overnight incubation in 8 M NaOH to allow tissue softening, the callose in the cell walls of the pollen tubes was stained by 0.1% (w/v) aniline blue in 108 mM  $K_3PO_4$  (pH 11) and 2% (v/v) glycerol. Prior to the observation under epifluorescent microscope (Nikon TE2000E, Tokyo, Japan), the samples were transferred to distilled water.

#### 4.7. Blue Dot Assay

The blue dot assay was performed as described previously [37]. Briefly, the young flower buds before anther dehiscence were emasculated, and the pistils were let mature for 1–2 days. The prepared pistils were pollinated with wild type pollen grains carrying pLAT52–GUS construct. The pistils were collected 24 hours after pollination, dissected under dissection microscope, and transferred to the GUS staining solution (50 mM sodium phosphate buffer, pH 7.0, 0.2% Triton X-100, 10 mM potassium ferrocyanide, 10 mM potassium ferricyanide, and 1 mM X-Gluc [5-bromo-4-chloro-3-indolyl-D-glucoronid acid]). After overnight staining at 37 °C, the samples were observed under the epifluorescent microscope (Nikon TE2000E, Tokyo, Japan) in the bright field.

#### 4.8. Cloning

The cloned DNA fragments were amplified from genomic DNA (extracted by CTAB as above for plant genotyping). At first, the DNA fragments were synthesized by a 2-step PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's instructions. The DNA fragments for cloning were synthesized by PCR primed by the primers (Sigma-Aldrich, Haverhill, United Kingdom) in Table S1. The second PCR reaction was in all cases except for the DD33 promoter primed by the AttB1 and AttB2 primer pair. The DD33 promoter was in the second PCR reaction amplified by the primer pair AttB4 and AttB1R. The amplified sequences except for the DD33 promoter bordered by attB1/B2-overhangs were recombined by Gateway™ BP Clonase™ II Enzyme mix (Thermo Fisher Scientific, Waltham, USA) to pDONR221 (Thermo Fisher Scientific, Waltham, USA), thus creating entry clones. The DD33 promoter with attB4/1R overhangs was recombined by BP reaction under the same conditions into the vector pDONRP4-P1r (Thermo Fisher Scientific, Waltham, USA). The vectors were transformed to self-prepared chemically competent *E. coli* of the TOP10 strain. After plasmid isolation by GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, USA), and verification by Sanger sequencing, the fragments were recombined by Gateway™ LR Clonase™ II Enzyme mix (Thermo Fisher Scientific, Waltham, USA) to the following destination vectors. The overexpression constructs tagged with GFP were formed from vectors pFAST-R05 (C-terminal GFP fusion), and pFAST-R06 (N-terminal GFP fusion) [75]. The constructs for complementation analysis and protein localization were derived from pB7FWG,0 (<https://gateway.psb.ugent.be>, Ghent, Belgium [76]) whereas the constructs for promoter activity were



created from pKGWFS7,0 (<https://gateway.psb.ugent.be>, Ghent, Belgium [76]). The overexpression experiments driven by LAT52 [39] and DD33 [40] promoters were carried out by destination vector pB7m34GW (Thermo Fisher Scientific, Waltham, USA) putting three fragments together. For LAT52 overexpression, pDONRP4-P1r with LAT52 promoter donated by David Twell's laboratory was recombined with pDONR221 carrying the coding genomic sequence of the *NACβ* genes, and with pDONRP2r-P3 with GFP sequence donated by David Twell's laboratory. For DD33 overexpression, pDONRP4-P1r with DD33 promoter prepared as above was recombined with pDONR221 carrying the coding genomic sequence of the *NACβ* genes, and with pDONRP2r-P3 with GFP sequence donated by David Twell's laboratory. The acquired destination vectors were again transformed to self-prepared chemically competent *E. coli* of the TOP10 strain. The vectors were isolated from the selected *E. coli* suspensions by GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, USA) and subsequently verified by Sanger sequencing.

#### 4.9. *Arabidopsis Thaliana* Transformation and Microscopy

The *Agrobacterium tumefaciens* strain GV3101 (homemade complementary cells) was transformed by electroporation with all the isolated constructs, and subsequently *Arabidopsis thaliana* wild type (protein localization, overexpression, promoter analysis) or *nacβ1/+nacβ2*, and *nacβ1nacβ2/+* (complementation analysis) flowers were dipped into *A. tumefaciens* cell suspension to acquire stable transformants [77], which were selected either on MS medium [70] with proper selection (BASTA 10  $\mu\text{g}\cdot\mu\text{L}^{-1}$ ; vector pB7FWG, 0 for protein localization and complementation analysis, and vector pB7m34GW for proDD33- and proLAT52-driven overexpression or kanamycin 50  $\mu\text{g}\cdot\mu\text{L}^{-1}$ ; vector pKGWFS7,0 for promoter analysis) or by screening for the red seeds under fluorescent microscope (vectors pFAST-R05, and pFAST-R06 for pro35S-driven overexpression). The presence of the inserts in the transformed plants was verified by PCR with GFP forward and reverse primer pair. Then, the fluorescent signal was checked under the epifluorescent inverted microscope (Nikon TE2000E, Tokyo, Japan). The best transformants were selected and observed in more detail under confocal fluorescent microscope (Zeiss LSM 880 with Airyscan detector, Oberkochen, Germany).

#### 4.10. Promoter Analysis

The promoter analysis was performed by constructs tagged with both GFP and GUS. The transformants were selected according to the presence of the insert detected by PCR as above and according to the presence of GFP signal under the epifluorescent microscope. The selected plants were screened for GUS signal [78]. Briefly, various tissue types (seedlings, leaves, cauline leaves, inflorescences, flower buds, flowers, pollen grains, siliques) were incubated usually for 2 h (in case of low signal, the duration was prolonged to 24 h) in GUS staining buffer (100 mM sodium phosphate buffer pH 7.0, 10 mM EDTA, 0.1% (v/v) Triton X-100, 0.5 mM potassium ferrocyanide, 0.5 mM potassium ferricyanide, and 1 mM X-Gluc), and then bleached with ethanol series (90%, 70%, 50%). The tissues from several representative plants were observed and photographed under the dissection microscope (Zeiss Stemi 508, Oberkochen, Germany).

#### 4.11. RNA Extraction, cDNA Library Preparation and Sequencing

The total RNA was extracted from harvested *A. thaliana* young flower buds (up to stage 12 [43]) of both single *NACβ* homozygous mutants (*nacβ1*, and *nacβ2*) together with double *NACβ* homozygous mutant (*nacβ1nacβ2*) and control wild type Col-0 plants (genotype *NACβ1NACβ2*) using the RNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. Collected RNA was treated with DNA-free™ DNA Removal Kit (Thermo Fisher Scientific, Waltham, USA) and the integrity, purity and concentration of samples was measured on bioanalyzer (IMG-CORE-AGILENT 2100, Santa Clara, USA). Three biological replicates of each genotype were used for sequencing. 3  $\mu\text{g}$  of high-quality RNA was used for library preparation. The libraries were prepared for paired-end sequencing by TruSeq Stranded mRNA LT Sample Prep Kit and TruSeq Stranded Total RNA LT Sample

Prep Kit (Plant) following the TruSeq Stranded mRNA Sample Preparation Guide, Part # 15031047 Rev. E and TruSeq Stranded Total RNA Sample Prep Guide, Part # 15031048 Rev. E and sequenced on Illumina platform (Illumina, California, USA) by Macrogen Inc. (Seoul, Republic of Korea). The data were deposited into BioProject ID PRJNA589533 (<https://www.ncbi.nlm.nih.gov/bioproject/589533>).

#### 4.12. Mapping and Assembly Of Reads

Raw reads were further processed as follows. The quality of pair-end raw reads was revised by FastQC ver. 0.11.8 [79] (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), and Cutadapt ver. 1.9.1 [80]. The quality reads (phred score > 20) were also trimmed of technical sequences using the same Cutadapt software. Paired-reads with at least one read shorter than 20 bp were also excluded. Consequently, the reads were mapped to reference genome of *A. thaliana* (ver. TAIR10) downloaded from the Araport database ([https://www.araport.org/downloads/TAIR10\\_genome\\_release](https://www.araport.org/downloads/TAIR10_genome_release)) by STAR software ver. 206.1a [81] and then, the reads were counted by the featureCounts program from the Subread package ver. 1.6.3 [82]. Differential expression analyses were performed with the DESeq2 software ver. 3.8 [83] with adjusted p-value < 0,05 and FoldChange  $\geq \pm 2$  used as thresholds for establishing significantly differentially expressed genes (DEGs).

#### 4.13. Annotation and Enrichment Analyses

To bring deeper insight into the functional significance of DEGs, GO enrichment analyses were performed using Panther database (<http://pantherdb.org>, annotation version: GO Ontology database Released 2019-07-03) using Fisher's exact test with false discovery rate < 0,05 as statistical significance threshold. ThaleMine, tool of Araport website, was used for functional annotation of DEGs list (<https://apps.araport.org/thalemine/begin.do>, ver. 1.10.4). To identify molecular pathways, list of DEGs was further analyzed using the KEGG database (<https://www.genome.jp/kegg>, Release 89.0).

#### 4.14. Confirmation of DEGs Expression by qRT-PCR

Genes encoding NAC $\beta$  subunits (At1g17880, At1g73230), NAC $\alpha$  subunits (At3g12390, At3g49470, At5g13850, At4g10480, At1g33040) and five other genes (At5g64120, At2g20142, At5g52390, At2g43510, At3g12580) were chosen for qRT-PCR verification. EARLY FLOWERING 4 (eLF4; At2G40080) and tubulin beta chain 3 (tub3; At5g62700) were added as references. Residual RNA collected for sequencing of each plant genotype (*nac $\beta$ 2*, *nac $\beta$ 1*, *nac $\beta$ 1nac $\beta$ 2*, *NAC $\beta$ 1NAC $\beta$ 2*) was used for the qRT-PCR in three biological replicas. The reactions were prepared with GoTaq<sup>®</sup>qPCR Master Mix (Promega, Madison, WI, USA) and analyzed on LightCycler 480 Instrument (Roche, Basel, Switzerland) according to the manufacturer's instructions using the primer sets from Table S2. cT values of each sample were recorded by the LightCycler 480 Software version 1.5 (Roche, Basel, Switzerland). Quality control of the obtained data was performed by the same software. Fragments per kilobase million (FPKM) values of transcript from RNA sequencing (RNA-seq) data were calculated with RSEM software [84]. Linear regression of cT values and log based FPKM values was projected in MS Excel (Microsoft, Redmond, USA).

#### 4.15. Protein Extraction and Proteome Analysis

The total proteins were extracted from the harvested *A. thaliana* young flower buds (up to stage 12 [43]) of double NAC $\beta$  homozygous mutant (*nac $\beta$ 1nac $\beta$ 2*) and control wild type Col-0 plants (genotype *NAC $\beta$ 1NAC $\beta$ 2*) by TRI-reagent (Sigma-Aldrich, St. Louis, USA) according to the manufacturer's instructions with slight modifications [85]. Briefly, 45 mg flower buds were snap-frozen in liquid nitrogen and then homogenized by a pestle in a mortar. To the grinded plant material, TRI reagent was added stepwise per 200  $\mu$ L to achieve a final volume of 1 mL to extract the proteins. Then, 200  $\mu$ L chloroform was added, and the sample was let on ice for 15 min. After centrifugation (20,000 $\times$ g, 20 min, 4  $^{\circ}$ C), the lower phase was mixed with 300  $\mu$ L ethanol. After another round of centrifugation (2,000 $\times$ g, 5 min, 4  $^{\circ}$ C), the supernatant was mixed with 1.5 volumes isopropanol and the proteins

were let precipitate at room temperature for 1 h. The proteins were pelleted (20,000×g, 20 min, 15 °C), washed 3 times with 1 mL cold ethanol and centrifuged again (20,000×g, 20 min, 4 °C). The final protein pellet was vacuum dried and stored at −20 °C until further use.

Dried pellets were solubilized by SDT buffer (4% SDS, 0.1 M DTT, 0.1 M Tris-HCl, pH 7.6) and protein concentration measured by native tryptophan fluorescence. Fifty micrograms of proteins were used for filter-aided sample preparation (FASP) method [86] with some modifications. The samples were mixed with 8M UA buffer (8 M urea in 100 mM Tris-HCl, pH 8.5), loaded onto the Microcon device with MWCO 30 kDa (Merck Millipore, Burlington, Massachusetts, USA) and centrifuged (7,000×g, 30 min, 20 °C). The retained proteins were washed (all centrifugation steps after sample loading were performed at 14,000× g) with 200 µL UA buffer. The washed protein concentrates kept in the Microcon device were mixed with 100 µL of UA buffer containing 50 mM iodoacetamide and incubated in the dark for 20 min. After the next centrifugation step, the samples were washed three times with 100 µL UA buffer and three times with 100 µL of 50 mM NaHCO<sub>3</sub>. Trypsin (sequencing grade, Promega, Madison, USA) was added onto the filter and the mixture was incubated for 18 h at 37 °C (enzyme:protein ratio 1:50). The tryptic peptides were eluted by centrifugation followed by two additional elutions with 50 µL of 50 mM NaHCO<sub>3</sub>. Peptides were extracted into LC-MS vials by 2.5% formic acid (FA) in 50% acetonitrile (ACN) and 100% ACN with addition of polyethylene glycol (20,000; final concentration 0.001%; [87]) and concentrated in a SpeedVac concentrator (Thermo Fisher Scientific, Waltham, USA) prior to LC-MS/MS analyses.

LC-MS/MS analyses of all peptide mixtures were performed by RSLCnano system (SRD-3400, NCS-3500RS CAP, WPS-3000 TPL RS) connected to Orbitrap Q Exactive HF-X spectrometer (Thermo Fisher Scientific, Waltham, USA). Prior to LC separation, tryptic digests were online concentrated and desalted by trapping column (Acclaim™ PepMap™ 100 C18, dimensions 300 µm × 5 mm, 5 µm particles; part number 160454). After washing of trapping column with 0.1% FA, the peptides were eluted in backflush mode (flow 300 nL·min<sup>−1</sup>) from the trapping column onto an analytical column (Acclaim Pepmap100 C18, 3 µm particles, 75 µm × 500 mm; Thermo Fisher Scientific, Waltham, USA) by 100 min gradient program (2%–35% of mobile phase B; mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 80% ACN). Equilibration of the trapping column and the column was done prior to sample injection to sample loop. The analytical column outlet was directly connected to the Digital PicoView 550 (New Objective) ion source with sheath gas option and SilicaTip emitter (New Objective; FS360-20-15-N-20-C12) utilization. ABIRD (Active Background Ion Reduction Device, ESI Source Solutions) was installed.

MS data were acquired in a data-dependent strategy selecting up to top 20 precursors based on precursor abundance in the survey scan (350–2000 m/z). The resolution of the survey scan was 120 000 (200 m/z) with a target value of 3×10<sup>6</sup> ions and maximum injection time of 50 ms. HCD MS/MS (28% relative fragmentation energy) spectra were acquired with a target value of 10,000 and resolution of 15,000 (200 m/z). The maximum injection time for MS/MS was 50 ms. Dynamic exclusion was enabled for 40 s after one MS/MS spectra acquisition. The isolation window for MS/MS fragmentation was set to 1.2 m/z.

The analysis of the mass spectrometric RAW data files was carried out by the MaxQuant software (version 1.6.2.10) using default settings unless otherwise noted. MS/MS ion searches were performed against modified cRAP database (based on <http://www.thegpm.org/crap>) containing protein contaminants like keratin, trypsin etc., and UniProtKB protein database for *Arabidopsis thaliana* ([ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/reference\\_proteomes/Eukaryota/UP000006548\\_3702.fasta.gz](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/Eukaryota/UP000006548_3702.fasta.gz); downloaded 8. 7. 2019, version 2019/07, number of protein sequences 27,476). Oxidation of methionine and proline, deamidation (N, Q) and acetylation (protein N-terminus) as optional modification, carbamidomethylation (C) as fixed modification and trypsin/P enzyme with 2 allowed miss cleavages were set. Peptides and proteins with FDR threshold < 0.01 and proteins having at least one unique or razor peptide were considered

only. Match between runs was set among all analyzed samples. Protein abundance was assessed according to protein intensities calculated by MaxQuant.

Protein intensities reported in proteinGroups.txt were further evaluated using the software container environment ([https://github.com/OmicsWorkflows/KNIME\\_docker\\_vnc](https://github.com/OmicsWorkflows/KNIME_docker_vnc); version 3.7.2a). Processing workflow is available upon request and it covered decoy hits and removal of contaminant protein groups, protein group intensities  $\log_2$  transformation, normalization, imputation of missing values (imp4p R package) and statistical analysis (LIMMA; p values adjustment using Benjamini and Hochberg approach). The final list of differently expressed proteins was acquired according to the following criteria: the fold change > 2, adjusted p value < 0.05; protein groups had to have at least 2 peptides and non-zero protein group intensity in at least 2 replicates of at least one sample.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016315.

#### 4.16. Plant Cultivation Under The Salt and Osmotic Stress

The seeds of Columbia-0 (wild type), and double NAC $\beta$  homozygous mutant (*nac $\beta$ 1nac $\beta$ 2*) were sterilized as above and sown onto the plates with MS medium (1% w/v sucrose, 1% w/v agar, 2.3 g MS basal salts from Sigma-Aldrich, M5524 to 1 L media [70] supplemented either with 50 mM, 100 mM, 125 mM, and 150 mM NaCl, respectively or 250 mM, 300 mM, and 350 mM mannitol, respectively. The plates were stratified in the cold room for 1 day and then transferred to the cultivation room. The proportion of grown seedlings was calculated 4, 6, 8, 11, and 13 days after the transfer to the cultivation room.

## 5. Conclusions

Collectively, this paper presented the functional analysis of *nac $\beta$ 1nac $\beta$ 2* mutants and proposed the role of the  $\beta$ -subunit of the nascent polypeptide-associated complex during flower and silique development of *Arabidopsis thaliana*. Moreover, the flower bud transcriptomic and proteomic data suggested that NAC $\beta$  subunits are likely involved in stress responses and male gametophyte development. The role of NAC complex during stress is consistent with the data acquired previously on different plant species and our experimental data.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/6/2065/s1>.

**Author Contributions:** Conceptualization, J.F., B.K. and D.H.; Data curation, D.P., M.J. and Z.Z.; Investigation, J.F., B.K., D.P. and P.Š.; Methodology, J.F., B.K., D.P., M.J. and Z.Z.; Project administration, J.F. and D.H.; Resources, J.F., Z.Z. and D.H.; Supervision, D.H.; Writing—original draft, J.F. and B.K.; Writing—review & editing, D.P., Z.Z. and D.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was funded by the Czech Ministry of Education, Youth and Sports [LTC18043] and the Czech Science Foundation [18-02448S, and 19-01723S]. The work was also supported from European Regional Development Fund-Project “Centre for Experimental Plant Biology” (No. CZ.02.1.01/0.0/0.0/16\_019/0000738). CIISB research infrastructure project LM2018127 funded by MEYS CR is gratefully acknowledged for financially supporting the LC-MS/MS measurements at the Proteomics Core Facility. Computational resources were supplied by the project “e-Infrastruktura CZ” (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

**Acknowledgments:** Special thanks also belong to Nina Lindstrøm Friggens for her help during plant genotyping. The authors also thank to Said Hafidh, and Katarína Kulichová for fruitful discussions and demonstration of in vivo pollen tube staining and blue dot assay, and Karel Müller for the supervision of qRT-PCR experiments.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

<i>ag</i>	<i>agamous</i>
<i>ap3</i>	<i>apetala 3</i>
BASTA	(RS)-2-Amino-4-(hydroxy(methyl)phosphonoyl)butanoic acid
bp	base pair
Col-0	Columbia-0
CaMV	cauliflower mosaic virus
<i>clv1</i>	<i>clavata 1</i>
<i>clv3</i>	<i>clavata 3</i>
CTAB	cetyltrimethylammonium bromide
DE	differential expression
DEG	differentially expressed gene
DEP	differentially expressed protein
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
GFP	green fluorescent protein
FPKM	fragments per kilobase million
GO	gene ontology
GUS	glucuronidase
HSP	heat shock protein
MS	Murashige and Skoog growth medium
NAC	nascent polypeptide-associated complex
<i>nacβ1nacβ2</i>	NACβ double homozygous mutant
PCR	polymerase chain reaction
<i>pi</i>	<i>pistillata</i>
qRT-PCR	quantitative reverse transcriptase polymerase chain reaction
RNA-seq	RNA sequencing
<i>sep</i>	<i>sepallata</i>
TAE	electrophoresis buffer composed of Tris base, acetic acid and EDTA
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
<i>wig</i>	<i>wiggum</i>
wt	wild type

## References

1. Kogan, G.L.; Gvozdev, V.A. Multifunctional nascent polypeptide-associated complex (NAC). *Mol. Biol.* **2014**, *48*, 189–196. [[CrossRef](#)]
2. Spreter, T.; Pech, M.; Beatrix, B. The crystal structure of archaeal nascent polypeptide-associated complex (NAC) reveals a unique fold and the presence of a ubiquitin-associated domain. *J. Boil. Chem.* **2005**, *280*, 15849–15854. [[CrossRef](#)] [[PubMed](#)]
3. Rospert, S.; Dubaquié, Y.; Gautschi, M. Nascent-polypeptide-associated complex. *Cell. Mol. Life Sci.* **2002**, *59*, 1632–1639. [[CrossRef](#)] [[PubMed](#)]
4. Reimann, B.; Bradsher, J.; Franke, J.; Hartmann, E.; Wiedmann, M.; Prehn, S.; Wiedmann, B. Initial characterization of the nascent polypeptide-associated complex in yeast. *Yeast* **1999**, *15*, 397–407. [[CrossRef](#)]
5. Ott, A.-K.; Locher, L.; Koch, M.; Deuerling, E. Functional dissection of the nascent polypeptide-associated complex in *Saccharomyces cerevisiae*. *PLoS ONE* **2015**, *10*, e0143457. [[CrossRef](#)] [[PubMed](#)]
6. Wang, S.; Sakai, H.; Wiedmann, M. NAC covers ribosome-associated nascent chains thereby forming a protective environment for regions of nascent chains just emerging from the peptidyl transferase center. *J. Cell Biol.* **1995**, *130*, 519–528. [[CrossRef](#)]
7. Wiedmann, B.; Sakai, H.; Davis, T.A.; Wiedmann, M. A protein complex required for signal-sequence-specific sorting and translocation. *Nature* **1994**, *370*, 434–440. [[CrossRef](#)]

8. Martin, E.M.; Jackson, M.P.; Gamedinger, M.; Gense, K.; Karamonos, T.K.; Humes, J.; Deuerling, E.; Ashcroft, A.E.; Radford, S.E. Conformational flexibility within the nascent polypeptide-associated complex enables its interactions with structurally diverse client proteins. *J. Biol. Chem.* **2018**, *293*, 8554–8568. [[CrossRef](#)]
9. Wegrzyn, R.D.; Hofmann, D.; Merz, F.; Nikolay, R.; Rauch, T.; Graf, C.; Deuerling, E. A conserved motif is prerequisite for the interaction of NAC with ribosomal protein L23 and nascent chains. *J. Biol. Chem.* **2005**, *281*, 2847–2857. [[CrossRef](#)]
10. Möller, I.; Beatrix, B.; Kreibich, G.; Sakai, H.; Lauring, B.; Wiedmann, M. Unregulated exposure of the ribosomal M-site caused by NAC depletion results in delivery of non-secretory polypeptides to the Sec61 complex. *FEBS Lett.* **1998**, *441*, 1–5. [[CrossRef](#)]
11. Funfschilling, U.; Rospert, S. Nascent polypeptide-associated complex stimulates protein import into yeast mitochondria. *Mol. Biol. Cell* **1999**, *10*, 3289–3299. [[CrossRef](#)] [[PubMed](#)]
12. Gamedinger, M.; Hanebuth, M.A.; Frickey, T.; Deuerling, E. The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science* **2015**, *348*, 201–207. [[CrossRef](#)] [[PubMed](#)]
13. Koplin, A.; Preissler, S.; Ilina, Y.; Koch, M.; Scior, A.; Erhardt, M.; Deuerling, E. A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. *J. Cell Biol.* **2010**, *189*, 57–68. [[CrossRef](#)] [[PubMed](#)]
14. Bloss, T.A.; Witze, E.S.; Rothman, J.H. Suppression of CED-3-independent apoptosis by mitochondrial  $\beta$ NAC in *Caenorhabditis elegans*. *Nature* **2003**, *424*, 1066–1071. [[CrossRef](#)] [[PubMed](#)]
15. Deng, J.M.; Behringer, R.R. An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Res.* **1995**, *4*, 264–269. [[CrossRef](#)] [[PubMed](#)]
16. Markesich, D.C.; Gajewski, K.M.; Nazimiec, M.E.; Beckingham, K. bicaudal encodes the *Drosophila beta* NAC homolog, a component of the ribosomal translational machinery. *Development* **2000**, *127*, 559–572. [[PubMed](#)]
17. Brockstedt, E.; Otto, A.; Rickers, A.; Bommert, K.; Wittmann-Liebold, B. Preparative high-resolution two-dimensional electrophoresis enables the identification of RNA polymerase B transcription factor 3 as an apoptosis-associated protein in the human BL60-2 Burkitt lymphoma cell line. *Protein J.* **1999**, *18*, 225–231. [[CrossRef](#)]
18. Moreau, A.; Yotov, W.V.; Glorieux, F.H.; St-Arnaud, R. Bone-Specific Expression of the alpha chain of the nascent polypeptide-associated complex, a coactivator potentiating c-Jun-mediated transcription. *Mol. Cell. Biol.* **1998**, *18*, 1312–1321. [[CrossRef](#)]
19. Yotov, W.V.; Moreau, A.; St-Arnaud, R. The alpha chain of the nascent polypeptide-associated complex functions as a transcriptional coactivator. *Mol. Cell. Biol.* **1998**, *18*, 1303–1311. [[CrossRef](#)]
20. Berardini, T.; Reiser, L.; Li, N.; Mezheritsky, Y.; Muller, R.; Strait, E.; Huala, E. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **2015**, *53*, 474–485. [[CrossRef](#)]
21. Ding, Y.; Jia, Y.; Shi, Y.; Zhang, X.; Song, C.; Gong, Z.; Yang, S. OST 1-mediated BTF 3L phosphorylation positively regulates CBF s during plant cold responses. *EMBO J.* **2018**, *37*, e98228. [[CrossRef](#)] [[PubMed](#)]
22. Park, H.L.; Cho, M.-H.; Shin, D.H.; Bhoo, S.H.; Hahn, T.-R. Molecular characterization of phytochrome-interacting nascent polypeptide associated complex protein in *Arabidopsis thaliana*. *J. Korean Soc. Appl. Boil. Chem.* **2012**, *55*, 323–326. [[CrossRef](#)]
23. Freire, M.A. Translation initiation factor (iso) 4E interacts with BTF3, the  $\beta$  subunit of the nascent polypeptide-associated complex. *Gene* **2005**, *345*, 271–277. [[CrossRef](#)] [[PubMed](#)]
24. Yang, K.-S.; Kim, H.-S.; Jin, U.-H.; Lee, S.S.; Park, J.-A.; Lim, Y.P.; Pai, H.-S. Silencing of NbBTF3 results in developmental defects and disturbed gene expression in chloroplasts and mitochondria of higher plants. *Planta* **2007**, *225*, 1459–1469. [[CrossRef](#)]
25. Ma, H.-Z.; Liu, G.-Q.; Li, C.-W.; Kang, G.; Guo, T.-C. Identification of the TaBTF3 gene in wheat (*Triticum aestivum* L.) and the effect of its silencing on wheat chloroplast, mitochondria and mesophyll cell development. *Biochem. Biophys. Res. Commun.* **2012**, *426*, 608–614. [[CrossRef](#)] [[PubMed](#)]
26. Wang, W.; Xu, M.; Wang, Y.; Jamil, M. Basal transcription factor 3 plays an important role in seed germination and seedling growth of rice. *BioMed Res. Int.* **2014**, *2014*, 465739. [[CrossRef](#)]
27. Maršálová, L.; Vitámvás, P.; Hynek, R.; Prášil, I.T.; Kosová, K. Proteomic response of *Hordeum vulgare* cv. Tadmor and *Hordeum marinum* to salinity stress: Similarities and differences between a glycophyte and a halophyte. *Front. Plant Sci.* **2016**, *7*, 2567. [[CrossRef](#)]

28. Karan, R.; Subudhi, P.K. Overexpression of a nascent polypeptide associated complex gene (Sa $\beta$ NAC) of *Spartina alterniflora* improves tolerance to salinity and drought in transgenic Arabidopsis. *Biochem. Biophys. Res. Commun.* **2012**, *424*, 747–752. [[CrossRef](#)]
29. Kang, G.; Ma, H.; Liu, G.; Han, Q.; Li, C.; Guo, T. Silencing of TaBTF3 gene impairs tolerance to freezing and drought stresses in wheat. *Mol. Genet. Genom.* **2013**, *288*, 591–599. [[CrossRef](#)]
30. Yin, X.; Nishimura, M.; Hajika, M.; Komatsu, S. Quantitative proteomics reveals the flooding-tolerance mechanism in mutant and abscisic acid-treated soybean. *J. Proteome Res.* **2016**, *15*, 2008–2025. [[CrossRef](#)]
31. Huh, S.U.; Kim, K.-J.; Paek, K.-H. Capsicum annuum basic transcription factor 3 (CaBtf3) regulates transcription of pathogenesis-related genes during hypersensitive response upon Tobacco mosaic virus infection. *Biochem. Biophys. Res. Commun.* **2012**, *417*, 910–917. [[CrossRef](#)] [[PubMed](#)]
32. Kaido, M.; Inoue, Y.; Takeda, Y.; Sugiyama, K.; Takeda, A.; Mori, M.; Tamai, A.; Meshi, T.; Okuno, T.; Mise, K. Downregulation of the NbNACa1 gene encoding a movement-protein-interacting protein reduces cell-to-cell movement of Brome mosaic virus in *Nicotiana benthamiana*. *Mol. Plant-Microbe Interactions* **2007**, *20*, 671–681. [[CrossRef](#)] [[PubMed](#)]
33. Reňák, D.; Dupl'áková, N.; Honys, D. Wide-scale screening of T-DNA lines for transcription factor genes affecting male gametophyte development in Arabidopsis. *Sex. Plant Reprod.* **2011**, *25*, 39–60. [[CrossRef](#)] [[PubMed](#)]
34. Boavida, L.C.; McCormick, S. Temperature as a determinant factor for increased and reproducible *in vitro* pollen germination in *Arabidopsis thaliana*. *Plant J.* **2007**, *52*, 570–582. [[CrossRef](#)] [[PubMed](#)]
35. Qin, Y.; Leydon, A.R.; Manziello, A.; Pandey, R.; Mount, D.; Denic, S.; Vasic, B.; Johnson, M.; Palanivelu, R. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.* **2009**, *5*, e1000621. [[CrossRef](#)] [[PubMed](#)]
36. Regan, S.M.; Moffatt, B.A. Cytochemical analysis of pollen development in wild-type Arabidopsis and a male sterile mutant. *Plant. Cell* **1990**, *2*, 877–889. [[CrossRef](#)] [[PubMed](#)]
37. Hafidh, S.; Potěšil, D.; Fila, J.; Čapková, V.; Zdráhal, Z.; Honys, D. Quantitative proteomics of the tobacco pollen tube secretome identifies novel pollen tube guidance proteins important for fertilization. *Genome Biol.* **2016**, *17*, 81. [[CrossRef](#)] [[PubMed](#)]
38. Apel, K.; Bohlmann, H. Comparison of different constitutive and inducible promoters for the overexpression of transgenes in *Arabidopsis thaliana*. *Plant Mol. Biol.* **1995**, *29*, 637–646.
39. Twell, D.; Yamaguchi, J.; McCormick, S. Pollen-specific gene expression in transgenic plants: Coordinate regulation of two different tomato gene promoters during microsporogenesis. *Development* **1990**, *109*, 705–713.
40. Steffen, J.G.; Kang, I.-H.; Macfarlane, J.; Drews, G.N. Identification of genes expressed in the Arabidopsis female gametophyte. *Plant J.* **2007**, *51*, 281–292. [[CrossRef](#)]
41. Nakabayashi, K.; Okamoto, M.; Koshiba, T.; Kamiya, Y.; Nambara, E. Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: Epigenetic and genetic regulation of transcription in seed. *Plant J.* **2005**, *41*, 697–709. [[CrossRef](#)] [[PubMed](#)]
42. Schmid, M.; Davison, T.S.; Henz, S.R.; Pape, U.J.; Demar, M.; Vingron, M.; Schölkopf, B.; Weigel, D.; Lohmann, J.U. A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **2005**, *37*, 501–506. [[CrossRef](#)] [[PubMed](#)]
43. Cardarelli, M.; Cecchetti, V. Auxin polar transport in stamen formation and development: How many actors? *Front. Plant Sci.* **2014**, *5*, 333. [[CrossRef](#)] [[PubMed](#)]
44. Supek, F.; Bosnjak, M.; Skunca, N.; Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [[CrossRef](#)] [[PubMed](#)]
45. Winter, D.; Vinegar, B.; Nahal, H.; Ammar, R.; Wilson, G.; Provart, N.J. An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* **2007**, *2*, e718. [[CrossRef](#)]
46. Tsukaya, H.; Byrne, M.E.; Horiguchi, G.; Sugiyama, M.; Van Lijsebettens, M.; Lenhard, M. How do ‘housekeeping’ genes control organogenesis?—unexpected new findings on the role of housekeeping genes in cell and organ differentiation. *J. Plant Res.* **2012**, *126*, 3–15. [[CrossRef](#)]
47. Riechmann, J.L. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucleic Acids Res.* **1996**, *24*, 3134–3141. [[CrossRef](#)]
48. Running, M.P.; Fletcher, J.C.; Meyerowitz, E.M. The WIGGUM gene is required for proper regulation of floral meristem size in Arabidopsis. *Development* **1998**, *125*, 2545–2553.

49. McKim, S.M.; Routier-Kierzkowska, A.-L.; Monniaux, M.; Kierzkowski, D.; Pieper, B.; Smith, R.S.; Tsiantis, M.; Hay, A.S. Seasonal regulation of petal number. *Plant Physiol.* **2017**, *175*, 886–903. [[CrossRef](#)]
50. Pelaz, S.; Ditta, G.S.; Baumann, E.; Wisman, E.; Yanofsky, M.F. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **2000**, *405*, 200–203. [[CrossRef](#)]
51. Yotov, W.V.; St-Arnaud, R. Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle-specific transcription factor. *Genes Dev.* **1996**, *10*, 1763–1772. [[CrossRef](#)] [[PubMed](#)]
52. Zheng, X.M.; Black, D.; Chambon, P.; Egly, J.M. Sequencing and expression of complementary DNA for the general transcription factor BTF3. *Nature* **1990**, *344*, 556–559. [[CrossRef](#)] [[PubMed](#)]
53. Liu, Y.; Hu, Y.; Li, X.; Niu, L.; Teng, M. The crystal structure of the human nascent polypeptide-associated complex domain reveals a nucleic acid-binding region on the NACA subunit. *Biochemistry* **2010**, *49*, 2890–2896. [[CrossRef](#)]
54. Miller, D.D.; Scordilis, S.P.; Hepler, P.K. Identification and localization of three classes of myosins in pollen tubes of *Lilium longiflorum* and *Nicotiana glauca*. *J. Cell Sci.* **1995**, *108*, 2549–2563. [[PubMed](#)]
55. Honys, D.; Reňák, D.; Feciková, J.; Jedelský, P.; Nebesářová, J.; Dobrev, P.; Čapková, V. Cytoskeleton-associated large RNP complexes in tobacco male gametophyte (EPPs) are associated with ribosomes and are involved in protein synthesis, processing, and localization. *J. Proteome Res.* **2009**, *8*, 2015–2031. [[CrossRef](#)]
56. Stinson, J.R.; Eisenberg, A.J.; Willing, R.P.; Pe, M.E.; Hanson, D.D.; Mascarenhas, J.P.; Anderson, J.A.; Ashworth, E.N. Genes expressed in the male gametophyte of flowering plants and their isolation. *Plant Physiol.* **1987**, *83*, 442–447. [[CrossRef](#)]
57. Hafidh, S.; Potěšil, D.; Müller, K.; Fila, J.; Michailidis, C.; Herrmannová, A.; Feciková, J.; Ischebeck, T.; Valášek, L.S.; Zdráhal, Z.; et al. Dynamics of the pollen sequestrome defined by subcellular coupled omics. *Plant Physiol.* **2018**, *178*, 258–282. [[CrossRef](#)]
58. Adamczyk, B.J.; Fernandez, D.E. MIK\**MADS* domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. *Plant Physiol.* **2009**, *149*, 1713–1723. [[CrossRef](#)]
59. Ge, Z.; Bergonci, T.; Zhao, Y.; Zou, Y.; Du, S.; Liu, M.-C.; Luo, X.; Ruan, H.; Garcia-Valencia, L.E.; Zhong, S.; et al. *Arabidopsis* pollen tube integrity and sperm release are regulated by RALF-mediated signaling. *Science* **2017**, *358*, 1596–1600. [[CrossRef](#)]
60. Ráček, N.; Wolf, S.; Kost, B.; Rausch, T.; Greiner, S.; Röckel, N. Elaborate spatial patterning of cell-wall PME and PME1 at the pollen tube tip involves PME1 endocytosis, and reflects the distribution of esterified and de-esterified pectins. *Plant J.* **2008**, *53*, 133–143. [[CrossRef](#)]
61. Brownfield, L.; Hafidh, S.; Borg, M.; Sidorova, A.; Mori, T.; Twell, D. A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet.* **2009**, *5*, e1000430. [[CrossRef](#)] [[PubMed](#)]
62. Takahashi, T.; Mori, T.; Ueda, K.; Yamada, L.; Nagahara, S.; Higashiyama, T.; Sawada, H.; Igawa, T. The male gamete membrane protein DMP9/DAU2 is required for double fertilization in flowering plants. *Development* **2018**, *145*, dev170076. [[CrossRef](#)]
63. Maruyama, D.; Sugiyama, T.; Endo, T.; Nishikawa, S.-I. Multiple BiP Genes of *Arabidopsis thaliana* are required for male gametogenesis and pollen competitiveness. *Plant Cell Physiol.* **2014**, *55*, 801–810. [[CrossRef](#)]
64. Sarkar, N.K.; Kim, Y.-K.; Grover, A. Rice sHsp genes: Genomic organization and expression profiling under stress and development. *BMC Genom.* **2009**, *10*, 393. [[CrossRef](#)] [[PubMed](#)]
65. Wehmeyer, N.; Klann, E.M.; Hall, B.; Bennett, A.B. Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Physiol.* **1996**, *112*, 747–757. [[CrossRef](#)] [[PubMed](#)]
66. Xu, Y.-H.; Liu, R.; Yan, L.; Liu, Z.-Q.; Jiang, S.-C.; Shen, Y.-Y.; Wang, X.-F.; Zhang, D.-P. Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in *Arabidopsis*. *J. Exp. Bot.* **2011**, *63*, 1095–1106. [[CrossRef](#)] [[PubMed](#)]
67. Varotto, C.; Pesaresi, P.; Meurer, J.; Oelmüller, R.; Steiner-Lange, S.; Salamini, F.; Leister, D. Disruption of the *Arabidopsis* photosystem I gene *psaE1* affects photosynthesis and impairs growth. *Plant J.* **2000**, *22*, 115–124. [[CrossRef](#)]
68. Jaiswal, H.; Conz, C.; Otto, H.; Wölfl, T.; Fitzke, E.; Mayer, M.P.; Rospert, S. The chaperone network connected to human ribosome-associated complex. *Mol. Cell. Boil.* **2011**, *31*, 1160–1173. [[CrossRef](#)]
69. Verma, A.K.; Diwan, D.; Raut, S.; Dobriyal, N.; Brown, R.E.; Gowda, V.; Hines, J.K.; Sahi, C. Evolutionary conservation and emerging functional diversity of the cytosolic Hsp70/J protein chaperone network of *Arabidopsis thaliana*. *G3: Genes/Genomes/Genetics* **2017**, *7*, 1941–1954. [[CrossRef](#)]



70. Murashige, T.; Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **1962**, *15*, 473–497. [[CrossRef](#)]
71. Weigel, D.; Glazebrook, J. *Arabidopsis: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2002.
72. Ogden, R.C.; Adams, D.A. Electrophoresis in agarose and acrylamide gels. *Methods Enzymol.* **1987**, *152*, 61–87. [[PubMed](#)]
73. Witham, F.H.; Blaydes, D.F.; Devlin, R.M. Experiment 15: Chlorophyll absorption spectrum and quantitative determinations. In *Experiments in Plant Physiology*; Van Nostrand Reinhold Co.: New York, NY, USA, 1971; p. 245.
74. Mori, T.; Kuroiwa, H.; Higashiyama, T.; Kuroiwa, T. GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nature* **2005**, *8*, 64–71. [[CrossRef](#)] [[PubMed](#)]
75. Shimada, T.L.; Shimada, T.; Hara-Nishimura, I. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *Plant J.* **2010**, *61*, 519–528. [[CrossRef](#)] [[PubMed](#)]
76. Karimi, M.; Inzé, D.; Depicker, A. GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.* **2002**, *7*, 193–195. [[CrossRef](#)]
77. Clough, S.J.; Bent, A. Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **1998**, *16*, 735–743. [[CrossRef](#)]
78. Hu, C.-Y.; Chee, P.; Chesney, R.; Zhou, J.; Miller, P.; O'Brien, W. Intrinsic GUS-like activities in seed plants. *Plant Cell Rep.* **1990**, *9*, 1–5. [[CrossRef](#)] [[PubMed](#)]
79. Andrews, S. FastQC. A quality control tool for high throughput sequence data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 6 December 2019).
80. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10. [[CrossRef](#)]
81. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2012**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
82. Liao, Y.; Smyth, G.K.; Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **2013**, *41*, e108. [[CrossRef](#)] [[PubMed](#)]
83. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 002832. [[CrossRef](#)] [[PubMed](#)]
84. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)]
85. Fila, J.; Čapková, V.; Feciková, J.; Honys, D. Impact of homogenization and protein extraction conditions on the obtained tobacco pollen proteomic patterns. *Biol. Plant.* **2011**, *55*, 499–506. [[CrossRef](#)]
86. Wiśniewski, J.R.; Ostasiewicz, P.; Mann, M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* **2011**, *10*, 3040–3049. [[CrossRef](#)] [[PubMed](#)]
87. Stejskal, K.; Potěšil, D.; Zdráhal, Z. Suppression of peptide sample losses in autosampler vials. *J. Proteome Res.* **2013**, *12*, 3057–3062. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

**The functional conservation of beta subunits of the nascent polypeptide associated complex in liverworts and angiosperms**

**My contribution:** In this study, I have proposed the tested hypothesis and planned the experiments. I carried out most of the experimental work. I have analysed all results and written the manuscript together with preparation of all figures.

Current status: on-going, to be submitted in the early 2024. Potentially to Plant and Cell Physiology.

# The functional conservation of beta subunits of the nascent polypeptide associated complex in liverworts and angiosperms

## Abstract

The evolutionary divergence of land plants from freshwater algae has given rise to diverse physiological adaptations necessary for terrestrial survival. Nascent polypeptide associated complex (NAC) is a chaperone complex present in all Eucaryots and Archaea that showed a significant high sequence and structure conservation. In plants, molecular mechanism of NAC is not known, however its role in stress adaptation and plant development has been reported in numerous species. This study examines the functional conservation of NAC $\beta$  subunit between the vascular angiosperm *Arabidopsis thaliana* and the non-vascular liverwort *Marchantia polymorpha*. We performed a complementation study on *A. thaliana* mutants of the NAC $\beta$  subunits by *M. polymorpha* NAC $\beta$  subunit. Our findings reveal a conserved function of NAC $\beta$  subunits across these divergent plant lineages. We propose that NAC function is conserved across all plant species and is crucial for plant development, as knock-out mutation seems to be lethal. Furthermore, we explored the functional divergence of *Arabidopsis* NAC subunits with the phylogeny and interaction analysis, addressing their relationship and potential sub-functionating. Overall, our findings support the theory that NAC $\beta$  function is conserved among land plants despite vast evolutionary distances, potentially due to its essential role in plant survival, as indicated by challenges in generating knock-out mutants in *M. polymorpha*.

## Introduction

Between 470 and 515 million years ago, land plants (Embryophyta) diverged from a freshwater algae, forming a monophyletic eukaryotic lineage (Bowman 2013; Mishler and Churchill 1985; Morris et al. 2018). Today, approximately 435,000 species of land plants exist on Earth, thriving across diverse ecosystems. Land plants are categorized into vascular and non-vascular groups. Non-vascular plants, such as Bryophytes, lack a vascular system. Vascular plants include angiosperms, with the model organism *Arabidopsis thaliana* (Doyle 1998). Bryophytes are considered a sister group to Angiosperms, sharing a common land plant ancestor (Harrison 2017; Mishler and Churchill 1985). By comparing their physiology, insights into shared and distinct strategies for colonization and survival on land can be gained, as well as understanding processes present in the common land plant ancestor. Certain physiological processes, such as apical stem cells, are found in both groups, whereas some functions have been lost or modified in one group (Bowman et al 2016; de Sousa et al. 2019). For example, stomata evolved in most vascular plants and were apparently lost in liverworts (Harris et al. 2020). Throughout evolution, gene regulatory networks have undergone various changes which can lead to consequences such as neofunctionalization or subfunctionalization. Therefore, studying regulatory networks in a narrow range of plant species may result in an incomplete understanding of land plant evolution and diversity.

*Marchantia polymorpha* is an emerging model for studying land plant adaptation. *M. polymorpha* belongs to the liverwort division Marchantiopsida, part of the Bryophyte

subfamily, which also includes leafy mosses (Bryopsida) and hornworts (Antherocerolesida). Several characteristics make *M. polymorpha* an excellent model for molecular biology. Genetically, the *M. polymorpha* genome is relatively small, approximately 280 Mbp, comprising of 19,421 genes organized into 9 chromosomes, including sex chromosomes—U-chromosome for females and V-chromosome for males, rendering *Marchantia* a dioecious plant (Bowman et al. 2017; Montgomery et al. 2020; Stotler and Bischler 1999). Furthermore, the *M. polymorpha* genome has undergone minimal duplications and exhibits low redundancy, which is uncommon in many angiosperms. Many genes in *Marchantia* exist as single paralogues, making it a suitable model for reverse genetic studies. For instance, the *Marchantia* genome encodes only 394 transcription factors and utilizes a minimalistic approach in many pathways and regulations compared to angiosperms (Bowman et al. 2017; Panchy et al. 2016). Thus, processes and gene families that are often very complicated to understand in angiosperms due to their complexity and/or redundancy can be more easily understood in *Marchantia*.

The Nascent polypeptide associated complex (NAC) is a family of proteins reported to play a role in the *de novo* folding of nascent polypeptides in yeast (Wiedmann et al. 1994). The family consists of the subunits NAC $\alpha$  and NAC $\beta$  (NAC $\beta$  is also sometimes referred to as basic transcription factor 3 BTF3) and is present in all Eukaryotes. NAC $\alpha$  homologues were found in Archaea genomes but there is no evidence of NAC in prokaryotes since there its role is most likely compensated by the unrelated chaperone Trigger factor (Deuerling et al. 1999). In animals, the presence of NAC seems to be vital, as knockout of its  $\beta$  subunits leads to embryonic lethality. Such a lethality is possibly caused by the mistargeting of nascent proteins to endoplasmic reticulum, as detailed mechanism of NAC steric competition with the SRP particle at the ribosomal exit site has been recently described through mathematical modelling (Hsieh et al. 2020). Here, the authors suggest that, besides the nascent protein folding, NAC's presence determines the specificity of protein targeting to endoplasmic reticulum.

In plants, most studies have focused on the  $\beta$  subunits. A reduced expression of NAC $\beta$  leads to various phenotypic defects. In *Arabidopsis thaliana*, NAC $\beta$  role in cold stress tolerance was reported (S. Yang et al. 2018). In another study of *A. thaliana*, the knockdown of  $\beta$  subunits resulted in plants with slow development, pale leaves, and male gametophytic defects, including slowly growing pollen tubes with low germination rate and ovule targeting efficiency, resulting in a dramatically limited seed set (Fíla et al. 2020). Interestingly, NAC $\beta$  subunits shows a high level of protein sequence conservation among eukaryotes (Kogan and Gvozdev 2014). Previously published data suggest NAC connection to the cytoplasmic chaperone system, supporting its conservation in function as a chaperone similarly to animals (Fíla et al. 2020). However, the detailed mechanism of molecular function has not been described in plants yet.

In *Arabidopsis thaliana*, NAC shows a high, housekeeping expression in all developmental stages (Fíla et al. 2020). The number of individual subunits in the *A. thaliana* genome is quite expanded, with 5 NAC $\alpha$  and 2 NAC $\beta$  subunits. *M. polymorpha* expression data from Evorepro database (<https://evorepro.sbs.ntu.edu.sg>) align with *Arabidopsis*, suggesting a housekeeping role for NAC in *Marchantia* as well. Notably, there is only one gene encoding each subunit in the *Marchantia* genome, making it a perfect candidate to study NAC

function. Possibly, the highly expanded chaperone system in *A. thaliana* could also be simplified in *M. polymorpha*, facilitating a more detailed understanding of NAC function.

We addressed the question whether the function of NAC subunits is conserved between *A. thaliana* and *M. polymorpha* with a complementation study of the *nacβ1/nacβ2* mutant, showing that a functional NACβ subunit from *M. polymorpha* can reverse the defective phenotype upon introduction to the mutant plant. We analysed the relationships between NAC subunits of *A. thaliana* and *M. polymorpha* by addressing their capacity to form heterodimers in and between the two species in order to broaden our understanding of their potential subfunctionalisation in Arabidopsis and the conservation of their binding capacity across the two species. Lastly, we analysed the distribution of the subunits in *M. polymorpha* by reporting their cellular localization and tissue expression. We attempted to acquire a knockout plant line in *M. polymorpha* but the complete mutation of even a single subunit appears to be lethal. As an ongoing experiment, we will try to acquire a conditional single mutant of the α and β subunit to prove its vital role in plant development.

## Materials and Methods

### Bioinformatic Analysis of NAC Sequence and Structure

Protein sequences for selected organisms were retrieved from the online Phytozome database. The secondary and tertiary structures of *M. polymorpha* and *A. thaliana* NAC subunits were obtained from UniProt database and analysed in iCn3D Structure Viewer (v. 3.28.4, <https://structure.ncbi.nlm.nih.gov/icn3d>). These protein sequences were aligned and the phylogeny tree was rendered using the Clustal Omega tool (Clustalo v. 1.2.4). The tree was visualized and graphically edited in the iTOL Tree of Life program. EMBOSS Needle was used for pair-wise alignment analysis. For Logo analysis, the online tool WebLogo 3 (version 3.7.12, Crooks et al. 2004; Thomas D Schneider and R Michael Stephens 1990) was used.

### DNA Extraction, Genotyping, Sequencing

Genomic DNA extraction from transgenic Marchantia lines thallus and gemmae was performed using the standard CTAB procedure (Porebski, Bailey, and Baum 1997). The extracted DNA was used for genotyping by PCR. PCR reactions included the following components: 0.4 μM of each primer (Sigma-Aldrich, Haverhill, United Kingdom), 0.2 mM dNTPs (Promega, Madison, USA), 3 μL DNA template (approximately 2000 ng), and 1U Taq DNA polymerase (Merck, Brno, Czech Republic) in a 1× reaction Merck buffer. The PCR cycle was as follows: 94 °C/2 min, followed by 40 cycles of 94 °C/30 s, 55 °C/30 s, 72 °C/1 min, with a final extension at 72 °C/5 min. The reaction mixture was combined with 6× Loading Dye (Thermo Fisher Scientific, Waltham, USA) and loaded on a 1% (w/v) agarose gel with ethidium bromide in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, mM EDTA). A GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) was used as a molecular marker, and gels were visualized using the Gel Documentation System G-Box (Syngene, Cambridge, United Kingdom).

For the detection of small nucleotide mutations in the CRISPR-Cas9 lines, the amplified DNA was purified by the QIAGEN PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, and then sent for Sanger sequencing (Eurofins Genomics

Europe Shared Services GmbH, Ebersberg, Germany). The resulting sequences were analysed in SnapGene software (GSL Biotech LLC, San Diego, California).

### **Protein Extraction, SDS-PAGE, and Western Blot**

The soluble fraction of total proteins was extracted from 50 mg of snap-frozen *M. polymorpha* thallus, ground by a pestle in a mortar. The extraction buffer (10 mM Tris-Cl pH 7.5, 14% glycerol, 150 mM NaCl, 0.5 mM EDTA, 0.5% IGEPAL® CA-630, 7 mM PMSF, EDTA-free Protease Inhibitor Cocktail (11873580001, Roche, Basel, Switzerland) was added to a total volume of 1 ml and the sample was transferred to Eppendorf tubes. Cellular debris was pelleted by centrifugation (20,000×g, 20 min, 15 °C), and the supernatant was transferred to fresh tubes. Samples mixed with sample buffer (0.15 M Tris-Cl pH 6.8, 30 % (v/v) glycerol, 6% (w/v) SDS, 0.15 M DTT, 0.03% bromophenol blue (1% in 50% isopropanol)) were loaded onto SDS-PAGE gels (resolving gel 11.25 %T, 2.6 %C, stacking gel 5 %T, 2.6 %C) and run at 75V for 30 minutes and 150V for another 60 minutes. Semi-dry Western blotting was employed for protein transfer. Protein detection was performed as described previously (Billey et al. 2021) with slight adjustments: proteins were detected using an Anti-RFP primary antibody from mouse (ab125244, Abcam, Cambridge, Great Britain; 1,000x diluted) after overnight staining, followed by 30 minutes with a goat anti-mouse IgG (H+L) secondary antibody conjugated with alkaline phosphatase (31320, Thermo Fisher Scientific, Waltham, Massachusetts, USA; 3,000x diluted). The alkaline phosphatase signal was detected in AP buffer (100 mM Tris-Cl pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>) containing BCIP (0.5 mM) and NBT (0.4 mM). A prestained protein marker (PageRuler™ Prestained Protein Ladder, cat. number: 26616, Thermo Scientific™) was used to determine the molecular weight of the proteins of interest.

### **RNA Extraction, cDNA Preparation, and RT-qPCR**

RNA was isolated from approximately 4-week-old *M. polymorpha* plants by the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The RNA was then treated with the DNA-free DNA Removal Kit (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's instructions. RNA from three independent CRISPR-Cas9 lines, each harbouring tri-nucleotide deletions, was used for subunit analysis in both tak-1 and tak-2 plants.

The qPCR samples were prepared with GoTaq qPCR Master Mix (Promega, Madison, WI, USA) using the primer sets from the Supplementary file 1 and analysed on a LightCycler 480 Instrument (Roche, Basel, Switzerland) according to the manufacturer's guidelines. The LightCycler 480 Software version 1.5 was used to obtain quality control and Ct values, and comparisons of  $\alpha$  or  $\beta$  subunit amplification between CRISPR-Cas9 lines and WT plants were made. A Student's t-test was employed for statistical significance assessment.

### **Cloning, *Marchantia polymorpha***

The following vectors for the Gateway cloning system were kindly provided by Takayuki Kohchi (Sugano et al. 2018): pmpGWB128 for protein localization with CaMV 35S promoter and TagRFP coding sequence (Addgene plasmid #68582), pmpGWB104 for promoter activity with GUS coding sequence (Addgene plasmid #68558), pmpGE010 for CRISPR/Cas9 genome editing (Addgene plasmid #71536), pmpGE-En03 for gRNA expression cassette construction

(Addgene plasmid #71535), and pmpGWB337 for conditional knock-out (Addgene plasmid #68061).

The gRNAs were designed using an online CasFinder tool (<https://marchantia.info/tools/casfinder/>, Aach, Mali, and Church 2014). The selected gRNA sequences targeting both exons and introns were designed and cloned as described in (Sugano et al. 2018) using the pmpGE\_En03 as entry vector, whereas pmpGE010 and pmpGWB337 were used as destination vectors for stable and inducible silencing, respectively (Supplementary file 1).

The coding sequences, including the 3' UTR region and promoter sequences of NAC $\beta$  and NAC $\alpha$  subunits, were amplified from cDNA using the primer sets provided in the Supplementary file 1. A two-step PCR with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) was conducted following the manufacturer's instructions. The amplicons were cloned into Gateway pDONR221 entry vectors (Thermo Fisher Scientific) using BP Clonase II enzyme mix. LR Clonase II Plus was used to create pBiFCt-2in1-CC expression constructs for Bimolecular Fluorescence Complementation (BiFC) tests of protein-protein interactions (Grefen and Blatt 2012). LR Clonase II was used to prepare pDEST22<sup>TM</sup> and pDEST32<sup>TM</sup> expression clones for Yeast Two-Hybrid (Y2H) and other expression clones for *M. polymorpha*, except for pmpGWB337. For pmpGWB expression clones with MpNAC $\alpha$  or MpNAC $\beta$  CDS sequences, the In-Fusion Snap Assembly cloning kit (Takara Bio USA) was utilized. The nucleotide sequences of all resulting clones were verified by Sanger sequencing. The homemade competent cells of *Agrobacterium tumefaciens* strain GV3101 were transformed by electroporation with the prepared expression clones and used further for plant transformation as described previously (Wiese et al. 2021).

### **Cloning, *Arabidopsis thaliana***

cDNA from *A. thaliana* leaves was acquired by the same procedure as described above and used as template for amplification of the coding sequences, coding sequence including the 3'UTR or 5'UTR region both NAC $\beta$  and all NAC $\alpha$  subunits for interaction analyses and promoter sequences of both NAC $\beta$  subunits. Used primers are stated in Supplementary file 1.

### **Plant Transformation**

For the creation of Tak-1 and Tak-2 *Marchantia polymorpha* transgenic lines, we followed the Gemmae AgarTrap protocol with slight modifications as described previously (Tsuboyama et al. 2018; Tsuboyama and Kodama 2014). Briefly, sterile cultures of *M. polymorpha* were grown on B5 Gamborg solid medium supplemented with 1% agar and 1% sucrose at pH 5.7. Approximately 30 gemmae from 1-month-old plants were placed onto a sterile Petri dish containing the same B5 Gamborg medium and left to grow for 2 days under continuous light. Prepared *Agrobacterium tumefaciens* cultures containing expression vectors grown on solid LB medium at 28°C for one day were skimmed from the medium surface and resuspended in transformation buffer (150  $\mu$ M acetosyringone, 10 mM MgCl<sub>2</sub>, 10 mM MES-NaOH, pH 5.7) to an OD<sub>600</sub> of 0.5. One millilitre of this transformation buffer with *A. tumefaciens* was added to each Petri dish containing gemmae and spread evenly. After one minute, the excess of *Agrobacterium* culture was removed, and dishes were sealed with micropore tape. The gemmae were co-cultivated with *Agrobacterium*

*tumefaciens* for 5 days in the dark at 22°C. Upon the co-cultivation, the plates were washed with sterile water and treated with selection buffer (10 mM MgCl<sub>2</sub>, 10 mM MES-NaOH, pH 5.7) containing 100 µg.ml<sup>-1</sup> hygromycin and 1 mg.ml<sup>-1</sup> cefotaxime. The plates were sealed with parafilm and incubated for 2 weeks under continuous light at 22°C. Transgenic plants were selected based on antibiotic resistance and transferred to new B5 Gamborg plates for further analysis.

### **Plant Cultivation**

*Marchantia polymorpha* Tak-1 and Tak-2 plants were cultured aseptically on half-strength Gamborg's B5 medium without vitamins, including 1% agar and 1% sucrose (Gamborg, Miller, and Ojima 1968). The plants were maintained at 22°C under continuous light. For soil growth, plants were transferred to Jiffy peat pellets and grown under identical conditions. Far-red light (Cosmo LED 20W, 24V, Inerared, Cosmorrow Secret Jardin, Manage, Belgium) was used to induce reproductive growth.

### ***in vitro* pollen tube growth**

The methodology for *in vitro* pollen tube growth followed the protocol established by Boavida and McCormick (2007). The growth medium (0.01% (w/v) H<sub>3</sub>BO<sub>3</sub>, 1 mM CaCl<sub>2</sub>, 1 mM KCl, 5% (w/v) sucrose, pH 7.5, 1% (w/v) low-melting agarose) was dispensed onto glass slides and left to solidify. Pollen grains from opened flowers were sown onto the set medium and then placed into small boxes in the cultivation rooms. Following a 6-hour incubation, pollen tube development was inspected using an inverted epifluorescent microscope (Nikon TE2000E, Tokyo, Japan). The images were then processed and analysed by Nikon's NIS Elements (Nikon, Tokyo, Japan) software and Fiji (Schindelin et al. 2012).

### **Microscopy Observation of NAC Subcellular Localization**

For nuclear visualization in 1-week-old gemmalings, the plants were stained by DAPI (4',6-diamidin-2-fenylindol, 1µg.mL<sup>-1</sup>) for 10 minutes and mounted them on slides in 50% glycerol. Confocal microscopy observations were performed by a Zeiss LSM880 with an Airyscan detector (Carl Zeiss AG, Jena, Germany). We used 561nm for RFP excitation and 604nm for emission. DAPI was excited with 353 nm and maximum emission was measured at 464 nm and chloroplast autofluorescence was visualized at 633 nm excitation and 650 nm emission. Images were processed in Fiji (Schindelin et al. 2012) and Adobe Photoshop CS6 (Adobe, San Jose, California, USA).

### **GUS Assay and Promoter Activity Analysis**

Various developmental stages of *Marchantia polymorpha*, such as gemmae, 4-week-old thalli, antheridiophores, archegoniophores, and 48-hour-old sporophytes, were stained in GUS staining solution and incubated overnight at 37°C in the dark. The samples were subsequently washed, cleared with ethanol series, and mounted for imaging under a LEICA stereomicroscope (Leica, Wetzlar, Germany).

### **Yeast-Two Hybrid (Y2H)**

The ProQuest Two-Hybrid System was employed for Y2H assays, performed according to the manufacturer's instructions. The β subunit of *M. polymorpha*, cloned into the pDEST22 vector, was co-transformed with each of the five *A. thaliana* α-subunits cloned into pDEST32, and vice versa. The strength of interactions was determined based on growth on



selective media and X-gal assays. The assay was replicated with three independent transformations, creating three independent biological replicas.

### **Transient Expression and Bimolecular Fluorescence Complementation (BiFC)**

We utilized the Multisite Gateway system to create pBiFCt-2in1 constructs (Grefen and Blatt 2012). *Agrobacterium tumefaciens* cells harboring selected expression clones were grown overnight, harvested, and prepared for leaf infiltration in *Nicotiana benthamiana*. Two days post-infiltration, fluorescence in leaves was examined using a Zeiss LSM880 laser scanning microscope (Carl Zeiss AG, Oberkochen, Germany) with excitation wavelengths appropriate for GFP/YFP and mRFP. Images were processed by Fiji/ImageJ (Schindelin et al. 2012) and Adobe Photoshop CS6 (Adobe, San Jose, California, USA).

## **Results**

### **Sequence conservation of NAC subunits in *A. thaliana* and *M. polymorpha***

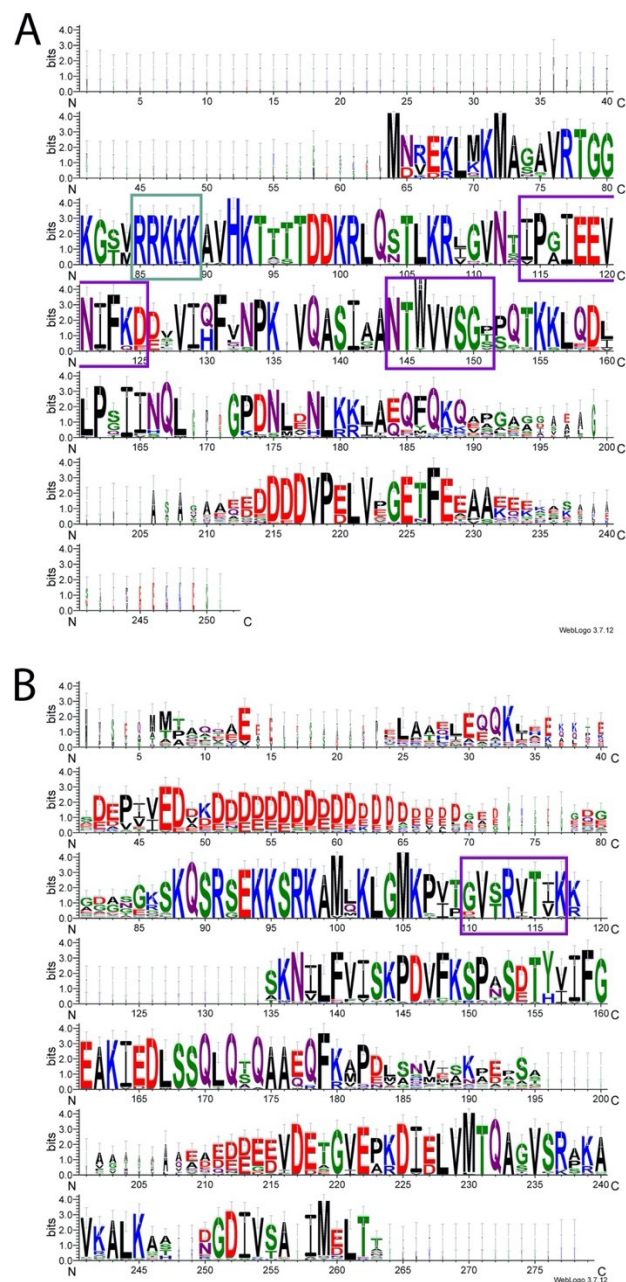
The amino acid sequence of NAC proteins is highly conserved across eukaryotic species, displaying significant similarity between plants and animals. For example, the alignment of 217 amino acids from the NAC $\beta$ 1 subunits reveals 41% identity and 53.9% similarity with the human BTF3 protein sequence (Supplementary File 2). The *Methanococcus aeolicus*  $\alpha$  subunit also shares a relatively high level of similarity with the NAC $\alpha$  subunits of *Arabidopsis thaliana*, ranging from 14.8 to 31.6% identity among the five NAC $\alpha$ s, and 27.8 to 33.5% similarity (Supplementary File 2).

The secondary structure of the  $\beta$  subunits from selected plants, human, and yeast shows conserved features (Fig. 1A). In plants, starting approximately 50 residues from the beginning, there are 6  $\beta$ -sheets. Upstream of these  $\beta$  sheets, two  $\alpha$  helices are conserved across all plant species except for MpNAC $\beta$ , where only one  $\alpha$  helix is identified. Downstream of the  $\beta$  sheets, both the primary and secondary sequences tend to be less conserved, with the number of  $\alpha$  helices varying between 2 and 4. The NAC domain consists of 116 or 115 residues in most analysed plants, starting at the 5th amino acid from the N-terminus of the protein.

In certain species, including human, yeast, *Chlamydomonas reinhardtii*, *Physcomitrium patens*, *Selaginella moellendorffii*, and *Solanum lycopersicum*, the heterodimerization site and the ribosomal binding site was predicted *in silico* (Supplementary Fig. 1). The heterodimerization sequence, predominantly present in the  $\beta$ -sheet structures, shares two long conserved motifs: IPGIEEVNIFKD/E and NTW/VVV/ISG-S in plants (Fig. 1A). The ribosomal binding site is situated near the N-terminus of the protein, around the 22nd residue, with the conserved motif: RRKH/KK in both plants and human. The RRKKK motif is present in all  $\beta$  subunits of the plants tested in this analysis except for two  $\beta$  subunits of *P. patens* (Pp3c9\_3350, Pp3c15\_17590), where the RRKHK motif is present (Fig. 1A). The dimerization motif PGIEEVNIFK/QD/E is found in *P. patens*, *C. reinhardtii*, *Selaginella*, *M. polymorpha*, and 4 subunits of *Z. mays* and 2 subunits of *O. sativa* (with an exception in Pp3c15\_17590 where I is substituted with V, resulting in PGIEEVNVFK). Conversely, the motif PAIEEVNIFK is present in *A. thaliana*, *S. lycopersicum*, *A. trichopoda*, *T. plicata* and one subunit of *O. sativa* and one subunit of *Z. mays*. The second conserved part of the

putative dimerization site NTW/VVV/ISG is conserved in all plants except for the alga *C. reinhardtii*. In *C. reinhardtii* the motif NTWVVSG and NTYVISG is present (Fig. 1A).

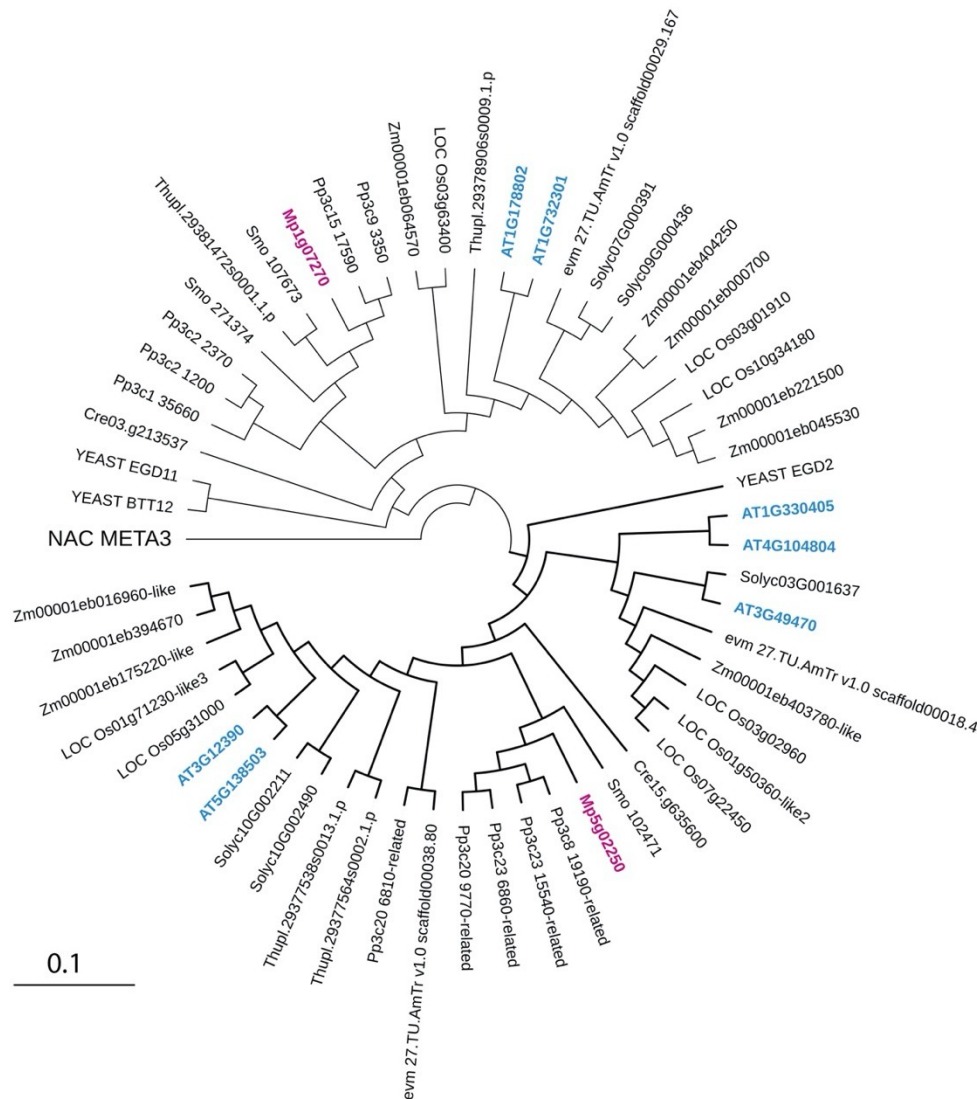
The analysis of the secondary structure of NAC $\alpha$  proteins also shows a high level of conservation (Supplementary Fig. 1). Most plant NAC $\alpha$ s, as well as yeast BTT1 (NAC $\beta$  subunit) and human NAC  $\alpha$  contain 6  $\beta$  sheets. At the N-terminus, one or two  $\alpha$  helices are described, while the C-terminus is composed of 4  $\alpha$  helices in plants. The first heterodimerisation site is located in the first and second  $\beta$  sheets with the consensus sequence tVoRV/ITV/IK, and the second between the fifth and sixth  $\beta$  sheets, but overall, the motifs seem less conserved than in the case of NAC $\beta$  (Fig. 1B). It is important to note that the  $\beta$ -sheet structure and the adjacent  $\alpha$  helix are also the regions predicted with very high fidelity by AlphaFold 2 (pLDDT > 90).



**Figure 1** LOGO analyses of NAC $\alpha$  and NAC $\beta$  subunits in selected plant species show a high level of sequence conservation. A. LOGO analysis of sequence conservation of the NAC $\beta$  subunit. The green rectangle highlights the conserved region of the

ribosomal binding site and the magenta rectangles highlight the heterodimerisation sites. B. LOGO analysis of sequence conservation of the NAC $\alpha$  subunit. The magenta rectangle highlights the conserved region of the first heterodimerization site.

In the basal land plant *M. polymorpha*, there is only one representative of each NAC subunit in its genome. In contrast, numerous angiosperms, as well as the moss *P. patens*, exhibit an expansion in the number of subunits. The genome of *A. thaliana* encodes for two NAC $\beta$  subunits and five NAC $\alpha$  subunits. To elucidate the phylogenetic relationships among *A. thaliana* NAC subunits, a phylogenetic tree was constructed that includes all NAC subunits available on Phytozome database, namely from the following species *M. polymorpha*, *P. patens*, *O. sativa*, *Z. mays*, *S. moellendorffii*, *A. trichopoda*, *S. lycopersicum*, *Ch. reinhardtii*, *T. plicata*, and *A. thaliana* to represent algae, liverworts, mosses, gymnosperms, and angiosperms including both monocots, dicots and lycophytes (Fig. 2). The tree was rooted with an Archaeal NAC $\alpha$  of *Methanococcus aeolicus* as an ancestral sequence. Yeast NAC subunits were also included to illustrate the divergence of plant branches.



**Figure 2** Phylogenetic tree of NAC subunits from selected plant species, including subunits from *S. cerevisiae* and rooted with the NAC $\alpha$  subunit of the archaeon *Methanococcus aeolicus* (NAC META3). Branches in bold indicate the NAC $\beta$  subunit branch. *A. thaliana* and *M. polymorpha* NAC subunits are highlighted in blue and magenta, respectively. The scale bar indicates the branch length. The names of the genes encoding the analysed proteins are corresponding to following organisms: AT = *A. thaliana*, Mp = *M. polymorpha*, Pp = *P. patens*, Os = *O. sativa*, Zm = *Z. mays*, Smo = *S. moellendorffii*, evm27.TU.AmTr = *A. trichopoda*, Solyc = *S. lycopersicum*, Cre = *Ch. reinhardtii*, Thupl = *T. plicata*.

The subunits were distinctly divided into two clades NAC $\beta$  and NAC $\alpha$ , indicating a significant evolutionary divergence between the two subunits. Focusing on the NAC $\beta$  branch, the yeast NAC $\beta$ s form a separate outgroup. Further divisions within the plant NAC $\beta$  branch show a separation based on species lineage. The first subgroup with NAC $\beta$  from the alga *C. reinhardtii* represents the most basal lineage among the plant species, suggesting that this protein may retain the most ancient characteristics of plant NAC $\beta$ s. Furthermore, the NAC $\beta$  subunits clustered into a clade that included all subunits from *P. patens*, *S. moellendorffii*, *M. polymorpha*, and one subunit from *T. plicata*. This branch likely represents an evolutionary path separate from the angiosperms. This indicates that although they share a common ancestor with angiosperm NAC $\beta$ s, they have diverged sufficiently and may have unique functions or regulatory mechanisms. Furthermore, the five NAC $\beta$  subunits of *P. patens* grouped into two clades; Pp3c15\_17590 and Pp3c9\_3350 were more closely related

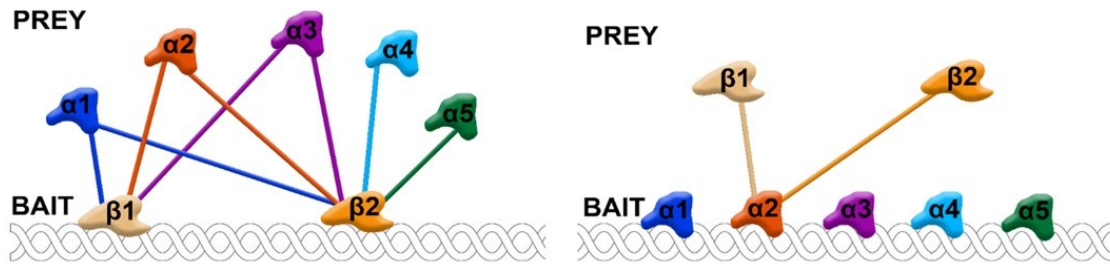
to the subunits from the other three species, while Pp3c2\_1200, Pp3c1\_35660, and Pp3c2\_2370 formed a distinct clade, indicating their divergent evolution in *P. patens*. The second subgroup includes monocots, dicots and *A. trichopoda*. Both *A. thaliana* NAC $\beta$  subunits were most closely related to each other than to other NAC $\beta$ s, suggesting that they evolved after angiosperm divergence.

The phylogenetic relationships of the *A. thaliana* NAC $\alpha$  subunits are also noteworthy. The plant NAC $\alpha$  subunits were divided between two branches. The first one included the NAC $\alpha$  subunits of *C. reinhardtii* suggesting a branch more similar to the algae subunits. This clade includes all *P. patens* NAC $\alpha$ s and the *M. polymorpha* NAC $\alpha$ . Notably, some monocot and dicot subunits also fall within this group. However, further subdivisions indicate that some NAC $\alpha$  subunits from the monocots *Z. mays* and *O. sativa* are grouped with subunits from dicots and other plants. This mixed grouping in the sister branches may imply multiple duplication and divergence events throughout the evolution of these subunits undergoing a convergent evolution with dicots in some cases, showing non-linear divergence events in NAC $\alpha$ s. The presence of *A. trichopoda*, often considered an early-diverging lineage of flowering plants, suggests that this branch represents an ancient lineage of NAC $\alpha$  subunits. For *A. thaliana*, NAC $\alpha$ 3 and NAC $\alpha$ 1 cluster within this clade, whereas NAC $\alpha$ 2, NAC $\alpha$ 5, and NAC $\alpha$ 4 belong to the sister branch that contains angiosperms exclusively. This suggests that NAC $\alpha$ 3 and NAC $\alpha$ 1 may be more closely related than NAC $\alpha$ 2, NAC $\alpha$ 5, and NAC $\alpha$ 4, potentially indicating their different levels of subfunctionalization or dimerization capabilities.

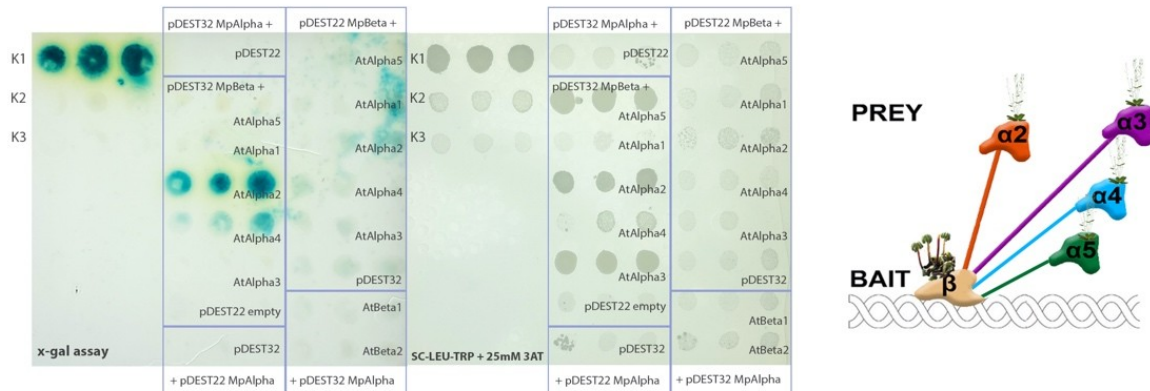
Consequently, protein-protein interaction studies were performed to determine the dimerization capabilities of the NAC subunits (Klodová 2019). In summary, the bimolecular fluorescence complementation (BiFC) revealed no preference for heterodimerization; all NAC $\alpha$ s were capable to dimerize with any NAC $\beta$  with cytoplasmic localization of the heterodimers. However, yeast two-hybrid (Y2H) assays demonstrated some specificity; NAC $\alpha$ 4 and NAC $\alpha$ 5 showed a preference for binding, dimerizing exclusively with NAC $\beta$ 2. Other subunits were capable of forming dimers with both NAC $\beta$ 1 and NAC $\beta$ 2 (Fig. 3A).

To test whether AtNAC and MpNAC subunits can form heterodimers between each other, constructs for Y2H were prepared. The interaction was tested between both MpNAC subunits and all 7 AtNAC subunits. The setting was carried out in both directions i.e. MpNAC cloned with activation domain and AthNAC with DNA binding domain and *vice versa*, resulting in 14 tested interactions. The presence and strength of the interaction was determined using the x-gal reporter gene in x-gal assay and by auxotrophic growth of the transformed yeast on selective medias. Heterodimerisation was observed on x-gal assay reporting strong interaction for MpNAC $\beta$  cloned into the pDEST32 with NAC $\alpha$ 2, and NAC $\alpha$ 4 cloned into pDEST22. A weaker interaction on the 25mM and 50 mM 3AT selection plates was further observed between MpNAC $\beta$  and NAC $\alpha$ 5 or NAC $\alpha$ 2 (Fig. 3B). No auto-activity causing false positivity was reported for either NAC subunit.

A



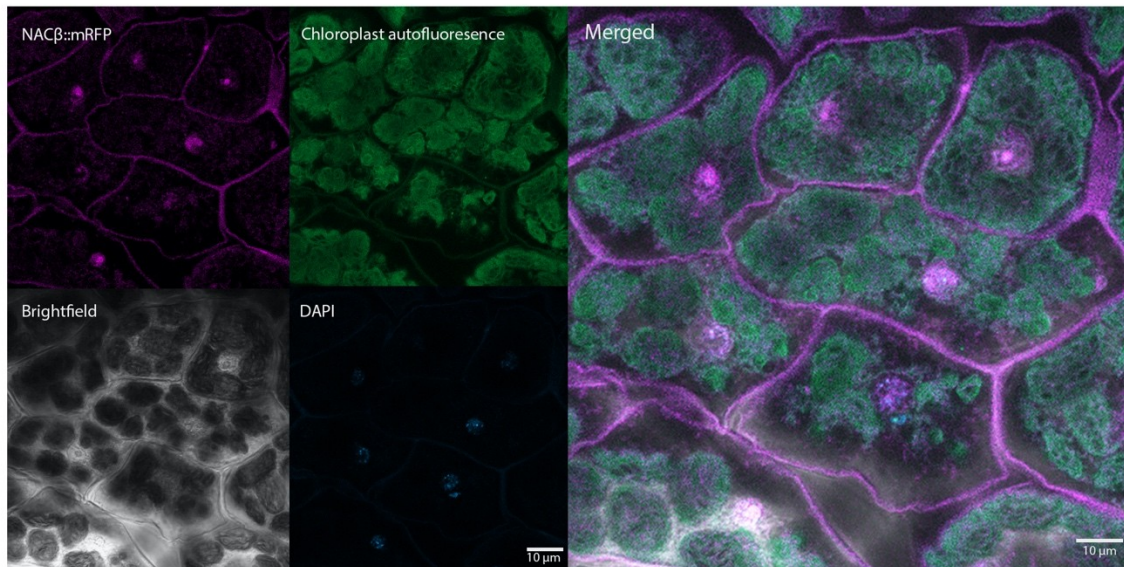
B



**Figure 3** analysis of heterodimerization interactions among NAC subunits. A. Yeast two-hybrid (Y2H) assay results demonstrated the binding affinities between NAC $\alpha$  and NAC $\beta$  subunits of *A. thaliana* (Adapted from Klodová, 2019). B. Y2H assay results showing the capacity for heterodimer formation between the NAC subunits from *A. thaliana* and *M. polymorpha*. Positive interactions were observed when MpNAC $\beta$  was used as bait in combination with AtNAC $\alpha$ 2, AtNAC $\alpha$ 3, AtNAC $\alpha$ 4, and AtNAC $\alpha$ 5. Control K1 represents pEXP<sup>TM</sup> 32/Krev1 paired with pEXP<sup>TM</sup> 22/RalGDS-wt, indicating a strong positive interaction. Control K2 represents pEXP<sup>TM</sup> 32/Krev1 paired with pEXP<sup>TM</sup> 22/RalGDS-m1, indicating a weak positive interaction, Control K3 involves pEXP<sup>TM</sup> 32/Krev1 paired with pEXP<sup>TM</sup> 22/RalGDS-m2, serving as a negative interaction control. The vectors pDEST22 and pDEST32 are shown as empty plasmid controls to check for auto-activation of the proteins under study.

**NAC subunits of *M. polymorpha* show cytoplasmic and nuclear localisation and house-keeping promoter activity (ongoing)**

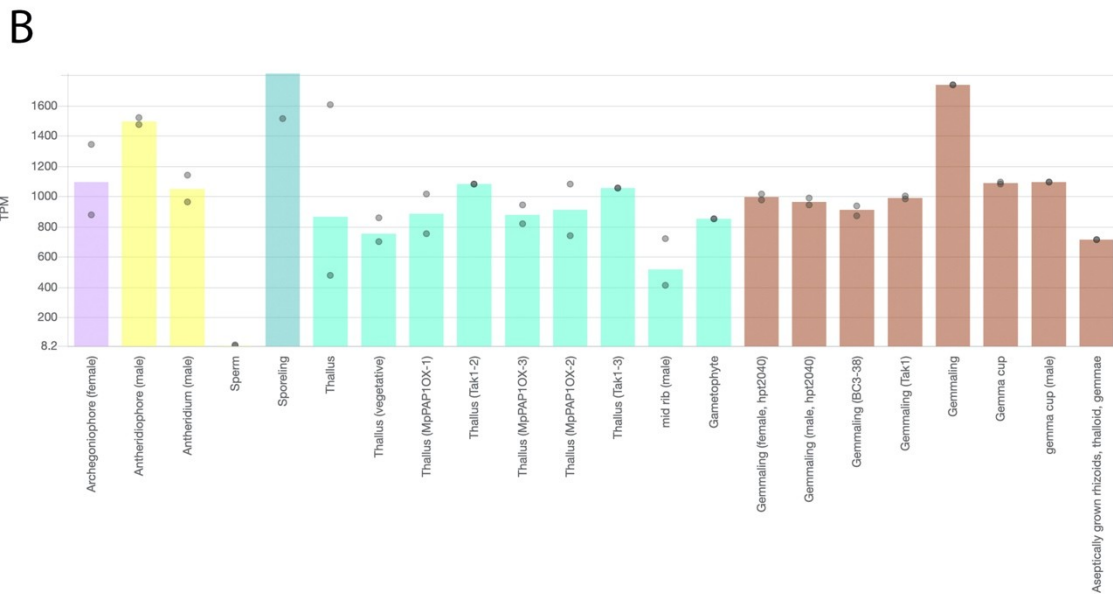
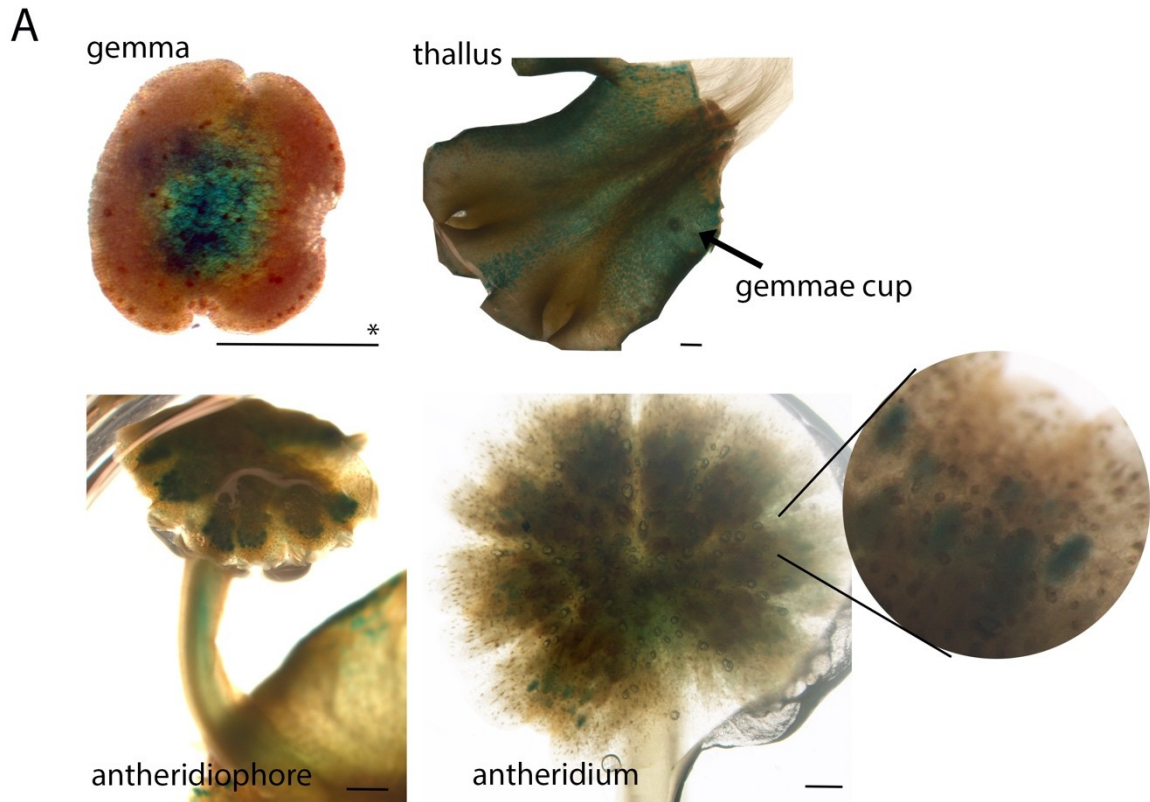
To determine the subcellular localization of NAC $\beta$  and NAC $\alpha$  subunits in *M. polymorpha*, plant lines containing the NAC sequence under the CaMV 35S promoter with a C-terminal



tRFP tag were prepared. Subcellular localization of NAC subunits was observed under a confocal microscope. DAPI staining provided nuclear visualization, while chloroplast autofluorescence helped discern the signal localization. The localisation was observed in young gemmalings. NAC $\beta$  subunit predominantly localized to the cytoplasm, with occasional presence in the nucleus (Fig. 4). The data of NAC $\alpha$  are only preliminary but the photos will be acquired after

**Figure 4** Localisation of NAC $\beta$  subunit of *M. polymorpha*. The NAC $\beta$  shows mainly cytoplasmic and nuclear localisation. Scale bar represents 10  $\mu$ l.

the prepared line reaches second gemmaling generation, but a similar localisation like NAC $\beta$  is expected. To dissect the promoter activity of NAC subunits, lines with approximately 1000 bp promoter region of NAC $\beta$  and NAC $\alpha$  with downstream glucuronidase marker gene were prepared. The GUS assay was performed to observe the promoter activity in various tissues and developmental stages. In case of NAC $\beta$ , the promoter was active in 3 day old gemmaling, 3 week old thallus including rhizoids and gemma cup, antheridiophore and antheridium (Fig. 5A). Then, NAC $\alpha$  promoter was active in archegoniophore, archegonium. The sporeling data will be acquired after the remaining transformed lines will reach reproductive growth. Overall, the activity of NAC $\beta$  was reported in all measured stages. Strong expression was visible in the mature gemmaling. In thallus, the expression was most prominent in the central part including the gemma cup and the apical notch with the dividing and differentiating zone. In the male reproductive organs, light expression was visible in the antheridiophore stalk and in the antheridia. However, the expression seems to be limited to the antheridia present close to the edge of antheridiophore disk. These represent the more mature stages, so the NAC $\beta$  expression could differentiate during the antheridium development. The resulting data correlate with expression data available at the Connekt database (Fig. 5B). Altogether, the expression pattern of NAC subunits shows a house-keeping nature, as they are expressed in majority of the developmental stages and tissues.



**Figure 5** Analysis of the tissue-specific expression of  $NAC\beta$  from *Marchantia polymorpha*. A. Results of the glucuronidase assay illustrating  $NAC\beta$  expression in gemmalings, thalli, antheridiophores, and antheridia. The scale bar indicates 1 mm. Scale bar with asterisk indicated 0.1 mm. B. The expression profile of  $NAC\beta$  sourced from the EvoRe pro database (<https://evorepro.sbs.ntu.edu.sg/>) indicating pleiotropic expression patterns.

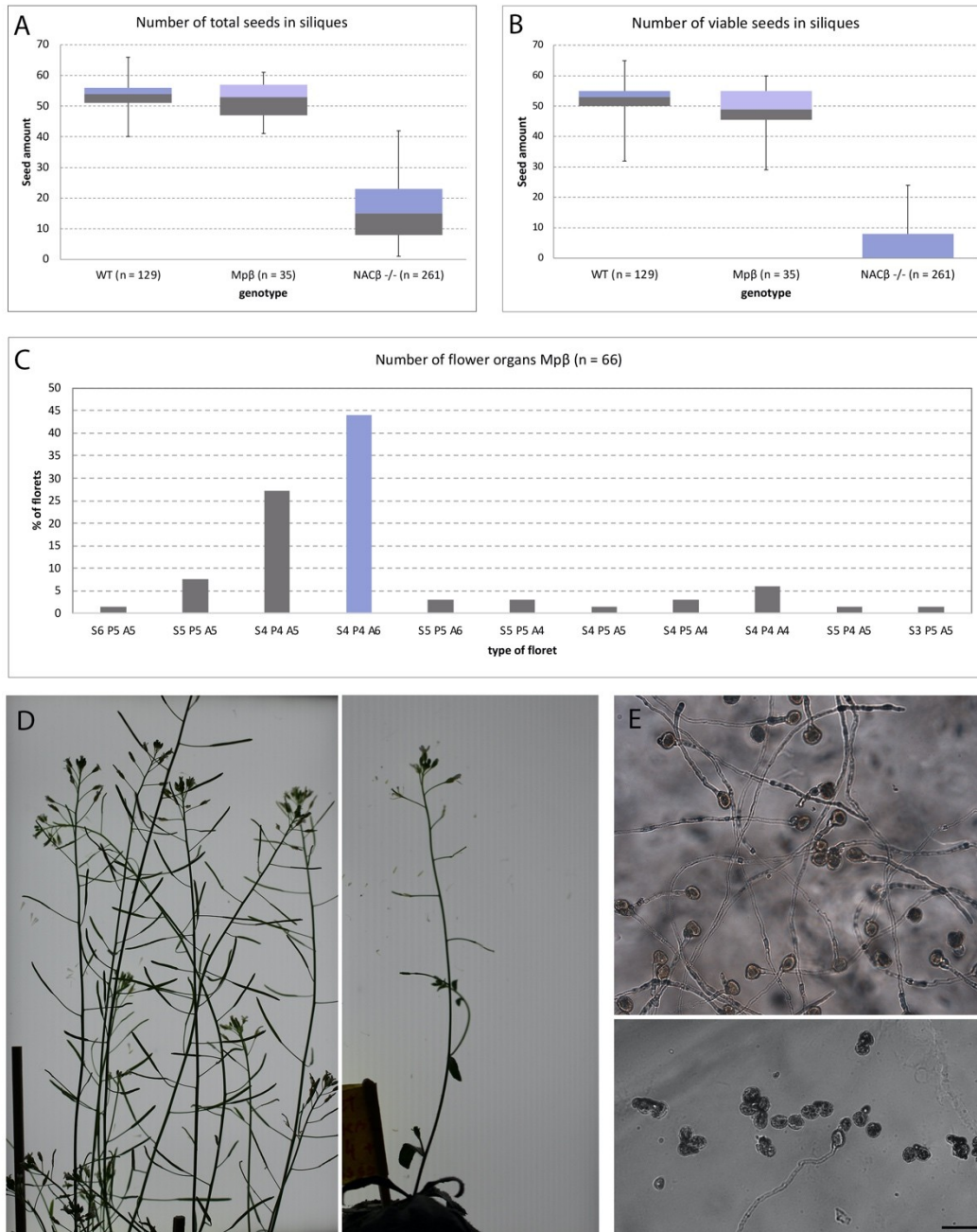
### Functional Complementation of $NAC\beta$ Subunits in *A. thaliana*

Previously characterized double homozygous *A. thaliana* mutants of both  $NAC\beta$  subunits ( $nac\beta 1/nac\beta 2$ ) exhibited various phenotypical defects, such as disturbed floral organ numbers, shorter siliques with less seeds, slow-growing pollen tubes with low germination rates, and delayed development compared to Col-0 WT plants (Fila et al. 2020). In the current study, we introduced the *M. polymorpha*  $NAC\beta$  subunit under the *A. thaliana*  $NAC\beta 1$



promoter with a C-terminal GFP tag into the *nacβ1/nacβ2* mutants to evaluate whether the *M. polymorpha* NACβ subunit could functionally complement the *A. thaliana* NACβ subunits and rescue the defective phenotype. Similar analysis was conducted using a short version of *A. thaliana* NACβ subunits, lacking the C-terminal part spanning across the whole length of the predicted NAC domain site. This resulted in a truncated transcript of NACβ1 lacking the NAC domain (noNAC line). These constructs were also used to complement the *nacβ1/nacβ2* phenotype to assess the necessity of NAC domain for functional NAC. The resulting double mutants expressing the Mp:NACβ:GFP fusion (*nacβ1/nacβ2*/Mp:NACβ:GFP) were assessed for various phenotypic traits: A) number of total seeds in siliques, B) number of viable seeds in siliques, C) number of floral organs, D) growth rate, E) *in vitro* pollen tube germination with comparisons made to previously acquired data from the *nacβ1/nacβ2* and Col-0 WT plants (Fíla et al. 2020).

Number of total seeds of siliques was calculated for 129 Col-0 WT plants, 261 *nacβ1/nacβ2* plants and 35 complemented plants. The median values were 54 for the WT, 15 for the *nacβ1/nacβ2* mutants, and 53 for the complemented line (Fig. 6A). Further, the number of viable seeds based on phenotypical observation was calculated showing 97.97%, 18.57% and 94.78% with standard deviation 4.41%, 24.10% and 6,49% in the three genotypes (Fig. 6B). In case of the *nacβ1/nacβ2* line, the predominant (25%) organ set was 5 sepals, 5 petals and 6 anthers, with only 5% of plants having the WT like organ set (4 sepals, 4 petals, 6 anthers) (Fíla et al. 2020). 47% of the plants of the complemented line showed the WT distribution with second most abundant being the 5 petals, 5 sepals, 6 anthers present in 24% of flowers (Fig. 6C). When comparing the overall habitus of the plants and the growth rate, the complemented lines did not differ from the WT plants, whereas the noNAC line retained the complete defective phenotype of the *nacβ1/nacβ2* mutant short siliques, delayed growth resulting in much smaller plants when compared to the WT or complemented plants of the same age (Fig. 6D). The germination rate of pollen tubes was also enhanced in the complemented lines (Fig. 6E). On average the germination rate of complemented pollen reached 61.3% but on the other hand only a very limited amount of pollen germinated in the noNAC complemented line (Fig. 6E).



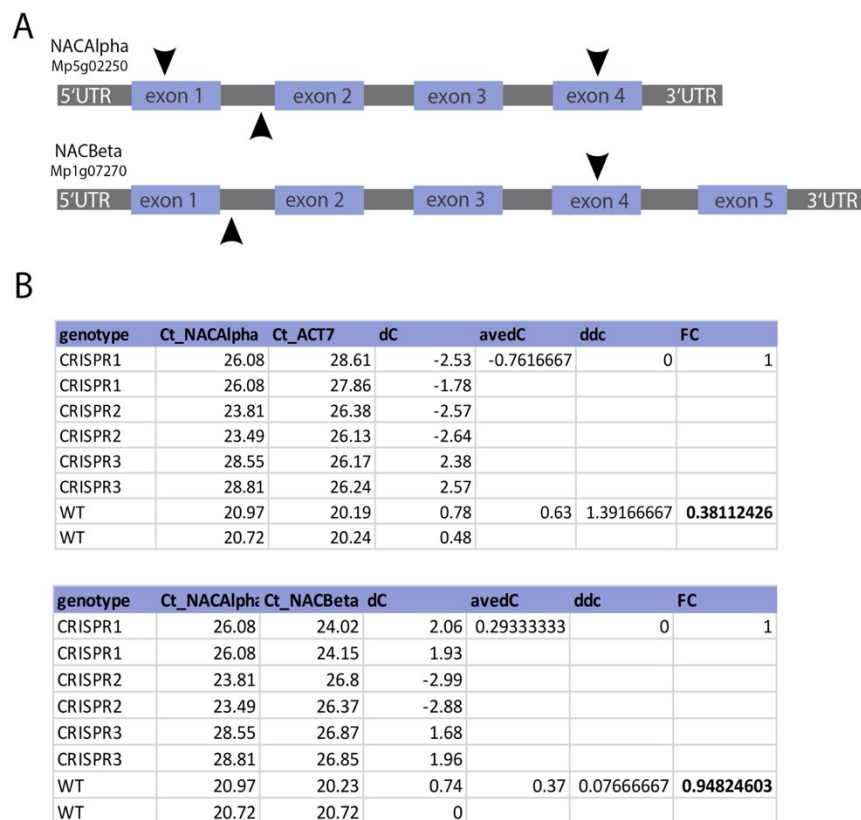
**Figure 6** Results of complementation assays indicate functional conservation of NACβ subunits between *M. polymorpha* and *A. thaliana*. **A.** Box-and-whisker plot showing the median and quartiles of the total number of seeds per silique in the Col-0 WT, Mppβ complemented, and nacβ1/nacβ2 lines. **B.** Box-and-whisker plot showing the median and quartiles of the total number of viable seeds per silique in the Col-0 WT, Mppβ complemented, and nacβ1/nacβ2 lines. **C.** Box plot showing the representation of different flower organ numbers in percentage in the flowers of NACβ complemented lines. **D.** Left - Mppβ complemented line exhibiting WT-like phenotype, including growth size and silique length. Right - noNAC complemented line exhibiting nacβ1/nacβ2 phenotype with reduced growth and short siliques. **E.** In vitro grown pollen tubes after 6 hours show a high germination rate in the Mppβ complemented line (up) and almost no germination in the noNAC complemented line (down). Scale bar = 50 μm.

### Generation of NAC knock-out lines in *M. polymorpha*

We utilized the CRISPR-Cas9 system to generate single mutants of the NACβ and NACα subunits in *Marchantia polymorpha*. Suitable targets were chosen within the first and fourth exons of NACα and the third exon of NACβ (Fig. 7A). Regrettably, no alternative target sites

within the coding region of NAC $\beta$  were identified. Transformants harbouring the Cas9 cassette were selected based on their resistance to antibiotic selection on B5 Gamborg medium. A total of eight NAC $\beta$  transformants and sixteen NAC $\alpha$  transformants targeting the first exon were obtained. However, we failed to acquire any NAC $\alpha$  lines utilizing co-infiltration with both targets. The occurrence of insertions or deletions was confirmed via Sanger sequencing. Analysis of the eight NAC $\beta$  transformants revealed exclusively wild-type (WT) plants. Among the NAC $\alpha$  transformants, nine were WT and seven were mutants, exhibiting the multiplications of three-nucleotide deletions, including one 15 bp, one 18 bp, one 6 bp, and four 12 bp deletions. Notably, none of the mutants exhibited deletions that resulted in a frameshift of the open reading frame. The in-frame mutant lines did not display any phenotypic alterations.

To ascertain whether expression was preserved in the in-frame mutants, we selected three lines of NAC $\alpha$  and assessed their expression levels using RT-qPCR. Expression levels were normalized to those of a housekeeping gene ACT7 (Saint-Marcoux et al. 2015) and NAC $\beta$  in two separated calculations. The data did not show significant decrease in NAC $\alpha$  transcript level, with in case of ACT7 as reference, the difference between the CRISPR lines and WT was FC = 0.38 and in case of NAC $\beta$  as a reference, FC = 0.94 (Fig. 7B).



**Figure 7** gRNA targets for CRISPR-Cas9 editing of *M. polymorpha* NAC subunits and RT-qPCR analysis of NAC $\alpha$  mutant lines. A. gRNA targets for CRISPR-Cas9, the arrow on the upper side indicates exon target sites for unconditional knock-out. The arrows on the lower side indicate intron target sites for conditional knock-out. B. Results of RT-qPCR experiment for 3 NAC $\alpha$  CRISPR lines show no significant change in the expression of NAC $\alpha$  transcript.

### **Generation of NAC conditional knock-out lines in *M. polymorpha* (ongoing)**

For the generation of the conditional mutant, we used the system developed by Sugano *et al.* 2018, which in one step introduces the complementation cDNA sequence of target gene in the inducible Cre-lox system and CRISPR-Cas9 cassette targeting an intron of the target gene, causing its mutation (Sugano *et al.* 2018). Upon treatment with heat shock and dexamethasone, the complementing sequence is removed and the expression of tdTomato-NLS is triggered, which enables fluorescence selection of the mutants. Target sites in the 1<sup>st</sup> intron on NAC $\beta$  and NAC $\alpha$  were designed and inserted to the pMpGE010 destination vector. The cDNA sequences were inserted to the pmpGWB337 vector. The *Agrobacterium tumefaciens* was co-transformed with both constructs and colonies harbouring both constructs were selected with colony PCR and used for *Marchantia polymorpha* infiltration.

## **Discussion**

### **NAC is structurally conserved between plant lineages**

Remarkably, the structural conservation of NAC subunits are preserved even between archaea and eukaryotes. For instance, archeal NAC (aeNAC) shares a 28% sequence similarity with human NAC $\alpha$  and 26% with human NAC $\beta$  (Liu *et al.*, 2010). Intriguingly, human NAC and aeNAC compete for binding to nascent polypeptide chains, suggesting a level of functional overlap (Spreter *et al.*, 2005). In plants, the NAC proteins exhibit high sequence and secondary structure conservation, especially in the NAC domain, which also contains potential dimerization sites.

When examining NAC subunits in *Marchantia polymorpha* and *Arabidopsis thaliana*, the sequence similarity for NAC $\beta$  is 83.1% for NAC $\beta$ 1, and 83.2% NAC $\beta$ 2, and 81.2% for NAC $\alpha$ 1, 66.5% for NAC $\alpha$ 2, 80.6% for NAC $\alpha$ 3, 72.7% for NAC $\alpha$ 4, and 74.4% for NAC $\alpha$ 5 (Supplementary file 2). The proteins display conserved secondary structures, with the same numbers of  $\alpha$  helices and  $\beta$  sheets forming the NAC domain. However, structural conservation does not necessarily equate to functional conservation (Harris *et al.* 2020). The number of NAC subunits in *A. thaliana* and *M. polymorpha* genomes suggests most likely gene duplication in *A. thaliana* or possibly gene loss in *M. polymorpha*, potentially leading to subfunctionalization or neofunctionalization in *A. thaliana* NAC subunits.

A phylogenetic analysis involving gymnosperms, angiosperms (monocots and dicots), mosses, bryophytes, and algae, and including an archeal NAC $\alpha$  subunit and a yeast NAC $\beta$  subunit as root points, was conducted to elucidate the relationships among *A. thaliana* NAC subunits. This analysis reveals two distinct AtNAC $\alpha$  subgroups: one more closely related to evolutionarily older plants, including NAC $\alpha$ 3 and NAC $\alpha$ 1, and the other comprising angiosperms, including NAC $\alpha$ 2, NAC $\alpha$ 4, and NAC $\alpha$ 5, with the latter two being more closely related to each other than to NAC $\alpha$ 2.

Y2H interaction experiments investigating the dimerization capacity between MpNACs and AtNACs did not align with their phylogenetic relationship. Based on phylogenetic relationships, one might anticipate stronger interactions between MpNAC $\beta$  and AtNAC $\alpha$ 1 and AtNAC $\alpha$ 3, as these subunits appear to be more closely related in sequence, suggesting a closer relationship to the common ancestor shared with liverworts. However, MpNAC $\beta$  demonstrated a strong interaction with AtNAC $\alpha$ 2 and AtNAC $\alpha$ 4. Additionally, it exhibited

weaker interactions with AtNAC $\alpha$ 5 and AtNAC $\alpha$ 3. On the other hand, the binding preferences among AtNACs showed specificity, with AtNAC $\alpha$ 4 and AtNAC $\alpha$ 5 only dimerizing with AtNAC $\beta$ 2. This specificity may indicate a potential functional divergence, particularly for AtNAC $\alpha$ 4 and AtNAC $\alpha$ 5, which could assume roles beyond the heterodimeric chaperone, as observed in humans where single NAC $\alpha$  subunits or homodimers participate in transcriptional regulation (Liu et al. 2010).

### **NAC $\beta$ is functionally conserved between *A. thaliana* and *M. polymorpha*, and shows a similar role**

In *A. thaliana*, NAC functions as a housekeeping protein throughout all developmental stages and tissues, exhibiting cytoplasmic and nuclear localization in leaves (Fila et al. 2020). BiFC analysis shows that NAC dimers are predominantly cytoplasmic (Klodová 2019). Expression profiling in *M. polymorpha* indicates a significant expression across all stages, except for early sperm cells, with sporelings showing peak expression (Fig. 5B). GUS assays using 1000 bp promoter sequences of NAC $\beta$  and NAC $\alpha$  also indicated robust expression in both sporophytic and gametophytic tissues (Fig. 5A). Similarly, cytoplasmic and nuclear localizations for both NAC $\beta$  subunit of *M. polymorpha* was reported, indicating a similar subcellular localization and tissue expression levels between the two species.

To assess functional conservation, the NAC $\beta$  subunit from *M. polymorpha* was introduced into *nac $\beta$ 1/nac $\beta$ 2* knockdown mutants of *A. thaliana*. The transgene successfully rescued the phenotypic defects, demonstrating functional complementarity. However, only partial rescue was observed in flower organ numbers, suggesting a potentially specialized or divergent role of the two  $\beta$  subunits in meristem development of *A. thaliana*.

Phenotypical studies in plants with altered expression of NAC $\beta$  subunits, including *A. thaliana*, *O. sativa*, *N. benthamiana*, *Z. mays*, *T. aestivum*, sugar beet (monosomic addition line M14), *M. domestica*, *S. lycopersicum*, *A. hypogaea*, *C. sativus* L., *Suaeda asparagoids*, *Spartina alterniflora*, *Gerbera hybrid*, and *Salvia aegyptica*, show modified NAC expression in response to various stress treatments (Jamil et al. 2015). For example, in *A. thaliana* and transgenic *Jatropha curcas* overexpressing *A. thaliana* NAC $\beta$ , the proposed mechanism for NAC $\beta$  in cold stress involves stabilization of newly synthesized cold-stress responsive proteins (Peng et al. 2017; S. Yang et al. 2018). Conversely, plants with lower NAC $\beta$  expression exhibit a reduced tolerance to cold stress in *Triticum aestivum* and *A. thaliana* (Kang et al. 2013; S. Yang et al. 2018) and delayed seed germination under osmotic stress in *A. thaliana nac $\beta$ 1/nac $\beta$ 2* mutants (Fila et al. 2020). Despite the roughly 400 million years that have passed since the last common ancestor of *A. thaliana* and *M. polymorpha*, our observations, coupled with the functional similarities of NAC $\beta$  proteins across various plant species reported from literature, lead us to propose that the role of NAC $\beta$  has remained largely unchanged throughout the evolution of land plants. However, the diversification of subunits in angiosperms may indicate additional subfunctionalization in some plant species.

### **NAC is essential for plant development**

In animals, including *Mus musculus* or *Caenorhabditis elegans*, knockout of NAC $\beta$  subunits resulted in embryonic lethality (Deng and Behringer 1995; Markesich et al. 2000). In *Arabidopsis thaliana*, double homozygous mutants for NAC $\beta$  subunits showed continued less abundant NAC $\beta$ 2 expression, leading to a knockdown mutant (Fila et al. 2020). CRISPR-

Cas9 attempts to create *M. polymorpha* mutants for each NAC subunit have been challenging, with no plants obtained that have out-of-frame deletions or insertions. In contrast, CRISPR-edited non-essential genes typically yielded a mix of in-frame ( $3n$ ) and out-of-frame ( $3n + x$ ) mutations, as well as wild-type lines (You et al., 2020). In *Dictyostelium discoideum*, *Dync1li1* is an essential gene, while *KIF1A* and *fAR1* are non-essential, as evidenced by the proportions of out-of-frame mutations and mutant phenotypes observed (You et al., 2020).

The research estimated that the ratio of  $3n$  and  $3n + x$  mutations, where 'x' can be any integer from -2 to 2 in non-essential gene mutations is random. This hypothesis was confirmed, with the non-essential genes *KIF1A* and *fAR1* exhibiting out-of-frame mutations in 42% of 19 tested lines and 80% in 5 tested lines for each gene, respectively. Notably, lines with out-of-frame mutations presented the previously characterized mutant phenotype. The behaviour of *Dync1li1* mutants differed, acquiring only 4 wild-type like sequences and various in-frame mutations out of 18 tested lines. The probability of such an occurrence, under the assumption of random mutation generation by CRISPR-Cas9, was calculated to be  $p = 0.00518$ , a figure derived from averaging the results of two gRNA targets for *Dync1li1* tested. In *M. polymorpha*, only viable mutants of NAC $\alpha$ , regardless of the gRNA targets, had  $3n$  deletions, allowing for residual transcript expression. This suggests that NAC's presence is crucial for plant growth and survival, as well as in animals.

Further experiments using an inducible silencing system aim to confirm the lethality of NAC knockout. However, these experiments were not completed when the thesis was submitted.

## Supplementary material:

**Supplementary File 1:** List of primers used in this study, including names, sequences, and purposes of use.

**Supplementary File 2:** Results of pairwise alignment analysis of different NAC subunits from human, *Arabidopsis thaliana*, Archaea *Methanococcus aeolicus* (Meta), and *Marchantia polymorpha*. The sheet includes pictures of the entire alignment, and a table with the percentage of similarity for each comparison.

**Supplementary Figure 1:** Secondary structures of NAC subunits from the studied organisms. The red spirals indicate  $\alpha$ -helices, and the green arrows indicate  $\beta$ -sheets. If present, putative conserved structures including the NAC domain, ribosomal binding site, and heterodimerization site are shown.

## Bibliography

Aach, J., Mali, P., & Church, G. M. (2014). CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes John Aach, Prashant Mali, George M. Church Department of Genetics, Harvard Medical School, Boston, MA 02115, USA Corresponding authors. *BioRxiv*, 1–8.

Billey, E., Hafidh, S., Cruz-gallardo, I., Jr, C. G. L., Jean, V., Carpentier, M., Picart, C., Kumar, V., Kulichova, K., & Mare, E. (2021). *LARP6C orchestrates posttranscriptional*

- reprogramming of gene expression during hydration to promote pollen tube guidance. 2637–2661. <https://doi.org/10.1093/plcell/koab131>
- Boavida, L. C., & McCormick, S. (2007). Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. *Plant Journal*, 52(3), 570–582. <https://doi.org/10.1111/j.1365-313X.2007.03248.x>
- Bowman, J. L. (2013). Walkabout on the long branches of plant evolution. *Current Opinion in Plant Biology*, 16(1), 70–77. <https://doi.org/10.1016/j.pbi.2012.10.001>
- Bowman, J. L., Araki, T., & Kohchi, T. (2016). Marchantia: Past, present and future. *Plant and Cell Physiology*, 57(2), 205–209. <https://doi.org/10.1093/pcp/pcw023>
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S. S., Althoff, F., Araki, T., Arteaga-Vazquez, M. A., Balasubramanian, S., Barry, K., Bauer, D., Boehm, C. R., ... Schmutz, J. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell*, 171(2), 287–304.e15. <https://doi.org/10.1016/j.cell.2017.09.030>
- Crooks, G., Hon, G., Chandonia, J., & Brenner, S. (2004). NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Research*, 14, 1188–1190. <https://doi.org/10.1101/gr.849004.1>
- de Sousa, F., Foster, P. G., Donoghue, P. C. J., Schneider, H., & Cox, C. J. (2019). Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *The New Phytologist*, 222(1), 565–575. <https://doi.org/10.1111/nph.15587>
- Deng, J. M., & Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Research*, 4(4), 264–269. <https://doi.org/10.1007/BF01969120>
- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., & Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400(6745), 693–696. <https://doi.org/10.1038/23301>
- Doyle, J. A. (1998). Phylogeny of vascular plants. *Annual Review of Ecology and Systematics*, 29(1), 567–599. <https://doi.org/10.1146/annurev.ecolsys.29.1.567>
- Fíla, J., Klodová, B., Potěšil, D., Juříček, M., Šesták, P., Zdráhal, Z., & Honys, D. (2020). The beta subunit of nascent polypeptide associated complex plays a role in flowers and siliques development of *Arabidopsis thaliana*. *International Journal of Molecular Sciences*, 21(6), 1–30. <https://doi.org/10.3390/ijms21062065>
- Gamborg, O. L., Miller, R. A., & Ojima, K. (1968). Nutrient requirements of suspension cultures of soybean root cells. *Experimental Cell Research*, 50(1), 151–158. [https://doi.org/10.1016/0014-4827\(68\)90403-5](https://doi.org/10.1016/0014-4827(68)90403-5)
- Grefen, C., & Blatt, M. R. (2012). A 2in1 cloning system enables ratiometric bimolecular fluorescence complementation (rBiFC). *BioTechniques*, 53(5), 311–314. <https://doi.org/10.2144/000113941>
- Harris, B. J., Harrison, C. J., Hetherington, A. M., & Williams, T. A. (2020). Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Current Biology*, 30(11), 2001–2012.e2. <https://doi.org/10.1016/j.cub.2020.03.048>
- Hsieh, H. H., Lee, J. H., Chandrasekar, S., & Shan, S. O. (2020). A ribosome-associated chaperone enables substrate triage in a cotranslational protein targeting complex. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-19548-5>
- Jamil, M., Wang, W., Xu, M., & Tu, J. (2015). Exploring the roles of basal transcription factor 3 in eukaryotic growth and development. *Biotechnology and Genetic Engineering*

- Reviews*, 31(1–2), 21–45. <https://doi.org/10.1080/02648725.2015.1080064>
- Jill Harrison, C. (2017). Development and genetics in the evolution of land plant body plans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1713), 20–25. <https://doi.org/10.1098/rstb.2015.0490>
- Kang, G., Ma, H., Liu, G., Han, Q., Li, C., & Guo, T. (2013). Silencing of TaBTF3 gene impairs tolerance to freezing and drought stresses in wheat. *Molecular Genetics and Genomics*, 288(11), 591–599. <https://doi.org/10.1007/s00438-013-0773-5>
- Klodová, B. (2019). Studying dimer formation and effectors of *Arabidopsis thaliana* nascent polypeptide-associated complex. *Diploma Thesis, Charles University*.
- Kogan, G. L., & Gvozdev, V. A. (2014). Multifunctional nascent polypeptide-associated complex (NAC). *Molecular Biology*, 48(2), 189–196. <https://doi.org/10.1134/s0026893314020095>
- Liu, Y., Hu, Y., Li, X., Niu, L., & Teng, M. (2010). The crystal structure of the human nascent polypeptide-associated complex domain reveals a nucleic acid-binding region on the NACA subunit. *Biochemistry*, 49(13), 2890–2896. <https://doi.org/10.1021/bi902050p>
- Markesich, D. C., Gajewski, K. M., Nazimiec, M. E., & Beckingham, K. (2000). bicaudal encodes the *Drosophila* beta NAC homolog, a component of the ribosomal translational machinery. *Development (Cambridge, England)*, 127(3), 559–572. <http://www.ncbi.nlm.nih.gov/pubmed/10631177>
- Mishler, B. D., & Churchill, S. P. (1985). Transition To a Land Flora: Phylogenetic Relationships of the Green Algae and Bryophytes. *Cladistics*, 1(4), 305–328. <https://doi.org/10.1111/j.1096-0031.1985.tb00431.x>
- Montgomery, S. A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S., Bowman, J. L., Cognat, V., Maréchal-Drouard, L., Ekker, H., Hong, S. F., Kohchi, T., Lin, S. S., Liu, L. Y. D., Nakamura, Y., Valeeva, L. R., Shakirov, E. V., Shippen, D. E., ... Berger, F. (2020). Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Current Biology*, 30(4), 573–588.e7. <https://doi.org/10.1016/j.cub.2019.12.015>
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(10), E2274–E2283. <https://doi.org/10.1073/pnas.1719588115>
- Panchy, N., Lehti-Shiu, M., & Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiology*, 171(4), 2294–2316. <https://doi.org/10.1104/pp.16.00523>
- Peng, X., Wang, Q., Liu, H., & Shen, S. (2017). Phylogenetic and functional analysis of the basic transcription factor gene BTF3 from *Jatropha curcas*. *Plant Growth Regulation*, 82(2), 247–257. <https://doi.org/10.1007/s10725-017-0255-6>
- Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, 15(1), 8–15. <https://doi.org/10.1007/BF02772108>
- Saint-Marcoux, D., Proust, H., Dolan, L., & Langdale, J. A. (2015). Identification of reference genes for real-time quantitative PCR experiments in the liverwort *marchantia polymorpha*. *PLoS ONE*, 10(3), 1–14. <https://doi.org/10.1371/journal.pone.0118678>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9, 676. <https://doi.org/10.1038/nmeth.2019>



- Spreter, T., Pech, M., & Beatrix, B. (2005). The crystal structure of archaeal nascent polypeptide-associated complex (NAC) reveals a unique fold and the presence of a ubiquitin-associated domain. *Journal of Biological Chemistry*, 280(16), 15849–15854. <https://doi.org/10.1074/jbc.M500160200>
- Stotler, R. E., & Bischler, H. (1999). Systematics and Evolution of the Genera of the Marchantiales. *The Bryologist*, 102, 774. <https://api.semanticscholar.org/CorpusID:87156332>
- Sugano, S. S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., Shimada, T., Hara-Nishimura, I., Osakabe, K., & Kohchi, T. (2018). Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in *Marchantia polymorpha*. *BioRxiv*, 1–22. <https://doi.org/10.1101/277350>
- Thomas D Schneider, & R Michael Stephens. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20), 6097–6100.
- Tsuboyama, S., & Kodama, Y. (2014). AgarTrap: A simplified agrobacterium-mediated transformation method for sporelings of the liverwort *Marchantia polymorpha* L. *Plant and Cell Physiology*, 55(1), 229–236. <https://doi.org/10.1093/pcp/pct168>
- Tsuboyama, S., Nonaka, S., Ezura, H., & Kodama, Y. (2018). Improved G-AgarTrap: A highly efficient transformation method for intact gemmalings of the liverwort *Marchantia polymorpha*. *Scientific Reports*, 8(1), 1–10. <https://doi.org/10.1038/s41598-018-28947-0>
- Wiedmann, B., Sakai, H., Davis, T. A., & Wiedmann, M. (1994). A protein complex required for signal-sequence-specific sorting and translocation. *Nature*, 370(6489), 434–440. <https://doi.org/10.1038/370434a0>
- Wiese, A. J., Steinbachová, L., Timofejeva, L., Čermák, V., Klodová, B., Ganji, R. S., Limones-Mendez, M., Bokvaj, P., Hafidh, S., Potěšil, D., & Honys, D. (2021). Arabidopsis bZIP18 and bZIP52 accumulate in nuclei following heat stress where they regulate the expression of a similar set of genes. *International Journal of Molecular Sciences*, 22(2), 1–24. <https://doi.org/10.3390/ijms22020530>
- Yang, S., Song, C., Zhang, X., Jia, Y., Shi, Y., Ding, Y., & Gong, Z. (2018). OST1-mediated BTF3L phosphorylation positively regulates CBFs during plant cold responses. *The EMBO Journal*, e98228. <https://doi.org/10.15252/embj.201798228>
- You, Y., Ramachandra, S. G., & Jin, T. (2020). A CRISPR-based method for testing the essentiality of a gene. *Scientific Reports*, 10(1), 1–8. <https://doi.org/10.1038/s41598-020-71690-8>

**Beta subunits of the nascent polypeptide associated complex play role in protein folding and chloroplast transport in *A. thaliana***

**My contribution:** In this article I participated in planning the experiments and stating the hypothesis. I conducted most of the experiments myself except for the protein aggregate extraction and LC MS/MS analysis. I analysed the results and wrote the whole manuscript including preparation of all figures.

Current status: In progress, to be submitted during 2024.

# **$\beta$ subunits of the nascent polypeptide associated complex play a role in protein folding and protein sorting to chloroplasts in *A. thaliana***

## Abstract

Translation represents a pivotal cellular process where ribosomes synthesize proteins that are crucial for cell survival and function. The nascent polypeptide-associated complex (NAC), a heterodimeric structure comprising  $\alpha$  and  $\beta$  subunits, plays a significant role in the early stages of protein maturation, folding, and sorting by associating with nascent polypeptide chains emerging from the ribosomal exit tunnel. Both NAC subunits form a  $\beta$ -barrel-like structure, with their N-termini positioned in the ribosomal tunnel, facilitating the first interaction with emerging polypeptides. NAC involvement in protein trafficking, particularly its antagonistic role to the Signal Recognition Particle, positions it as a key player in directing nascent protein pathways and mitigating their mistargeting. Moreover, the knockout of NAC has been associated with embryonic lethality in several model animals, underscoring its essential role in development. Despite its established presence and importance in metazoan models, the detailed mechanisms of NAC role in plants remain unexplored. This study investigates NAC role in translation and protein sorting in *Arabidopsis thaliana*. By analysing NAC $\beta$  interactome, we revealed NAC collaboration with the cytoplasmic chaperone system and confirm its association with ribosomes and the components of the translation machinery. Furthermore, our data suggest NAC involvement in chloroplast protein import, building on previous observations in other organisms. However, we suggest that NAC function is pleiotropic and to a high extent redundant with other plant chaperones and protein sorting pathways.

## Introduction

Translation, a key cellular process, happens on ribosomes. During this protein synthesis, nascent polypeptides emerge from the ribosomal exit tunnel and are immediately engaged with various ribosome-associated factors, which are critical for the protein maturation, folding, and sorting (Hartl and Hayer-Hartl 2002; Kramer et al. 2009; Rospert and Chacinska 2006).

An essential early interactor with the nascent polypeptide is represented by the nascent polypeptide-associated complex (NAC), which forms a conserved heterodimeric structure comprising  $\alpha$  and  $\beta$  subunit, connected by a conserved NAC domain into a  $\beta$ -barrel-like structure with the N-termini of the proteins reaching the inside of the ribosomal tunnel, facilitating early interaction with freshly synthesized nascent polypeptides (Gamerding et al. 2019; Y. Liu et al. 2010; L. Wang et al. 2010; Wiedmann et al. 1994). Structure of NAC subunits is conserved across eukaryotes, and homologs of the NAC $\alpha$  paralogs have been discovered even in archaeal genomes (Macario and Conway De Macario 2001; Makarova et al. 1999). On the other hand, NAC complex (and even one of its subunits) is absent from bacteria, where the structurally unrelated chaperone Trigger factor possible fulfils a similar

role (Deuerling et al. 1999). NAC is highly abundant in a cell since it is present in an equimolar ratio with ribosomes (del Alamo et al. 2011; Raue et al. 2007). NAC complex associates closely with the ribosomal exit site, and intriguingly, both  $\alpha$  and  $\beta$  subunits seem to be capable of binding to ribosomes independently. The interaction between NAC complex and ribosomes is mediated through ribosomal proteins Rpl31/Rpl17, with Rpl31 being the binding site for NAC $\beta$  and Rpl17 for NAC $\alpha$  in yeast (Pech et al. 2010). This binding site is also targeted by the ribosome-associated complex (RAC) (Gamerdinger et al. 2019).

The function of NAC complex has remained an intriguing topic for over two decades since its initial description as a ribosome-bound protein interacting with nascent polypeptides (Wiedmann et al. 1994). Its role appears to be multifaceted, potentially pleiotropic. It has been detailed that NAC binding to specific ribosomal sites is competitively antagonistic to other complexes, such as the Signal Recognition Particle (SRP) and RAC, directing the trajectory of nascent proteins (Gamerdinger et al. 2019; Hsieh et al. 2020; Wickner 1995). NAC interaction with nascent polypeptide chains prevents inappropriate SRP associations, thus mitigating mistargeting of proteins to endoplasmic reticulum. In the absence of NAC, SRP and its receptor are rendered unnecessary, as ribosomes alone can direct nascent chains to the Sec61 complex, a protein transport channel to endoplasmic reticulum (Hsieh et al. 2020). NAC was also implicated in the attachment of cytosolic ribosomes to mitochondria and in facilitating translation-coupled protein import. However, this was shown only under *in vitro* conditions (Gamerdinger et al. 2015; Lesnik et al. 2014; Walsh et al. 2002). The NAC knock-out has been linked to embryonic lethality in various metazoans, including *Mus musculus*, *Drosophila melanogaster*, and *Caenorhabditis elegans* (Bloss et al. 2003; Deng and Behringer 1995; Markesich et al. 2000).

To date, the molecular mechanisms underlying the involvement of NAC complex in protein folding and sorting have not been explored in plants. However, the NAC role in stress tolerance and plant development was indicated. Its molecular function has been proposed to relate to cold stress tolerance through phosphorylation (Yang et al. 2018). Furthermore, other studies have associated NAC with chloroplast development (Ma et al. 2012). Our prior research showed that silencing of NAC $\beta$  subunits reduces the abundances of photosynthetic proteins but leads to the upregulation of other cytoplasmic chaperones (Fíla et al. 2020). Coupled with phenotypic defects in processes that heavily rely on a functional and efficient translation apparatus, such as slow-growing and poorly germinating pollen tubes and diminished chlorophyll content, we hypothesized that NAC is crucial for efficient translation and protein import in plants, as was previously observed in metazoans and yeast (Duttler et al. 2013; Fíla et al. 2020; Koplín et al. 2010; L. Wang et al. 2010; Wiedmann et al. 1994). A recent study proposed a model where NAC is essential for chloroplast protein import in rice (*O. sativa*), suggesting that NAC post-translationally assist in delivery of nascent polypeptides to the chloroplast (Liu et al. 2021). However, several previous reports suggest cytoplasmic and nuclear localization in several plant species including *A. thaliana*, *O. sativa* and *Triticum aestivum* (Fíla et al. 2020; Ma et al. 2012; Ya Wang et al. 2012).

In our current study, we aim to elucidate NAC role in translation and protein sorting. By analysing its interactome, we have linked NAC to the cytoplasmic chaperone system and confirmed its association with ribosomes and other components of the translation

machinery. We propose that NAC collaborates with other chaperones in targeting the nascent polypeptides to chloroplasts.

## Material and methods

### Plant cultivation

Plants (*Arabidopsis thaliana*, ecotype Columbia-0 wild type (WT), double homozygous mutant *nacβ1/nacβ2*, complemented lines for NACβ1 and NACβ2) were grown on Jiffy peat pellets. The seeds underwent a three-day stratification at 4 °C before they were moved to a growth chamber set to long-day conditions, with a cycle of 16 hours of light followed by 8 hours of darkness, at a temperature of 21 °C. Fourteen days after sowing, the seedlings were transferred into new pots containing Jiffy peat pellets.

### DNA extraction, genotyping

Genomic DNA was isolated from *Arabidopsis* leaves using the established CTAB extraction (Porebski, Bailey, and Baum 1997). The extracted DNA served as a template for PCR-based genotyping. The PCR mix comprised 0.4 μM concentration of primers (Sigma-Aldrich, Haverhill, UK), 0.2 mM of dNTPs (Promega, Madison, USA), 1 μL of the genomic DNA (app. 330 ng), and 1 unit of Taq DNA polymerase (Merck, Brno, Czech Republic) in the reaction buffer provided by the manufacturer. The thermocycling conditions were as follows: denaturation at 94 °C for 2 minutes, followed by 40 cycles at 94 °C for 30 seconds, annealing at 55 °C for 30 seconds, extension at 72 °C for 1 minute, and a final elongation step at 72 °C for 5 minutes. For electrophoretic analysis, the amplification products were mixed with 6× Loading Dye (Thermo Fisher Scientific, Waltham, USA) and electrophoresed on a 1% agarose gel stained with ethidium bromide (0.5 μg/mL) in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, mM EDTA). The GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Waltham, USA) was used as a size standard, and the gels were imaged with the G-Box Gel Documentation System (Syngene, Cambridge, UK).

### Cloning and plant transformation

The coding sequences of all NAC subunits and genes selected as candidate proteins for protein-protein interaction analyses were amplified from leaf cDNA of Col-0 wild type plants using a set of primers listed in Supplementary file 1. The GoldenBraid system 3.0 was utilized to construct NACβ the coding sequence (CDS) tagged with the FLAG epitope, driven by the native promoter and terminated with the NOS terminator (Sarrion-Perdigones et al. 2013). Constructs for yeast two-hybrid, bimolecular fluorescence complementation (BiFC), and Förster resonance energy transfer (FRET) experiments were prepared by the Gateway™ cloning system. Sequence integrity of the prepared constructs was verified by Sanger sequencing (Eurofins, Germany). Plant lines expressing NACβ subunits tagged with GFP under the native promoter were obtained from a previous study (Fila et al. 2020). Expression vectors were electroporated into homemade competent *Agrobacterium tumefaciens* cells, strain GV3101 (pMP90RK). Stable transformed *Arabidopsis thaliana* lines were generated by the floral dip (Clough and Bent 1998). The acquired transgenic lines were selected based on genotyping with primers from Supplementary file 1 and by Western blot analysis (described subsequently).

### Protein extraction, Western blot and co-immunoprecipitation

The extraction of the soluble protein fraction was conducted using 50 mg of snap-frozen *A. thaliana* unopened flower buds, which were grinded by a pestle in a mortar. The proteins were then extracted by the added extraction buffer (10 mM Tris-Cl pH 7.5, 14% glycerol, 150 mM NaCl, 0.5 mM EDTA, 0.5% IGEPAL® CA-630, 7 mM PMSF, and a Protease Inhibitor Cocktail without EDTA (11873580001, Roche, Basel, Switzerland) to a final volume of 1 ml. The acquired mixture was transferred to microtubes. Cellular residue was segregated by centrifugation at 20,000×g for 20 minutes at 15 °C, and the clear supernatant was decanted into new tubes. For electrophoresis, the samples were mixed with a 3x sample buffer (Tris-Cl pH 6.8 0.15 M, glycerol 30 %, SDS 6 % w/v, DTT 0.15 M, bromophenol blue 0.03%) and loaded onto SDS-PAGE gels (11.25%T, 2.6%C resolving gel and 5%T, 2.6%C stacking gel). The gels were run at 75V for 30 minutes followed by 150V for 60 minutes. For the transfer of proteins, a semi-dry Western blotting technique was utilized. The detection of proteins was achieved with a mouse anti-GFP primary antibody (gfms, Chromotek, Planegg, Germany, diluted 1,000 times) or mouse anti-FLAG primary antibody (Monoclonal ANTI-FLAG® M2, F3165, Merck, Rahway, New Jersey, USA) following an overnight incubation, then a 30-minute incubation with a goat anti-mouse IgG (H+L) secondary antibody conjugated to alkaline phosphatase (31320, Thermo Fisher Scientific, Waltham, Massachusetts, USA; diluted 3,000 times) as described previously (Billey et al. 2021). The alkaline phosphatase signal was visualized by AP buffer (100 mM Tris-Cl pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>) with BCIP (5-Bromo-4-chloro-3-indolyl phosphate, 0.5 mM) and NBT (4-Nitro blue tetrazolium chlor, 0.4 mM). Molecular weights of the proteins of interest were assessed by a prestained protein marker (PageRuler™ Prestained Protein Ladder, cat. number: 26616, Thermo Scientific™). For co-immunoprecipitation, the same protein extraction method was used and then followed by an over-night incubation of protein extracts with anti-GFP agarose beads (GFP-Trap® Agarose, 001152-02-01, Chromotek, Planegg, Germany) or anti-FLAG magnetic beads (Anti-FLAG® M2 Magnetic Beads, M8823, Merck, Rahway, New Jersey, USA) at 4°C. After incubation, the remaining beads were removed and the samples were washed three times with the wash buffer (10 mM Tris-Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA (Ethylenediaminetetraacetic acid)) and subjected to LC MS/MS analysis.

### **LC MS/MS analysis of the interactome**

Following a wash with wash buffer, the proteins bound to the beads were digested in situ by 0.5 µg of trypsin (sequencing grade, Promega) in a 50 mM NaHCO<sub>3</sub> solution, followed by an incubation at 37°C for 18 hours. The acquired peptides were then transferred into vials for liquid chromatography-mass spectrometry (LCMS), using a mixture of 2.5% formic acid and acetonitrile with an added 0.01% final concentration of polyethylene glycol (20,000), and subsequently concentrated by a SpeedVac (Thermo Fisher Scientific).

The peptide mixtures were analysed by LCMS/MS using an Ultimate 3000 RSLCnano system coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). The tryptic digests were first concentrated and desalted on a trapping column (100 µm × 30 mm, 3.5 µm particles, X-Bridge BEH 130 C18 sorbent; Waters, Milford, MA, USA; maintained at 40 °C). The trapping column, washed after the run with 0.1% formic acid, facilitated the elution of peptides at a flow rate of 500 nl.min<sup>-1</sup> onto an analytical column (Acclaim Pepmap100 C18, 3 µm particles, 75 µm × 500 mm; with a column compartment temperature of 40 °C, Thermo Fisher Scientific) using a 135-minute non-linear gradient program ranging from 1% to 56% of mobile phase B (mobile phase A: 0.1% formic acid in

water; mobile phase B: 0.1% formic acid in 80% acetonitrile). The outlet of the analytical column was directly attached to a Digital PicoView 550 ion source with sheath gas, using a SilicaTip emitter (New Objective; FS360-20-15-N-20-C12). An Active Background Ion Reduction Device (ABIRD, ESI Source Solutions) was also employed.

MS data were gathered in a data-dependent mode, prioritizing scans by precursor abundance, with a survey scan range of  $m/z$  3502000. The survey scan's resolution was 120,000 (at  $m/z$  200) with a target ion count of  $4 \times 10^5$  and a maximum injection time of 100 ms. Higher energy collisional dissociation (HCD) MS/MS spectra were obtained with 30% relative fragmentation energy at a resolution of 15,000 (at  $m/z$  200), targeting  $5.0 \times 10^4$  ions and a maximum injection time of 22 ms. A dynamic exclusion period of 30 seconds was set post one MS/MS spectra acquisition, with an MS/MS fragmentation isolation window of 1.2  $m/z$ .

Mass spectrometric RAW data files were processed by MaxQuant software (version 1.6.2.0) with standard settings, unless specified otherwise. Searches against the cRAP database (sourced from <http://www.thegpm.org/crap>) included common contaminants like keratin and trypsin, and the UniProtKB database for *Arabidopsis thaliana* ([ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/reference\\_proteomes/Eukaryota/UP000006548\\_3702.fasta.gz](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/reference_proteomes/Eukaryota/UP000006548_3702.fasta.gz); accessed May 2018, version 2018/05 with 27,582 protein sequences). The search parameters allowed for optional modifications such as oxidation of methionine and proline, deamidation of asparagine and glutamine, and acetylation of the protein N-terminus, using trypsin/P with up to two missed cleavages. Peptides and proteins were validated with a false discovery rate below 0.01, with inclusion criteria of at least one unique or razor peptide. Cross-analysis consistency was ensured using the 'Match between runs' feature for all samples. Protein quantification was based on intensity measurements from MaxQuant. The protein intensity data reported in the proteinGroups.txt output file from MaxQuant were further analysed using a software container environment (<https://github.com/OmicsWorkflows>), version 3.7.2a.

### **Yeast-Two Hybrid**

The Yeast Two-Hybrid (Y2H) assays were conducted by the ProQuest Two-Hybrid System, following the manufacturer's instructions. Interaction intensity was evaluated by observing growth on selective media (SC-LEU-TRP-URA, SC-LEU-TRP + FoA, SC-LEU-TRP-URA +3AT) and assessing  $\beta$ -galactosidase activity through X-gal assays. The transformation was repeated thrice to ensure reliability.

### **Transient Expression, Bimolecular Fluorescence Complementation, Förster resonance energy transfer**

We utilized the Multisite Gateway system to create pBiFCt-2in1 constructs (Grefen and Blatt 2012). *Agrobacterium tumefaciens* cells harbouring selected expression clones were grown overnight, harvested, and prepared for leaf infiltration in *Nicotiana benthamiana* as described previously (Wiese et al. 2021). Two days after infiltration, the fluorescence in leaves was examined by a Zeiss LSM880 confocal microscope (Carl Zeiss AG, Oberkochen, Germany) with excitation wavelengths appropriate for GFP/YFP (excitation 488 nm, emission 507 nm) and mRFP (excitation 561 nm, emission 607 nm) in case of BiFC and for mVenus (excitation 488 nm, emission 527 nm) and mTorquoise2 (excitation 421 nm, emission

474 nm) for FRET AB experiment. The FRET AB was analysed using the ZEN black software (Carl Zeiss AG, Oberkochen, Germany). Images were processed by Fiji/ImageJ (Schindelin et al. 2012) and Adobe Photoshop CS6 (Adobe, San Jose, California, USA).

### **Protein aggregate isolation**

Protein aggregates were isolated from two tissues leaves and unopen flower buds following the procedure described in (Planas-Marquès, Saul Lema, and Coll 2016). Briefly, 200 mg of tissue was harvested, frozen in liquid nitrogen and grinded. After the addition of 4 ml of fractionation buffer, the sample was transferred to 15-ml Falcon tube and vortexed properly. The suspension was centrifuged ( $2000\times g$ , 5 min, 4 °C) to remove large debris particles. The supernatant was collected and centrifuged again ( $6,000g$ , 10 min, 4°C) to obtain the final total fraction. The supernatant was transferred to a Fresh Falcon tube and protein concentration was estimated by 2D-Quant Kit (GE Healthcare, Uppsala, Sweden) according to the manufacturer's instructions. The samples were diluted to get the same amount of proteins in each sample (mostly 1.1 mg proteins) and spun again ( $100,000\times g$ , 90 min, 4°C). The supernatant contained a final soluble fraction whereas the pellets were subjected to the subsequent steps. The pellet was mixed with 3 ml of fractionation buffer with added 0,3% Triton X-100, resuspended by pipetting and incubated for 1 hour at 4°C with mixing by rotation. Then, the samples were centrifuged ( $50,000\times g$ , 60 min, 4°C). The supernatants represented the final microsomal fraction whereas the pellet was again mixed with 3 ml of fractionation buffer including 0.3% Triton X-100 and sonicated on ice with Bioruptor set on 50% power intensity with the following parameters: 3 cycles, 30 s "ON", 30 s "OFF". The aggregate fraction was acquired that was subjected to LC-MS/MS analyses that were conducted as above.

### **Measurement of photosynthesis performance**

Thylakoid membranes were isolated as described previously (Mark Aurel Schöttler, Kirchhoff, and Weis 2004). Chlorophyll content in isolated thylakoids was assessed by 80% (v/v) acetone extractions, following the methodology of Porra et al. (1989). The abundance of photosystem II (PSII), photosystem I (PSI), cytochrome b6f (Cytb6f), and plastocyanin (PC) in the thylakoid samples was quantified employing techniques established by Schöttler *et al.* (2004). Quantification of PSI relied on the measurement of P700 difference absorption at wavelengths between 830 and 870 nm in solubilized thylakoids, conducted with a Dual-PAM-100 instrument (Walz, Effeltrich, Germany) as described previously (Mark A. Schöttler et al. 2007; Mark Aurel Schöttler et al. 2007). The determination of PSII and Cytb6f levels was based on difference absorption measurements of cytochrome b559 and Cytb6f, with the analytical methods and data analysis protocols described by Kirchhoff et al. (2002) (Kirchhoff, Mukherjee, and Galla 2002) and Schöttler et al. (2007a).

The maximal quantum efficiency of PSII (Fv/Fm) was gauged in leaves dark-adapted for one hour, using chlorophyll fluorescence metrics obtained with a Dual-PAM-100 fluorimeter on intact plants at ambient temperature. For recording the 77 K chlorophyll-a fluorescence emission spectra, an F-6500 fluorometer (Jasco, Tokyo, Japan) was utilized. This involved thylakoid membranes isolated to deliver a chlorophyll concentration of  $10 \mu\text{g}\cdot\text{mL}^{-1}$ . The samples were excited at 430 nm (10 nm bandwidth), and the emission was captured from 655 to 800 nm at 0.5 nm intervals (1 nm bandwidth).



Thylakoid proton conductivity ( $\text{gH}^+$ ), indicative of ATP synthase activity, was inferred from the dark-interval relaxation kinetics of the electrochromic shift. These electrochromic shift signals were recorded and analysed by a KLAS-100 spectrophotometer (Walz) (Rott et al. 2011).

### ***in vivo* chloroplast import assay**

Mesophyll protoplasts were extracted from the leaves of 6 weeks old *A. thaliana* plants. The protoplast isolation and transformation were carried out as described previously in Gross et al. (2011), with modifications that included the utilization of 10 mg of DNA per plasmid for the transformation process and the addition of ampicillin to the K3 solution in the concentration of  $0.1 \text{ mg}\cdot\text{mL}^{-1}$  to avoid bacterial contamination. The transformed protoplasts were left to incubate overnight at laboratory temperature and then observed under the confocal microscope. For the *in vivo* chloroplast import assay, protoplasts were isolated from the Col-0 WT and *nacβ1/nacβ2* mutant plants (Fila et al. 2020). The constructs for the assay were kindly provided by Dr. Sotirios Fragkostefanakis from Goethe University Frankfurt am Main (Germany). The *in vivo* assay was carried out as described previously (Sommer et al. 2013).

### **Preparation and observation of samples under the transmission electron microscope**

Freshly harvested leaves from 6 weeks old plants were cut to small, long pieces with razor and washed twice with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM  $\text{Na}_2\text{HPO}_4$ , 1.8 mM  $\text{KH}_2\text{PO}_4$ ). Then samples were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer (pH 7.2) at laboratory temperature for 2 hours. The samples were washed twice with 0.1 M Na-cacodylate buffer with 2% sucrose. Then, the samples were incubated with 1% reduced osmium tetroxide for 1 hour with gentle shaking and washed twice with the cacodylate buffer. Dehydration with ethanol dilution series was performed with 30-50-70-90% of ethanol and followed by two washes with 100% ethanol, propylene oxide. After they were washed, they were incubated overnight in 50:50 propylene oxide: araldite mixture. Then, the samples were washed twice with pure araldite for two hours and the pellet was transferred to beam capsule for hardening. The slides were cut to 50nm thickness on Leica ultramicrotome and observed under TEM microscope Zeiss 900 (Carl Zeiss AG, Oberkochen, Germany).

## **Results**

### **Interactome analysis of NACβ subunits**

To explore the interactome of NACβ1 and NACβ2 subunits, we utilized *A. thaliana* lines expressing C-terminally tagged NACβ proteins. One line was under the control of the 35S promoter tagged with GFP, and the other with a FLAG epitope tag. We included GUS-GFP and GUS-FLAG plant lines as controls to identify non-specific interactions. From these, we identified 1,249 proteins in GFP line and 3,775 in FLAG line. Significant quantitative and qualitative interactions were observed with proteins in the NACβ1-GFP, NACβ2-GFP, NACβ1-FLAG, and NACβ2-FLAG lines, revealing interaction counts of 483, 413, 1,477, and 1,542 respectively (Supplementary File 2). The interactomes of NACβ1 and NACβ2 shared a considerable number of proteins, with variations depending on the tag used (Fig. 1A).

First, only overlapping proteins were analysed. There were 177 proteins commonly identified in the NAC $\beta$ 1 interactomes and 173 in the NAC $\beta$ 2 interactomes across GFP and FLAG lines, with 137 proteins shared by both. This overlap included all five NAC $\alpha$  subunits, confirming co-immunoprecipitation across lines. Predominantly, co-immunoprecipitated proteins were ribosomal subunits, with 59 cytoplasmic, 3 mitochondrial, and 8 chloroplastic proteins (Fig. 1B, Supplementary File 2). Additional components of translation machinery included 4 proteins involved in the formation of translation initiation complex and various chaperones, including 3 HSP70 family chaperones involved in *de novo* folding, as well as chloroplast HSP70-1 and mitochondrial HSP60 chaperonin. Most proteins exhibited cytoplasmic localization, but numerous also localized to the mitochondrion, vacuole, chloroplast, or the Golgi apparatus and endoplasmic reticulum (Fig. 1C, Supplementary File 2).

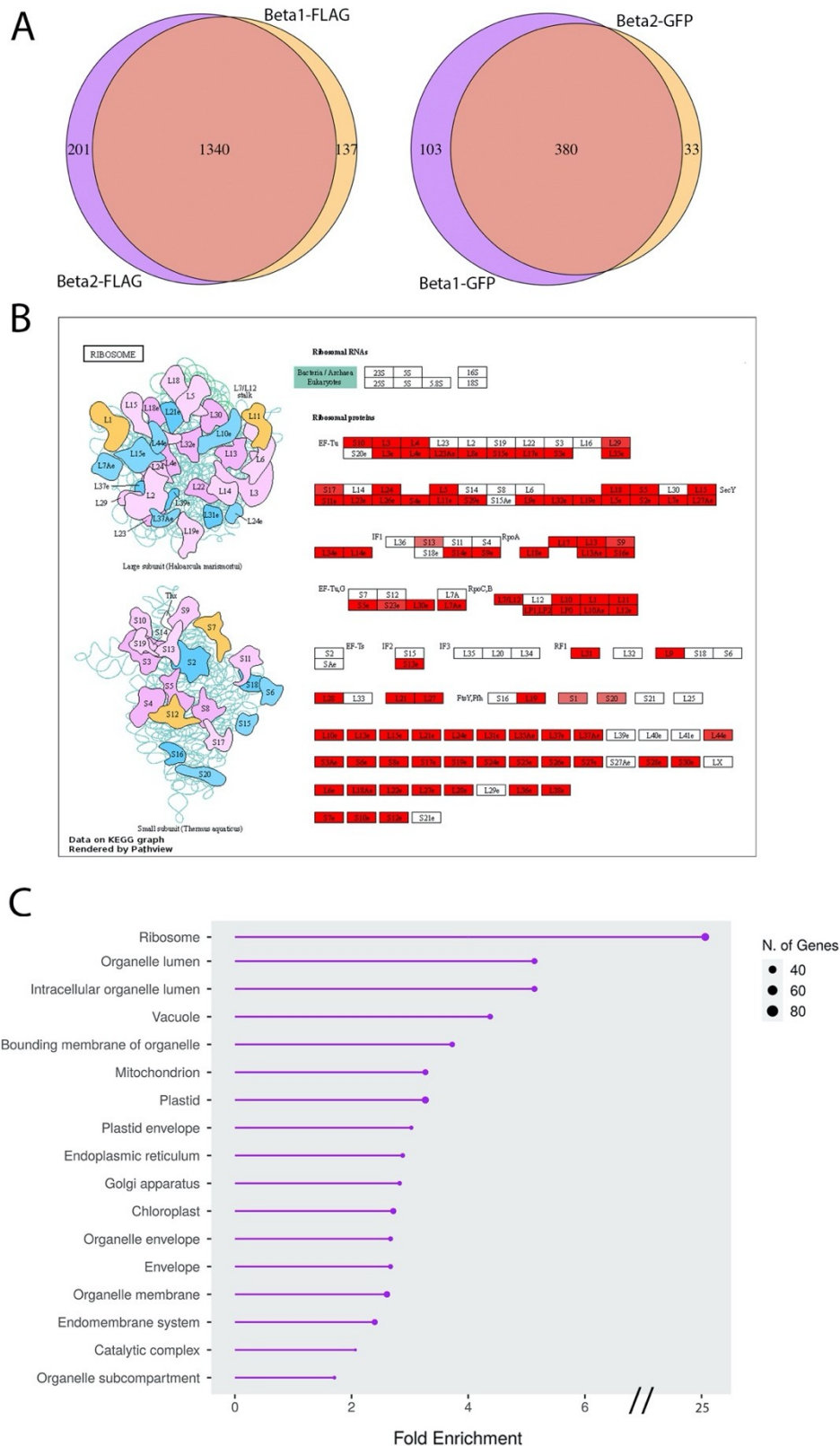
The NAC $\beta$ 2-specific interactome revealed 37 unique proteins, including ribosomal proteins, vesicle-associated proteins, chaperones, and TOM40 translocator. The NAC $\beta$ 1-specific interactome had 40 unique proteins, with additional ribosomal subunits.

Next, we analysed proteins present in the FLAG interactome. There were identified 113 proteins related to translation, including 101 ribosomal proteins and 14 proteins involved in translation initiation (Supplementary File 2). Protein transport-related proteins encountered a total of 89, 21 were involved in chloroplast transport, including the cytoplasmic chaperone contact site TOC64-III and the translocon TOC-III 75 (Supplementary File 2). Various inner transport proteins, such as TIC proteins, were also present. The majority of transport-related proteins are implicated in vesicular trafficking, including parts of COPI, COPII, and clathrin-coated vesicles or vesicle-processing proteins, along with proteins from the Golgi apparatus, endoplasmic reticulum, or vacuole. A total of 67 proteins play a direct role in vesicular transport (Supplementary File 2). Other enriched protein groups include those involved in lysosomal transport, ER organization, endosome to vacuole transport, and protein secretion, representing the entire protein pathway from nucleus to vacuole/exterior. Mitochondrial transport is also represented, including 3 inner membrane TIM importers and six components of the 21-protein TOM complex. The folding-related protein group consists of 44 proteins, including the entire cytoplasmic chaperonin T-complex and its 8 subunits (Supplementary File 2). The group contains proteins residing in the cytoplasm, mitochondria, chloroplast, and ER, including oxidoreductases ERO1 and ERO2, protein disulfide isomerase 5-2, endoplasmin, and two calnexin homologs, as well as SRP54.

Photosynthesis-related proteins included 45 proteins both nuclear and chloroplast-encoded, predominantly nuclear, with functions in the electron transport chain and photosystem repair and assembly (Supplementary File 2). Among the less abundant photosynthetic proteins in the *nac $\beta$ 1/nac $\beta$ 2* mutant compared to WT flower bud proteomes (Fila et al. 2020), 14 were co-immunoprecipitated, mainly PSII components.

There were 201 proteins co-immunoprecipitated exclusively in the NAC $\beta$ 2-FLAG and 137 in NAC $\beta$ 1-FLAG lines, with each group showing distinct biological process enrichment. The NAC $\beta$ 2 interactome included 86 chloroplast-localizing proteins, including four transporters

(TOC33, TIM22-3, OPE161, and ALBINO3), 41 mitochondrial, and 37 ER-localizing proteins, among other ribosomal proteins and nine plastid organizing proteins. The NAC $\beta$ 1 interactome revealed 54 chloroplast-localizing proteins, 37 mitochondrial, and 9 ER-targeted proteins, including three PIP proteins (PIP2, PIP1E, and PIP1B).



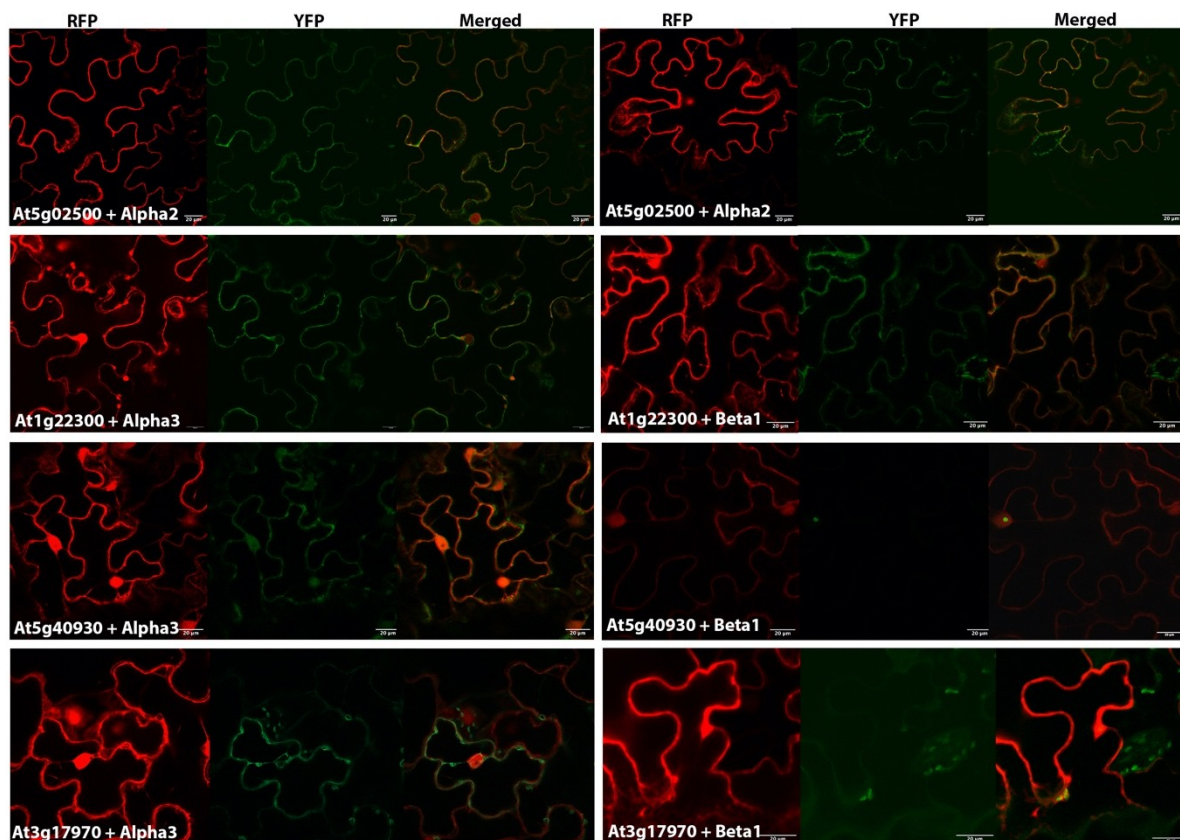
**Figure 1** Co-immunoprecipitation analysis of NAC $\beta$  subunits. **A.** Venn diagrams showing the number of co-immunoprecipitated proteins with NAC $\beta$ 1 and NAC $\beta$ 2 in FLAG and GFP lines, demonstrating a high overlap of co-immunoprecipitated proteins between NAC $\beta$ 1 and NAC $\beta$ 2. **B.** Pathview output indicating which ribosomal subunits (highlighted in red) were co-immunoprecipitated with the NAC $\beta$ 1 subunit. **C.** Gene Ontology (GO) cellular component analysis of proteins present in both NAC $\beta$  subunits for both FLAG and GFP lines, indicating localization to various cellular compartments.

### **Direct interaction of NAC subunits with selected candidates**

Cytoplasmic proteins and outer translocon complexes implicated in peptide sorting to mitochondria and chloroplasts identified within NAC $\beta$  interactomes, were examined for direct interactions by yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) assays. The rationale for employing both methods was to capture different types of NAC subunit's interaction capabilities: 1) Y2H assays will identify direct protein-protein interactions, indicating binding of NAC subunits to specific proteins; 2) BiFC assays will reveal the proximity of NAC complex to the translocon complex, which could result from direct interactions or indirect associations, such as those mediated by the polypeptide being transported, since BiFC requires two YFP halves to be within approximately 7 nm apart from each other for fluorescence complementation (Don et al. 2021).

Candidates that were further tested, included HSP70 (AT4G22670), HSP70-1 (AT5G02500), TOC64-III (AT3G17970), a 14-3-3 protein (AT1G22300), chloroplast-specific SRP (AT5G03940), and TOM20.4 (AT5G40930). Y2H interactions were assessed for both NAC $\beta$  subunits and selected NAC $\alpha$  subunits in both prey and bait configurations. BiFC vectors were chosen based on the position of important domains in the tested proteins. NAC proteins were engineered with the fluorophore attached to the C-terminus to avoid interference with NAC function and dimerization, which can occur if the fluorophore is attached to the N-terminus (unpublished data). For translocon components, the fluorophore was attached according to their membrane anchor orientation, with TOM20s tagged at the N-terminus and TOC64-III at the C-terminus. However, N-terminal fluorophore tagging on TOM20 proteins disrupted their mitochondrial localization; therefore, their interaction with NAC was only tested in a non-localized, cytoplasmic context.

BiFC indicated an interaction between HSP70 and NAC $\alpha$ 2 in the cytoplasm (Fig. 2). Both Y2H and BiFC confirmed the interaction of HSP70-1 with NAC $\alpha$ 2. BiFC also detected interaction with NAC $\alpha$ 4, which was cytoplasmically localized (Fig. 2), but there was no interaction with NAC $\alpha$ 1 and NAC $\alpha$ 3. The 14-3-3 protein displayed interactions with NAC $\beta$ 1 and NAC $\alpha$ 3 in BiFC assays, but no interaction was noted in Y2H assays. The interaction of TOM20.4 with cytoplasmic localization was confirmed with NAC $\alpha$ 3 and only nucleolus signal with NAC $\beta$ 1 (Fig. 2). TOC64-III interacted with NAC $\beta$ 1 and NAC $\alpha$ 3 at the chloroplast membrane and occasionally in cytoplasm (Fig. 2). As a negative control, a chloroplast-localized SRP homolog was tested and, as expected, no interaction was observed with any tested NAC subunit.

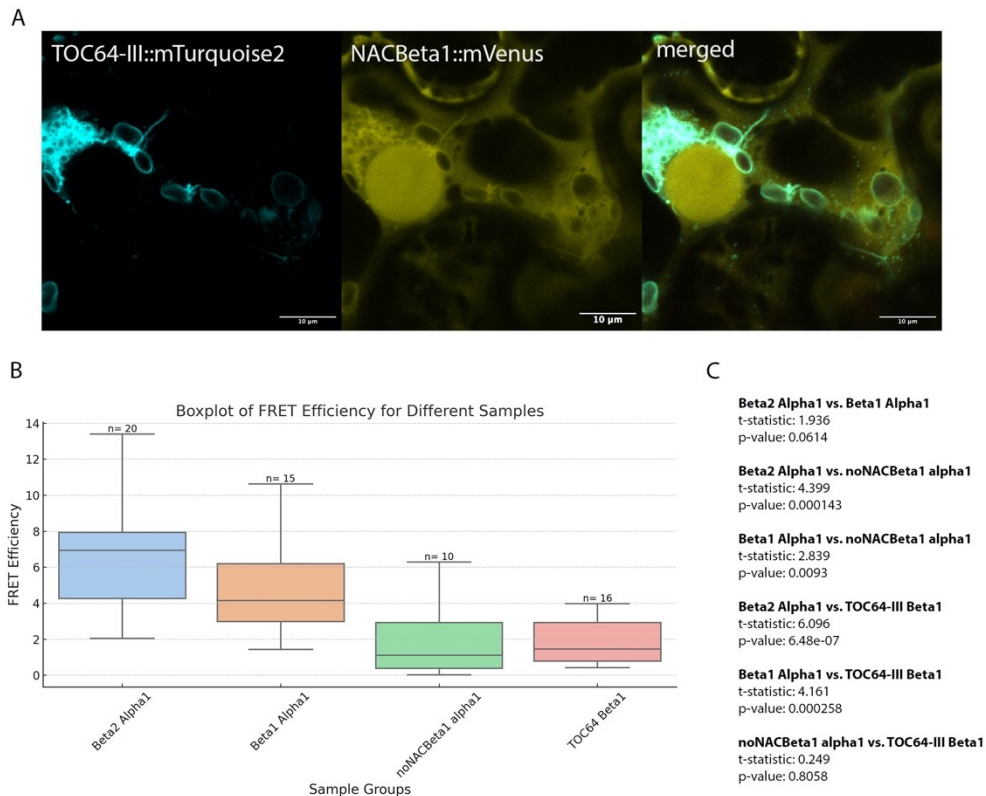


**Figure 2** Results of BiFC analyses testing interactions between selected candidates and NAC subunits. The proven interaction is represented by the YFP signal, whereas RFP channel indicates successful infiltration of *N. benthamiana* leaves. AT5G02500, AT1G22300, and AT5G40930 show cytoplasmic localization of dimers with NAC. AT3G17970 shows cytoplasmic and chloroplast/chloroplast membrane localization. The scale bar indicates 20  $\mu\text{m}$ .

### **NAC subunits co-localise with translocon TOC64-III on chloroplast membrane**

The relationship between NAC and TOC64-III was further examined using a 2in1 vector for FRET-AB analysis with TOC64-III with C-terminally bound mTourqois2 fluorophore and NAC subunits with C-terminally bound mVenus fluorophore, a fluorophore pair capable of Foerster Resonance Energy Transfer (FRET). The construct was transiently infiltrated to the *N. benthamiana* leaves and the localisation of the signals was observed under confocal microscope. The localisation of TOC64-III showed chloroplast membrane localisation (Fig. 3A), which corresponds to its translocon function. NAC-mVenus was localised to cytoplasm and nucleus. Interestingly, by comparing the localisation signals as co-localisation, overlap of signal around chloroplasts was apparent, showing a positive correlation of TOC64-III and NAC signal (Fig. 3A). We continued with the FRET-AB analyses and measured fluorescence changes in donor fluorescences to calculate FRET efficiency. The regions including chloroplast were selected as regions of interest for bleaching in case of TOC64-III and NAC $\beta$ 1. However, on average we did not get a positive FRET for TOC64-III and NAC (Fig. 3B). As a positive control, we used interaction between the NAC $\alpha$  and NAC $\beta$  subunits (NAC $\alpha$ 1-NAC $\beta$ 2 and NAC $\alpha$ 1-NAC $\beta$ 1), which show a higher FRET efficiency in the regions of interest in the cytoplasm (Fig. 3B). On the other hand, combination of NAC $\alpha$  subunit with NAC $\beta$  subunit lacking the NAC domain (NAC $\alpha$ 1-noNAC $\beta$ 1) showed low FRET efficiency (Fig. 3B). T-test with equal variance was used to evaluate the differences between samples, showing significant

difference between the positive and negative controls and also between the positive controls and the TOC64-III - NAC $\beta$ 1 test (Fig. 3C).



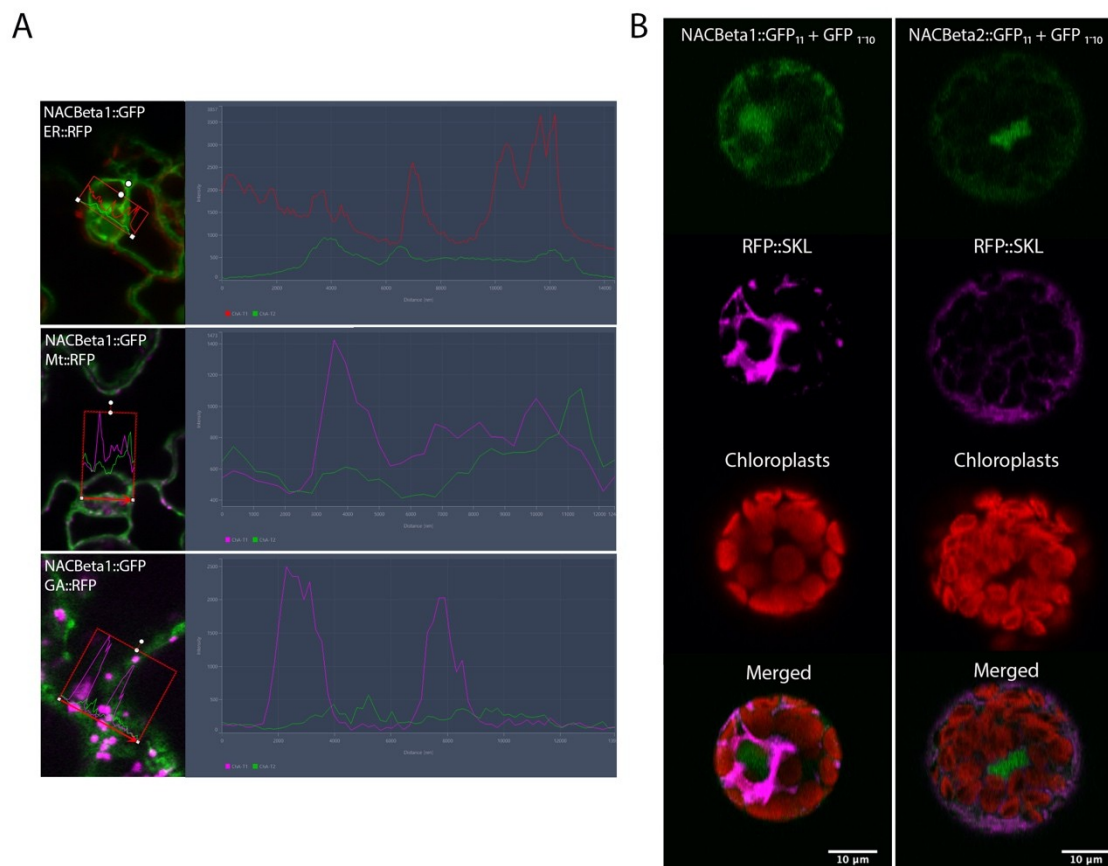
**Figure 3** Interaction analyses between NAC $\beta$ 1 and TOC64-III. **A.** TOC64-III localizes to the chloroplast membrane, whereas NAC $\beta$ 1 localization is observed in the cytoplasm, nucleus, and is also concentrated around the chloroplast membrane, colocalizing with the TOC64-III signal. **B.** Box plot showing the results of FRET AB analysis. **C.** Statistical evaluation of differences between FRET AB experiments.

### Co-localisation of NAC subunits with markers of subcellular compartments

In our previous work, we reported the localization of NAC $\beta$  subunits to the cytoplasm and nucleus (Fíla et al. 2020). The formation of NAC heterodimers predominantly in the cytoplasm was also documented with BiFC (Klodová 2019). To further specify the localization of NAC, we generated stable *A. thaliana* lines expressing organelle-specific markers fused with fluorescent tags alongside GFP-tagged NAC $\beta$  protein to assess co-localization. These lines included markers targeting the Golgi apparatus (GA), endoplasmic reticulum (ER), and mitochondria, as established by Nelson et al. (2007). We computed the Pearson correlation coefficient and analysed signal profiles of fluorescence intensities within seedling cotyledons and roots to determine co-localization.

Our analyses did not reveal any linear correlation between the NAC signal and the targeted organelles. The co-localization was evaluated using the Pearson correlation coefficient in the ZEN software, examining both the entire image and selected regions of interest. Averaging the data across measured cells for whole image data of each experiment yielded Pearson coefficients of 0.19, 0.03, and 0.10 for mitochondria, GA, and ER, respectively. Next, we analysed regions of interest, that yielded average coefficients of -0.7, 0.23, and -0.28 for

mitochondria, GA, and ER, respectively. The assessments of signal profile within these regions further corroborated the absence of significant correlation (Fig. 4A).



**Figure 4** Analysis of NAC $\beta$  subunits localization. A. Signal profile of colocalization of NAC $\beta$ 1 with the endoplasmic reticulum (ER), mitochondria (Mt), and Golgi apparatus (GA) in *A. thaliana* leaves. The profile does not show colocalization of the signals. B. Split-GFP assay to determine NAC $\beta$ 's localization within *A. thaliana* protoplasts. The only positive signal was observed in combination with GFP<sub>1-10</sub>, suggesting that NAC $\beta$  subunits localize exclusively to the cytoplasm and nucleus.

### Split GFP assay to assess NAC $\beta$ 's subcellular localisation to cytoplasm

To clarify the discrepancies in the localization data of NAC $\beta$  in *Arabidopsis thaliana* compared to *Oryza sativa*, where NAC is reported to localize to the chloroplast in *O. sativa* (X. J. Liu et al. 2021) but not confirmed in *A. thaliana* through localization assays and proteomic analysis, we performed a split-GFP assay in *A. thaliana* chloroplasts. For this assay, NAC $\beta$  was tagged with GFP<sub>11</sub>, while various organelle markers were fused with GFP<sub>1-10</sub>. The organelle markers included a free GFP<sub>1-10</sub> as a cytoplasmic marker, MGD1 (AT4G31780) intermembrane space of chloroplast, pFlb (plant ferredoxin-like protein, AT1G10960) for chloroplast stroma, Tim50 for mitochondria intermembrane space (AT3g46560), and ATPb (AT5g08670) for mitochondria matrix. Additionally, we employed an mCherry-SKL to verify successful protoplast infiltration.

Three independent transformations were conducted, and in each, over 200 protoplasts for each construct were observed. Despite this analysis, we detected a signal only for NAC $\beta$  localization to the cytoplasm (Fig. 4B). This evidence further substantiates the cytoplasmic localization of NAC complex.



### **The knock down of NAC $\beta$ subunits leads to accumulation of organelle proteins in aggregates**

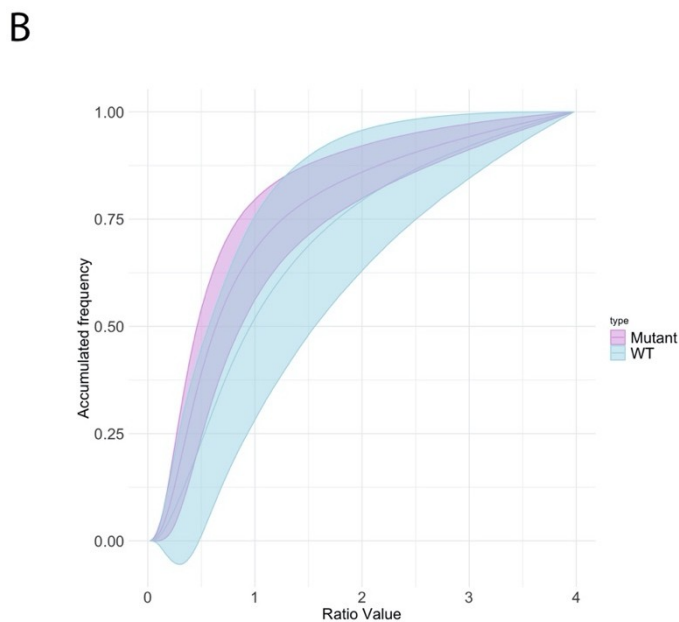
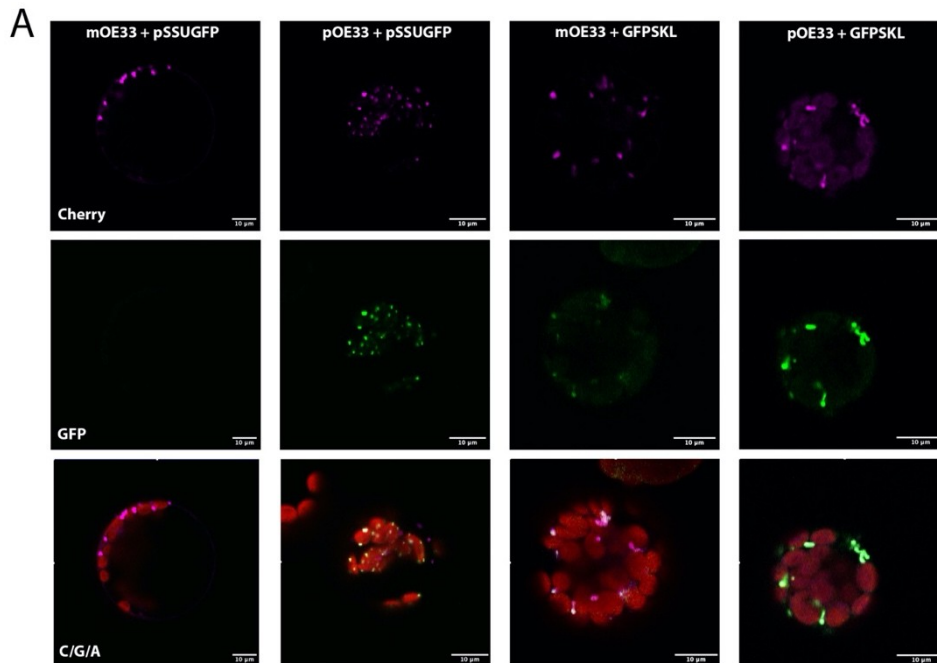
Previous research indicated that silencing of the NAC complex in yeast resulted in a higher accumulation of protein aggregates, particularly ribosomal proteins. To investigate if this phenomenon extends to *Arabidopsis thaliana*, we isolated and analysed protein aggregates from the leaves and flower buds of Col-0 wild-type (WT) plants and *nac $\beta$ 1/nac $\beta$ 2* mutants and analysed their composition by LC-MS/MS. This comparison aimed to determine whether NAC $\beta$  silencing in *A. thaliana* leads to a higher incidence of protein aggregation and/or of aggregation of specific protein groups.

In the analysis of flower buds, we identified 7,251 proteins in at least two replicates, each based on a minimum of two peptides. Among these, 260 proteins were significantly more aggregated in the *nac $\beta$ 1/nac $\beta$ 2* mutant. Similarly, in the leaves, we identified 2,919 proteins that passed the significance threshold, with 348 of these showing increased aggregation in the mutant.

The aggregated proteins predominantly consisted of those with various functions. Notably, a significant fraction was localized to the chloroplasts and mitochondria in both the flower buds and leaves. Specifically, chloroplast-localized proteins comprised 112 (25.3%) of the up-regulated proteins in the leaves. In the flower buds, 69 (26%) of the proteins were identified as chloroplast-residing. Mitochondrial proteins were represented by 88 proteins. Other prominently represented localizations included the plasma membrane (83 proteins) and the cytosol (72 proteins).

### **The knock down of NAC $\beta$ subunits does not lead to defects in the photosynthetic performance**

In our previous article describing the *nac $\beta$ 1nac $\beta$ 2* mutant phenotype, we discovered that the silencing of the NAC $\beta$  subunits leads to lower abundance of photosynthetic proteins, specifically parts of Light harvesting complex II (LHCII) including for instance two photosystem repair proteins D1 (AtC00020) and D2 (AtC00270) (Fíla et al. 2020). Furthermore, the mutant plants had a significantly lower amount of total chlorophyll in their leaves when normalised on weight of the measured tissue (Fíla et al. 2020). To test whether these defects influence photosynthetic efficiency of the mutant plants, we measured light response curves of chlorophyll-a fluorescence parameters such as linear electron transport, photoprotective non-photochemical quenching (NPQ), and the redox state of the PSII acceptor side (qL), light response curves of Photosystem-I-related parameters including its donor- and acceptor-side limitations, and of plastocyanin redox reactions, light response curves of the redox state of cytochrome f, light-induced proton motive force (pmf) across the thylakoid membrane, ATP synthase activity and in the isolated thylakoids, absolute complex contents per chlorophyll and compared them to the values obtained from Col-0 WT. Also, two complemented lines, for NAC $\beta$ 1 and NAC $\beta$ 2 were included in the analyses. Averaged values for 5 independent measurements were compared. However, no significant difference in any measured values were found between Col-0 WT, both complemented lines and *nac $\beta$ 1/nac $\beta$ 2* plants (Supplementary Tab. 1).



**Figure 5** *In vivo* chloroplast import assay comparing the efficiency of protein import into chloroplasts between Col-0 WT and *nacbeta1/nacbeta2* mutant. A. Combination of constructs used in the assay, showing their expected localization as described in Sommer et al., 2013. The C/G/A channel represents the merge of the Cherry, GFP, and chloroplast autofluorescence channels. B. Plot of the cumulative frequency of GFP/Cherry signal ratio from *pOE33 + GFP-SKL* expressing protoplast from WT Col-0 and *nacbeta1/nacbeta2*. The data were obtained from at least 20 protoplasts.

### ***In vivo* chloroplast import assay**

To evaluate the efficiency of protein import into chloroplasts in *nacβ1/nacβ2* mutants compared to wild-type (WT) plants, we conducted an *in vivo* sorting assay in mesophyll protoplasts as described by Sommer et al. (2013).

Briefly, the primary component, the small rubisco subunit pSSU, contains a pre-sequence that directs it to the chloroplast and is fused on its C-terminus with mCherry and GFP<sub>1-10</sub> fluorophores, along with the SKL peroxisomal sorting signal. Protoplasts are co-transformed with a GFP<sub>1-10</sub>-SKL construct. Efficient chloroplast targeting results in the majority of pSSU being transported to the chloroplast, yielding a high GFP but low mCherry signal throughout the protoplast. Conversely, if the import is impaired, more pSSU accumulates in peroxisomes, leading to a higher mCherry and lower GFP signal. Therefore, the GFP to mCherry signal ratio can be used to determine the efficiency of pSSU protein sorting to chloroplasts; the higher the ratio, the more disrupted the import is. The mSSU was used as a positive control for the system, as utilized by Sommer *et al.* (2013).

Data from individual chloroplasts were normalized to the background and averaged for each genotype. The results were visualized as the cumulative frequency of each normalized GFP/mCherry ratio between 0 and 1. Control experiments confirmed the correct localization of each construct, as previously described, verifying the credibility of the method (Fig. 5). However, on average, there was no significant difference observed in the import efficiency of pSSU to chloroplasts in *nacβ1/nacβ2* mutants compared to WT protoplasts (Fig. 5).

## Discussion

### **NAC is part of the cellular chaperone and sorting network**

To acquire insights into the potential interactome of NAC and the pathways that it is potentially involved in, we co-immunoprecipitated proteins with both NACβ1 and NACβ2 subunits from *Arabidopsis thaliana* flower buds in two independent lines—those expressing NACβ tagged with GFP and those expressing NACβ tagged with FLAG epitope. A significant overlap of nearly 200 proteins reported a list of proteins with a high fidelity of their functional connections to NAC. Not surprisingly, the majority of these included ribosomal and translation-associated proteins. As previously reported in yeast and by computational modelling, NAC is bound to ribosomes, and ribosomal binding sites are conserved in plant NAC subunits, leading to the plausible conclusion that NAC is involved in translation also in plants. The set of co-immunoprecipitated proteins in both GFP and FLAG lines includes all five NACα subunits, suggesting co-immunoprecipitation is consistent across the lines used, aligning with prior findings in yeast (Wiedmann *et al.* 1994) and our protein-protein interaction study results (Klodová 2019).

The co-immunoprecipitation of numerous proteins residing in chloroplasts and mitochondria, including photosystem subunits or inner membrane translocons, is a bit confusing. However, it may be caused by the firm interactions inside such complexes, allowing for their collective co-immunoprecipitation. Furthermore, it's plausible that after the disruption of the cellular compartments during protein isolation, the ability of NAC to accommodate various substrates by itself and its potential interaction with other chaperones can cause the crosslinking of complexes, which would normally be separated by residing in different cellular compartments. These interactions can thus be considered as non-specific.

The list also included several cytoplasmic, chloroplastic, mitochondrial, and endoplasmic reticulum chaperones, along with proteins involved in sorting pathways to these compartments. Overall, we estimate that NAC interacts with and hands over the nascent polypeptide to multiple chaperone and translocon pathways. For instance, in the FLAG interactome, we co-immunoprecipitated the whole *A. thaliana* chaperonin T-complex, which plays a role in the folding of actin or tubulin proteins and potentially a wider range of substrates (Ahn et al. 2019). Components of the HSP70/HSP40 complex, including 4 cytoplasmic HSP70s and also chloroplastic and mitochondrial chaperones (for instance chloroplastic HSP70-1 and mitochondrial HSP60 chaperonin) are crucial for folding and importing targeted proteins to chloroplasts or mitochondria (Murcha et al. 2015; Su and Li 2010). Chaperones localising to the endoplasmic reticulum that are involved secretory pathways, such as BIP1 and others, were also present. SRP54, which competes with NAC at the ribosomal exit binding site (Hsieh et al. 2020) was also found in the interactome. Interactions of NAC complex with secretory proteins have been previously reported (del Alamo et al. 2011). In archaea, the pull-down of NAC $\alpha$  also consisted of other chaperones. We have confirmed the direct interaction of several NAC subunits with HSP70 cytoplasmic chaperones by Y2H or BiFC and one 14-3-3 protein, linking NAC with the organelle sorting chaperone system. NAC was shown to co-localize with chloroplast translocon subunit TOC64-III and also showed positive signal interaction in BiFC. However, we did not obtain positive results for FRET-AB analyses. There are two possible explanations for that: (i) the interaction of TOC64-III with NAC can be very dynamic and short-lived. (ii) Alternatively, the interaction may be facilitated not directly, but by the transferred polypeptide, and thus the distance or orientation of the two proteins may not be sufficient for FRET. Interestingly, silencing of the transporters is usually not lethal, as alternative pathways can supplement the loss, but it does cause defects in respective organelles, such as pale flowers with impaired chloroplast protein import and reduced photosynthesis rates in case of TOC64-III silencing (Sommer et al. 2013). In the follow-up experiments, we will conduct FRAP analysis of the NAC subunits to establish the dynamics of NAC inside the cytoplasm.

Overall, these data suggest the involvement of NAC $\beta$  in cytoplasmic translation, potentially through direct association with ribosomes, as previously indicated in the literature (Pech et al. 2010; Wegrzyn et al. 2006). The data also point to its role in conjunction with other chaperones and in protein sorting to various organelles. These findings do not support a role for NAC in the nucleus as a transcription regulator, as was previously reported for NAC function in human (Y. Liu et al. 2010).

### **NAC complex likely plays a role in translation regulation under stress conditions**

We hypothesized that silencing of NAC $\beta$  could lead to increased protein aggregation, similar to observations made in yeast (Kirstein-Miles et al. 2013; Shen et al. 2019) To investigate this, we analysed protein aggregates and compared them between *nac $\beta$ 1/nac $\beta$ 2* mutants and Col-0 WT plants across two tissues: flower buds and leaves. The difference was more pronounced in the leaves.

It is not surprising that the composition of the aggregates encompasses a wide range of functional protein categories. This is consistent with reports that NAC can accommodate different protein structures (Martin et al. 2018) has been associated with various types of mRNAs serving different functions and is potentially involved in multiple protein sorting

pathways. Notably, a high proportion of aggregated proteins enriched in the mutants were targeted to chloroplasts, supporting the role of NAC in chloroplast protein targeting. However, the overall protein aggregation profile does not appear to change dramatically in case of NAC $\beta$  knockdown. This suggests that the functions of NAC $\beta$  may be compensated for by other chaperones, and differences might become more evident under stress conditions.

Interestingly, protein aggregation has been studied in *C. elegans*, where researchers observed how the translation machinery worked in the presence and absence of both NAC subunits under both acute and chronic stress (Kirstein-Miles et al. 2013). They made a notable discovery: while the abundance of NAC complex was unaltered by stress, its relocalization to aggregates formed during stress led to a subsequent decrease in translation efficiency. Furthermore, the presence of NAC facilitated to restore the function of the translation machinery (the polysome fraction) after recovery from an acute heat stress. In contrast, mutants lacking NAC complex were much less efficient in restoring original polysome fractions. This suggests that NAC functions as a regulator of translation, and the sequestration of NAC into protein aggregates during stress responses may fine-tune the mechanism for regulating the translation rate during stress response.

Therefore, it may be more insightful to investigate protein aggregation in *nac* mutants and wild types under stress conditions in conjunction with polysome profiling to assess differences in polysome distribution and, by extension, translation efficiency. We hypothesize that aggregated NAC proteins may be found in WT plants under stress, accompanied by a lower fraction of polysomes. However, unexpectedly, a lesser impact on the translation machinery may be observed in the *nac $\beta$ 1/nac $\beta$ 2* mutants, as the regulatory mechanism that slows translation machinery during stress could be absent, likely resulting in alternative protein folding mechanisms compensating for the lack of NAC and thus not altering the polysome fraction.

### **NAC silencing does not lead to defects in photosynthesis performance**

To investigate whether the pale leaf phenotype, the downregulation of photosynthesis-related proteins—including components of both photosystems—and the upregulation of cytoplasmic chaperones, coupled with delayed growth observed in *A. thaliana nac $\beta$ 1/nac $\beta$ 2* double homozygous mutants, could be indicative of reduced photosynthetic efficiency, we assessed the photosynthetic performance of the mutants in more detail. Despite our hypothesis, no differences were observed between the *nac $\beta$ 1/nac $\beta$ 2* mutant and wild type in any measured parameters, suggesting that the plants might compensate for the down-regulated proteins in other ways (Supplementary Tab. 1).

Possible compensatory mechanisms for the downregulation of photosynthetic proteins could include photorelocation of chloroplasts or increased energy investment into photosynthesis, accompanied by the upregulation of additional chaperones. In the knockdown *nac $\beta$ 1/nac $\beta$ 2* mutant, these adjustments could suffice to offset the reduced abundance of some photosystem proteins. Notably, mutants with a downregulation in the stress-related refolding chaperone Universal Stress Protein (USP, AT3G53990) also exhibited lower chlorophyll content, while its overexpression led to an increase in chlorophyll content (Jung et al. 2015). This suggests that USP encodes for proteins that might influence chlorophyll biosynthesis or stability. The USP protein was found to be up-regulated in the

*nacβ1/nacβ2* mutant. One explanation for the up-regulation could have a compensatory role in NACβ's function.

The challenges in protein delivery to chloroplasts and the lower chlorophyll content, therefore, could be counterbalanced by various cellular mechanisms, and do not necessarily result in diminished photosynthetic performance. To further explore whether disparities in plant coloration and chlorophyll content could stem from differences in the organization, structure, or number of chloroplasts within the cells, we prepared samples of the mutant and Col-0 WT plants for analysis with a transmission electron microscope. The analyses however was not finished at the time of thesis submission.

### **NAC works as a branch point of protein targeting by interacting with different pathways**

In recently published article about rice (*O. sativa*), the colocalization of NACβ with chloroplast was reported (Liu et al. 2021). The authors also observed disturbed chloroplast development and lower numbers of pollen grains in the Osj10gBTF3 mutant of rice. They proposed that NAC delivers the cargo with the TIC and TOC translocons directly to the chloroplasts. However, we find this model less probable for *A. thaliana* due to several reasons. (i) First, NACβ was not reported to be identified in any *A. thaliana* chloroplast proteomes (Ferro et al. 2003; Kleffmann et al. 2004; Wang et al. 2016). (ii) Second, we have confirmed its previously reported cytoplasmic localisation by split-GFP assay. (iii) Third, being of bacterial origin, which lack NAC complex as a whole, chloroplasts possess most likely an alternative folding machinery. However, it may be still possible that different mechanisms of sorting were adapted in *O. sativa* and *A. thaliana*. We propose that NAC complex is bound to ribosomes in plants in the same manner as was observed in animals and yeast. There it interacts with the nascent polypeptides and competes with SRP particle, by which it influences the secretory protein pathway. Furthermore, NAC is the docking site for other chaperones, which further influence the protein sorting. However, NAC complex does not change its localisation to the compartments, it more likely transfers the nascent peptide to other chaperones, such as HSP70 or 14-3-3 or translocons, such as TOC64-III. Moreover, NAC silencing does not lead to significant disruption in protein sorting, as it can be complemented by alternative pathways and chaperones, which are up-regulated in the absence of NAC proteins.

Generally, the lack of NAC complex results in developmental defects in plants, particularly in stages that require efficient translation, like pollen tube growth. This is attributed to less effective translation processes that necessitate compensatory mechanisms due to the absence of NACβ, rather than a defect in a specific developmental pathway. The absence of NAC thus leads in general delay in development especially during processes highly demanding for functional translation, including pollen tube growth, due to the less efficient translation and requirement for NAC absence compensation but not in a specific defect in one pathway.

## **Supplementary material:**

**Supplementary File 1:** List of primers used in this study, including names, sequences, and purposes of use.

**Supplementary File 2:** Results of co-immunoprecipitation of NAC $\beta$  subunits. The file includes 6 tables. The first describes proteins presented in all four analyses (NAC $\beta$ 1-FLAG, NAC $\beta$ 2-FLAG, NAC $\beta$ 1-GFP, NAC $\beta$ 2-GFP, „common\_all“), and the second table includes proteins shared between NAC $\beta$ 1-FLAG and NAC $\beta$ 2-FLAG („common\_FLAG“). These tables contain columns including protein UniProt identifiers, gene symbols, Ensembl gene identifiers, descriptions, and categorizations based on Gene Ontology. The other four tables include lists of proteins for each independent analysis with UniProt protein identifiers and descriptions.

**Supplementary Table 1:** Table showing results of the analysis of photosynthetic performance of WT, *nac $\beta$ 1/nac $\beta$ 2*, and complemented lines. No significant difference was observed in any measured aspect.

## Bibliography

- Ahn, H. K., Yoon, J. T., Choi, I., Kim, S., Lee, H. S., & Pai, H. S. (2019). Functional characterization of chaperonin containing T-complex polypeptide-1 and its conserved and novel substrates in Arabidopsis. *Journal of Experimental Botany*, 70(10), 2741–2757. <https://doi.org/10.1093/jxb/erz099>
- Billey, E., Hafidh, S., Cruz-gallardo, I., Jr, C. G. L., Jean, V., Carpentier, M., Picart, C., Kumar, V., Kulichova, K., & Mare, E. (2021). LARP6C orchestrates posttranscriptional reprogramming of gene expression during hydration to promote pollen tube guidance. 2637–2661. <https://doi.org/10.1093/plcell/koab131>
- Bloss, T. A., Witze, E. S., & Rothman, J. H. (2003). Suppression of CED-3-independent apoptosis by mitochondrial betaNAC in Caenorhabditis elegans. *Nature*, 424(6952), 1066–1071. <https://doi.org/10.1038/nature01920>
- Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*, 16(6), 735–743. <https://doi.org/10.1046/j.1365-313X.1998.00343.x>
- del Alamo, M., Hogan, D. J., Pechmann, S., Albanese, V., Brown, P. O., & Frydman, J. (2011). Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes. *PLoS Biology*, 9(7). <https://doi.org/10.1371/journal.pbio.1001100>
- Deng, J. M., & Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Research*, 4(4), 264–269. <https://doi.org/10.1007/BF01969120>
- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., & Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400(6745), 693–696. <https://doi.org/10.1038/23301>
- Don, E. K., Maschirow, A., Radford, R. A. W., Scherer, N. M., Vidal-itriago, A., Hogan, A., Maurel, C., Formella, I., Stoddart, J. J., Hall, T. E., Lee, A., Shi, B., Cole, N. J., Laird, A. S., Badrock, A. P., Chung, R. S., & Morsch, M. (2021). *In vivo Validation of Bimolecular Fluorescence Complementation ( BiFC ) to Investigate Aggregate Formation in Amyotrophic Lateral Sclerosis ( ALS )*. 2061–2074.
- Duttler, S., Pechmann, S., & Frydman, J. (2013). Principles of cotranslational ubiquitination and quality control at the ribosome. *Molecular Cell*, 50(3), 379–393.

- <https://doi.org/10.1016/j.molcel.2013.03.010>
- Ferro, M., Salvi, D., Brugière, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J., & Rolland, N. (2003). Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. *Molecular & Cellular Proteomics : MCP*, 2(5), 325–345. <https://doi.org/10.1074/mcp.M300030-MCP200>
- Fíla, J., Klodová, B., Potěšil, D., Juříček, M., Šesták, P., Zdráhal, Z., & Honys, D. (2020). The beta subunit of nascent polypeptide associated complex plays a role in flowers and siliques development of *arabidopsis thaliana*. *International Journal of Molecular Sciences*, 21(6), 1–30. <https://doi.org/10.3390/ijms21062065>
- Gamerding, M., Hanebuth, M. A., Frickey, T., & Deuerling, E. (2015). The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science (New York, N.Y.)*, 348(6231), 201–207. <https://doi.org/10.1126/science.aaa5335>
- Gamerding, M., Kobayashi, K., Wallisch, A., Kreft, S. G., Sailer, C., Schlömer, R., Sachs, N., Jomaa, A., Stengel, F., Ban, N., & Deuerling, E. (2019). Early Scanning of Nascent Polypeptides inside the Ribosomal Tunnel by NAC. *Molecular Cell*, 1–11. <https://doi.org/10.1016/j.molcel.2019.06.030>
- Grefen, C., & Blatt, M. R. (2012). A 2in1 cloning system enables ratiometric bimolecular fluorescence complementation (rBiFC). *BioTechniques*, 53(5), 311–314. <https://doi.org/10.2144/000113941>
- Groß, L. E., Machettira, A. B., Schleiff, E., & Sommer, M. S. (2011). GFP-based in vivo protein topology determination in plant protoplasts. *Journal of Endocytobiosis and Cell Research*, 89–97.
- Hartl, F. U., & Hayer-Hartl, M. (2002). Protein folding. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science*, 295(5561), 1852–1858. <https://doi.org/10.1126/science.1068408>
- Hsieh, H. H., Lee, J. H., Chandrasekar, S., & Shan, S. ou. (2020). A ribosome-associated chaperone enables substrate triage in a cotranslational protein targeting complex. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-19548-5>
- Jung, Y. J., Melencion, S. M. B., Lee, E. S., Park, J. H., Alinapon, C. V., Oh, H. T., Yun, D. J., Chi, Y. H., & Lee, S. Y. (2015). Universal stress protein exhibits a redox-dependent chaperone function in *arabidopsis* and enhances plant tolerance to heat shock and oxidative stress. *Frontiers in Plant Science*, 6(DEC), 1–11. <https://doi.org/10.3389/fpls.2015.01141>
- Kirchhoff, H., Mukherjee, U., & Galla, H.-J. (2002). Molecular Architecture of the Thylakoid Membrane: Lipid Diffusion Space for Plastoquinone. *Biochemistry*, 41(15), 4872–4882. <https://doi.org/10.1021/bi011650y>
- Kirstein-Miles, J., Scior, A., Deuerling, E., & Morimoto, R. I. (2013). The nascent polypeptide-associated complex is a key regulator of proteostasis. *EMBO Journal*, 32(10), 1451–1468. <https://doi.org/10.1038/emboj.2013.87>
- Kleffmann, T., Russenberger, D., Von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W., & Baginsky, S. (2004). The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Current Biology*, 14(5), 354–362. <https://doi.org/10.1016/j.cub.2004.02.039>
- Klodová, B. (2019). Studying dimer formation and effectors of *Arabidopsis thaliana* nascent polypeptide-associated complex. *Diploma Thesis, Charles University*.
- Koplin, A., Preissler, S., Llina, Y., Koch, M., Scior, A., Erhardt, M., & Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated



- complex on ribosomes. *Journal of Cell Biology*, 189(1), 57–68.  
<https://doi.org/10.1083/jcb.200910074>
- Kramer, G., Boehringer, D., Ban, N., & Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Structural & Molecular Biology*, 16(6), 589–597. <https://doi.org/10.1038/nsmb.1614>
- Lesnik, C., Cohen, Y., Atir-Lande, A., Schuldiner, M., & Arava, Y. (2014). OM14 is a mitochondrial receptor for cytosolic ribosomes that supports co-translational import into mitochondria. *Nature Communications*, 5, 1–10.  
<https://doi.org/10.1038/ncomms6711>
- Liu, X. J., Sun, J., Huang, Y., Li, C., Zheng, P., Yuan, Y., Chen, H., Jan, M., Zheng, H., Du, H., & Tu, J. (2021). Osj10gBTF3-Mediated Import of Chloroplast Protein Is Essential for Pollen Development in Rice. *Frontiers in Plant Science*, 12(August).  
<https://doi.org/10.3389/fpls.2021.713544>
- Liu, Y., Hu, Y., Li, X., Niu, L., & Teng, M. (2010). The crystal structure of the human nascent polypeptide-associated complex domain reveals a nucleic acid-binding region on the NACA subunit. *Biochemistry*, 49(13), 2890–2896. <https://doi.org/10.1021/bi902050p>
- Ma, H.-Z., Liu, G.-Q., Li, C.-W., Kang, G.-Z., & Guo, T.-C. (2012). Identification of the TaBTF3 gene in wheat (*Triticum aestivum* L.) and the effect of its silencing on wheat chloroplast, mitochondria and mesophyll cell development. In *Biochemical and biophysical research communications* (Vol. 426).  
<https://doi.org/10.1016/j.bbrc.2012.08.137>
- Macario, A. J., & Conway De Macario, E. (2001). The molecular chaperone system and other anti-stress mechanisms in archaea. *Frontiers in Bioscience : A Journal and Virtual Library*, 6, D262-83.
- Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I., & Koonin, E. V. (1999). *Comparative Genomics of the Archaea (Euryarchaeota): Evolution of Conserved Protein Families, the Stable Core, and the Variable Shell*. 608–628.
- Markesich, D. C., Gajewski, K. M., Nazimiec, M. E., & Beckingham, K. (2000). bicaudal encodes the Drosophila beta NAC homolog, a component of the ribosomal translational machinery\*. *Development (Cambridge, England)*, 127(3), 559–572.  
<http://www.ncbi.nlm.nih.gov/pubmed/10631177>
- Martin, E. M., Jackson, M. P., Gamerding, M., Gense, K., Karamonos, T. K., Humes, J. R., Deuerling, E., Ashcroft, A. E., & Radford, S. E. (2018). Conformational flexibility within the nascent polypeptide-associated complex enables its interactions with structurally diverse client proteins. *Journal of Biological Chemistry*, 293(22), 8554–8568.  
<https://doi.org/10.1074/jbc.RA117.001568>
- Murcha, M. W., Narsai, R., Devenish, J., Kubiszewski-Jakubiak, S., & Whelan, J. (2015). MPIC: A mitochondrial protein import components database for plant and non-plant species. *Plant and Cell Physiology*, 56(1), e10. <https://doi.org/10.1093/pcp/pcu186>
- Pech, M., Spreter, T., Beckmann, R., & Beatrix, B. (2010). Dual binding mode of the nascent polypeptide-associated complex reveals a novel universal adapter site on the ribosome. *Journal of Biological Chemistry*, 285(25), 19679–19687.  
<https://doi.org/10.1074/jbc.M109.092536>
- Planas-Marquès, M., Saul Lema, A., & Coll, N. S. (2016). Detection and quantification of protein aggregates in plants. *Methods in Molecular Biology*, 1450, 195–203.  
[https://doi.org/10.1007/978-1-4939-3759-2\\_15](https://doi.org/10.1007/978-1-4939-3759-2_15)
- Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction

- protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, 15(1), 8–15. <https://doi.org/10.1007/BF02772108>
- Porra, R. J., Thompson, W. A., & Kriedemann, P. E. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 975(3), 384–394. [https://doi.org/10.1016/S0005-2728\(89\)80347-0](https://doi.org/10.1016/S0005-2728(89)80347-0)
- Raue, U., Oellerer, S., & Rospert, S. (2007). Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence. *Journal of Biological Chemistry*, 282(11), 7809–7816. <https://doi.org/10.1074/jbc.M611436200>
- Rospert, S., & Chacinska, A. (2006). Distinct yet linked: Chaperone networks in the eukaryotic cytosol. *Genome Biology*, 7(3), 7–9. <https://doi.org/10.1186/gb-2006-7-3-208>
- Rott, M., Martins, N. F., Thiele, W., Lein, W., Bock, R., Kramer, D. M., & Schöttlera, M. A. (2011). ATP synthase repression in tobacco restricts photosynthetic electron transport, CO<sub>2</sub> assimilation, and plant growth by overacidification of the thylakoid lumen. *Plant Cell*, 23(1), 304–321. <https://doi.org/10.1105/tpc.110.079111>
- Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarolo, P., Blanca, J., Granell, A., & Orzaez, D. (2013). Goldenbraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiology*, 162(3), 1618–1631. <https://doi.org/10.1104/pp.113.217661>
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9, 676. <https://doi.org/10.1038/nmeth.2019>
- Schöttler, Mark A., Flügel, C., Thiele, W., Stegemann, S., & Bock, R. (2007). The plastome-encoded PsaI subunit is required for efficient Photosystem I excitation, but not for plastocyanin oxidation in tobacco. *Biochemical Journal*, 403(2), 251–260. <https://doi.org/10.1042/BJ20061573>
- Schöttler, Mark Aurel, Flügel, C., Thiele, W., & Bock, R. (2007). Knock-out of the plastid-encoded PetL subunit results in reduced stability and accelerated leaf age-dependent loss of the cytochrome b<sub>6</sub>f complex. *Journal of Biological Chemistry*, 282(2), 976–985. <https://doi.org/10.1074/jbc.M606436200>
- Schöttler, Mark Aurel, Kirchhoff, H., & Weis, E. (2004). The role of plastocyanin in the adjustment of the photosynthetic electron transport to the carbon metabolism in tobacco. *Plant Physiology*, 136(4), 4265–4274. <https://doi.org/10.1104/pp.104.052324>
- Shen, K., Gamerdinger, M., Chan, R., Gense, K., Martin, E. M., Sachs, N., Knight, P. D., Schlömer, R., Calabrese, A. N., Stewart, K. L., Leiendecker, L., Baghel, A., Radford, S. E., Frydman, J., & Deuerling, E. (2019). Dual Role of Ribosome-Binding Domain of NAC as a Potent Suppressor of Protein Aggregation and Aging-Related Proteinopathies. *Molecular Cell*, 729–741. <https://doi.org/10.1016/j.molcel.2019.03.012>
- Sommer, M., Rudolf, M., Tillmann, B., Tripp, J., Sommer, M. S., & Schleiff, E. (2013). Toc33 and Toc64-III cooperate in precursor protein import into the chloroplasts of *Arabidopsis thaliana*. *Plant, Cell and Environment*, 36(5), 970–983. <https://doi.org/10.1111/pce.12030>

- Su, P. H., & Li, H. M. (2010). Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. *Plant Cell*, 22(5), 1516–1531. <https://doi.org/10.1105/tpc.109.071415>
- Walsh, P., George, R., Lithgow, T., & Beddoe, T. (2002). The nascent polypeptide-associated complex (NAC) promotes interaction of ribosomes with the mitochondrial surface in vivo. *FEBS Letters*, 516(1–3), 213–216. [http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list\\_uids=11959135](http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=11959135)
- Wang, J., Yu, Q., Xiong, H., Wang, J., Chen, S., Yang, Z., & Dai, S. D. (2016). Proteomic insight into the response of Arabidopsis chloroplasts to darkness. *PLoS ONE*, 11(5), 1–24. <https://doi.org/10.1371/journal.pone.0154235>
- Wang, L., Zhang, W., Wang, L., Zhang, X. C., Li, X., & Rao, Z. (2010). Crystal structures of NAC domains of human nascent polypeptide-associated complex (NAC) and its  $\alpha$ NAC subunit. *Protein and Cell*, 1(4), 406–416. <https://doi.org/10.1007/s13238-010-0049-3>
- Wang, Y., Zhang, X., Lu, S., Wang, M., Wang, L., Wang, W., Cao, F., Chen, H., Wang, J., Zhang, J., & Tu, J. (2012). Inhibition of a basal transcription factor 3-like gene Osj10gBTF3 in rice results in significant plant miniaturization and typical pollen abortion. *Plant and Cell Physiology*, 53(12), 2073–2089. <https://doi.org/10.1093/pcp/pcs146>
- Wegrzyn, R. D., Hofmann, D., Merz, F., Nikolay, R., Rauch, T., Graf, C., & Deuerling, E. (2006). A conserved motif is prerequisite for the interaction of NAC with ribosomal protein L23 and nascent chains. *Journal of Biological Chemistry*, 281(5), 2847–2857. <https://doi.org/10.1074/jbc.M511420200>
- Wickner, W. (1995). The nascent-polypeptide-associated complex: Having a “NAC” for fidelity in translocation. *Proceedings of the National Academy of Sciences of the United States of America*, 92(21), 9433–9434. <https://doi.org/10.1073/pnas.92.21.9433>
- Wiedmann, B., Sakai, H., Davis, T. A., & Wiedmann, M. (1994). A protein complex required for signal-sequence-specific sorting and translocation. *Nature*, 370(6489), 434–440. <https://doi.org/10.1038/370434a0>
- Wiese, A. J., Steinbachová, L., Timofejeva, L., Čermák, V., Klodová, B., Ganji, R. S., Limones-Mendez, M., Bokvaj, P., Hafidh, S., Potěšil, D., & Honys, D. (2021). Arabidopsis bZIP18 and bZIP52 accumulate in nuclei following heat stress where they regulate the expression of a similar set of genes. *International Journal of Molecular Sciences*, 22(2), 1–24. <https://doi.org/10.3390/ijms22020530>
- Yang, S., Song, C., Zhang, X., Jia, Y., Shi, Y., Ding, Y., & Gong, Z. (2018). OST1-mediated BTF3L phosphorylation positively regulates CBFs during plant cold responses. *The EMBO Journal*, e98228. <https://doi.org/10.15252/embj.201798228>

## General discussion

Pioneer study of male gametophyte -omics, started with transcriptomics of *Arabidopsis thaliana* in 2003 describing the transcriptome of mature pollen grain and pinpointing male gametophyte-specific transcripts by comparing the data with the ones from sporophytic tissues. In 2004, the first transcriptomic data describing the dynamics of gene expression across the four stages of microgametogenesis established the main trends in pollen gene expression dynamics. Since then, -omics studies of male gametophyte have expanded substantially. Up to 2017, 19 transcriptomic and 6 proteomic studies have been conducted in *A. thaliana* (Fíla et al. 2017). Additionally, these studies have extended to various plant species, with 63 transcriptomic and 32 proteomic studies documented (Fíla et al. 2017).

Post-2017, the rapid accumulation of data has persisted. Methodological advancements have integrated more complex and sophisticated methods to describe additional steps in gene expression of male gametophyte, yielding more detailed analyses. With the advent of RNA-seq, analysis of alternative splicing and stage-specific isoforms has become feasible, alongside studies of both small and long non-coding RNAs (Golicz et al. 2021; Lohani et al. 2023; Nie et al. 2023; Rutley et al. 2020). Cell sorting techniques like FACS have enabled the separation of vegetative and sperm cells, facilitating the dissection of the processes occurring within each of the pollen nuclei (Misra et al. 2019, 2023; P. Qin, Loraine, and McCormick 2018). Furthermore, epigenetic regulation dynamics in developing pollen have been clarified through Chip-seq, ATAC-seq data and single cell RNA-seq (Borg et al. 2021; Ichino et al. 2022; Zhu et al. 2023). Multi-omics and comparative analyses have revealed new insights into gene expression dynamics and conservations of evolutionary pathway (Hafidh et al. 2018; Julca et al. 2020; Keller et al. 2018; Poidevin et al. 2020).

This thesis encompasses studies of the gene expression dynamics of male gametophyte development in *A. thaliana*. The ensuing discussion will summarize the primary outcomes, future prospects, and the expected challenges of male gametophyte research.

### Challenges of studying long non-coding RNAs in the male gametophyte

In our study, we described a remarkable increase in the availability of expression data for non-coding RNAs, highlighting the advantages of RNA-seq over the traditional Affymetrix platform (Klodová et al. 2023). We propose a significant role for long non-coding RNAs (lncRNAs) as regulators of male gametophyte development, identifying potential candidates based on their expression profiles.

The realm of non-coding RNA-seq, particularly its role in plant development, is increasingly gaining attention of the scientific community. A recent study on the gene expression dynamics of pollen development, with an emphasis on lncRNAs in *Brassica rapa*, aligns with our findings. This study analysed expression dynamics across five developmental stages: pollen mother cell, tetrad, microspore, bicellular pollen, and mature pollen. They identified 1,058 expressed long intergenic non-coding RNAs (lincRNAs) and 780 natural antisense transcripts (NATs) that overlap with protein-coding genes on the opposite strand (Lohani et al. 2023). They observed expression patterns in *B. rapa* pollen that are similar to what we found in *A. thaliana*. Generally, this includes lower levels of long non-coding RNAs (lncRNAs)

compared to protein-coding genes and distinct expression patterns that are specific to certain stages, suggesting that these patterns are synchronized with the overall expression program in developing pollen. The lncRNAs were categorized based on their interactions with mRNAs and microRNAs (miRNAs). They predicted 22 lincRNAs as potential targets of 18 miRNAs and 21 NATs as targets of 36 miRNAs. Ten lincRNAs and eight lncNATs were described as potential small RNA precursors. The analysis of miRNA-lncRNA-mRNA clusters serves as a robust tool to delineate regulatory pathways among different RNA groups.

However, the study of lncRNAs during microgametogenesis showed an additional challenge. There is a notable relationship between lncRNAs and transposable elements (TEs). While the evidence from plants is sparse and necessitates detailed analysis, several information on human and mouse lncRNAs are available. Often, TEs are integral to lncRNAs, with TE-derived sequences of lncRNAs exhibiting a higher conservation than non-TE-derived sequences (Kapusta et al. 2013). Reports indicate that 83% human and 66% mouse lncRNAs, respectively, contain at least one TE (Kelley and Rinn 2012). Conversely, only 5.5% of human protein-coding genes appear to be derived from TEs (Kelley and Rinn 2012). In general, lncRNAs are significantly enriched for TEs compared to protein-coding genes and TEs contribute to various functional capacities of lncRNAs (reviewed in (Fort, Khelifi, and Hussein 2021). For example, the tissue and expression specificity of some lncRNAs depend on their TE content (Kapusta et al. 2013; Kelley and Rinn 2012; Watanabe et al. 2015). TEs also play a role in lncRNA editing and influence their cellular localization (Chen, DeCervo, and Carmichael 2008; Lubelsky and Ulitsky 2018). Experiments involving mutations of TEs within lncRNAs have shown alterations in their function (Nguyen et al. 2020). Furthermore, TEs can facilitate protein or miRNA binding to lncRNAs (Fort et al. 2021). Thus, it is evident that TEs and lncRNAs are closely intertwined and lncRNAs are often functionally dependent on TEs.

Such a link is especially important in the context of the male gametophyte. During microgametogenesis, there is a transient relaxation of epigenetic repression, which leads to a vast TE activation. In the post-meiotic phase, CHH methylation in sperm cells is lost, resulting in the derepression of retrotransposons. The CHH methylation is only restored during embryo development (Calarco et al. 2012). Constitutive heterochromatin is lost along with histone replacement and DDM1 downregulation in the vegetative nucleus, leading to TE activation in early microsporogenesis (Calarco et al. 2012). This leads to the production of a high amount of 21 nt siRNAs, which accumulate in the sperm cells and possibly play a role in TE silencing in the germline (Slotkin et al. 2010).

The intricate nature of lncRNAs, which are typically characterized by their low abundance, precise regulation of expression, and diversity, combined with their minimal sequence conservation and complex interaction with TEs, presents a significant challenge in understanding the functional dynamics of lncRNAs and TEs. It complicates the determination of whether the observed genetic activities are attributable to lncRNAs or TEs.

Integrative approaches may offer solutions to overcome this challenge. For instance, employing advanced sequencing methods such as ChIP-seq or ATAC-seq to map epigenetic modifications and chromatin accessibility, in conjunction with RNA-seq data, can provide insights into TE activation. Such approach has recently been used to describe the high activity of TEs from pericentromeric regions upon the loss of heterochromatin in the

vegetative nucleus, undergoing transcription leading to the production of small non-coding RNAs (Borg et al. 2021).

Downstream analysis can distinguish the expression of lncRNAs and TEs by their potential differences in sequence properties. TEs often have target site duplications, terminal inverted repeats, and are usually flanked by direct repeats, whereas lncRNAs can exhibit features like polyadenylation and splice sites. Computational tools such as RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0.2013-2015 <<http://www.repeatmasker.org>>.), TETranscripts (Y. Jin et al. 2015), Repbase (Bao, Kojima, and Kohany 2015), and Cuffcompare (Trapnell et al. 2010) have been developed to discern TE-related sequences within RNA-seq data. Nonetheless, these resources are largely dependent on extant annotations of TEs. Although the databases cataloguing annotated TEs and ncRNAs do exist, the depth of functional annotation and the synchronization across databases for lncRNAs and TEs, particularly in plant species, are often insufficient. It is imperative that specialized bioinformatic tools will be developed to facilitate the differentiation between lncRNAs and TEs. Moreover, the establishment of comprehensive and standardized repositories for ncRNAs and TEs is paramount. Finally, functional assays of putative lncRNAs, employing reverse genetic approaches, would significantly enhance our understanding of their biological roles, with or without their association with TEs.

### **Translatomics filling the gap in translation dynamics**

In our study, and also in the study of tomato pollen heat stress response, the dynamics of mRNA translation have been inferred based on k-means clustering of transcriptomic and proteomic data across various pollen stages (Keller et al. 2018; Klodová et al. 2023). Combining transcriptomics with proteomics represents a powerful approach to analyse mRNA translation dynamics. However, the analysis offers only indirect evidence since the transcripts may undergo various post-transcriptional and post-translational modifications, including mRNA storage or alternative splicing. Moreover, the sensitivity of proteomic data in detecting low-abundant proteins can lead to misinterpretations of their absence (Yan et al. 2022).

Translatome analysis provides more direct information about the mRNA translation rate and specificity, particularly in processes such as pollen and seed development, where rapid changes in translation dynamics and mRNA storage play a pivotal role. Consequently, the correlation rate between the transcriptome and proteome can be generally very low. The translation machinery evolves throughout development from an active, polysome-enriched state to a quiescent state with a dominant monosome fraction. Understanding the role of mRNAs stored in the form of inactive monosomes can yield important information about the processes required for pollen tube germination, growth and fertilisation.

A method for co-immunoprecipitation of ribosome-associated mRNAs using FLAG-tagged ribosomal proteins was developed and facilitates the analysis of tissue-specific translomes. Originally, an atlas of 21 specific translomes for shoot and root tissues was acquired using this method (Mustroph and Bailey-Serres 2010). To analyse pollen-specific translomes, the transgenic *A. thaliana* plants (LAT52:HF-RPL18) with an epitope-tagged ribosomal protein L18 driven by the pollen-specific proLAT52 promoter for mRNA co-immunoprecipitation can be utilised. This approach was employed to obtain polysomal

complexes from self-pollinated, non-pollinated styles, and *in vitro*-grown pollen tubes to understand the mRNA dynamics during pollen tube germination (Lin et al. 2014). The study identified 519 transcripts present exclusively in *in vivo*-grown pollen tubes, which may be vital in male-female communication during sperm cell delivery, as they were not found in the *in vitro* pollen tube samples or the pollen grain samples. In the subsequent studies, transcriptome dynamics during pollen germination were also studied in response to heat stress (Poidevin et al. 2020). A notable phenomenon emerged in this study when comparing the transcriptome and proteome of germinated pollen under the normal conditions: 1,316 genes were present in the transcriptome but absent from the proteome. Functional characterization and comparative analysis revealed that these genes consist largely of sperm cell-expressed genes, suggesting that during pollen germination, the translation of proteins, unrequired for the germination, is highly restricted. However, they found a relatively high overall correlation of transcriptome and proteome in developing pollen ( $R=0.88$ ) (Poidevin et al. 2020). This contrasts with reports from transcriptome and proteome correlations, which tend to be low due to the high mRNA storage dynamics in pollen (Baerenfaller et al. 2008; Maier, Güell, and Serrano 2009; Vogel and Marcotte 2012). Under heat stress, the authors reported conserved mechanisms applied during pollen germination including the unfolded protein response and cytosolic protein response (Poidevin et al. 2020). The dynamics of proteomes and their coordination with the transcriptome under various stresses (wounding, dark-light transition, heat, hypoxia, phosphate starvation, and pathogen attack) were studied in various tissues (Y. Chen, Liu, and Dong 2021). A high correlation between the transcriptome and proteome was observed during the transition from dark to light. However, for other stresses, differential coordination of transcription and translation was noted. Generally, upregulation of stress-related genes at both the transcriptional and translational levels was reported, although regulation differed for the downregulation of genes encoding proteins that are not involved in stress response. These were often regulated only at one level either transcription or translation. This mechanism exemplifies the fine-tuning of gene expression on multiple levels, potentially saving energy costs.

Polysomal profiling allows tracking the ratio between monosomal and polysomal fractions in the sample and enables their separation, which can then be processed with RNA-seq to analyze the proteome of each fraction and LC-MS/MS to obtain information on translation machinery-associated proteins. Implementing this approach in developing male gametophyte and pollen tubes can fill the missing link in our analysis. By acquiring the proteomes of the pollen stages and comparing them with transcriptomes and proteomes, we could dissect the dynamics and outcomes of mRNA storage. Proteome analysis could provide specificities and dynamics of the translation machinery in developing pollen. However, the challenge posed here comes with the material requirement, as acquiring enough RNA from pollen stages for monosomal and polysomal fraction analysis would be difficult in *Arabidopsis thaliana*. As a pilot experiment to fine-tune the methodology, we are currently analysing the polysomal and monosomal fractions of mature pollen and pollen tubes of *A. thaliana*. This analysis also lays the groundwork for comparing the similarities and differences between germinating seeds and pollen grains, as samples from dormant, imbibed and germinated seeds were also analysed.

## Dissecting the pollen gene expression in sperm cells

In our study (Klodová et al. 2023), we concentrated on analysing gene expression during four stages of microgametogenesis. We isolated the individual stages by Percoll gradient, followed by RNA or protein extraction and subsequent RNA-sequencing or LC MS/MS analysis, respectively. Our approach enabled to obtain samples encompassing the entire spore tissues, including transcripts (or proteins) from both the vegetative cell and sperm cells. However, it is conceivable that the contribution of the sperm cell transcriptome might have been overshadowed by the predominant abundance of transcripts from the vegetative cell, leading our data to primarily reflect gene expression in the vegetative cell, a major contributor to the overall pollen transcriptome (Borg et al. 2011).

Numerous studies have indicated that vegetative cells, and potentially even the two sperm cells, exhibit distinct structural, biochemical, genetic, and epigenetic characteristics (reviewed in (Flores-Tornero and Becker 2023)). The specificities of the sperm cell transcriptome have been explored in various studies and species. The pioneering genome-wide study using Affymetrix technology involved the separation of *A. thaliana* sperm cells via fluorescence-activated cell sorting (FACS), enabling the comparison of their transcriptome with that of whole pollen and seedlings to identify enriched and sperm cell-specific transcripts (Borges et al. 2008). This analysis identified 5,829 positive calls for sperm cell transcripts, including 642 sperm cell-specific and approximately 2,400 sperm cell-enriched transcripts. Functional analysis suggested their involvement in processes including DNA repair, ubiquitin-mediated proteolysis, and cell cycle progression, providing a list of candidates essential for male gamete development (represented by the sperm cells) and fertilization. The study also proposed the possibility of distinct epigenetic regulation in sperm cells. Specific expression profiles of sperm cells were confirmed also in *O. sativa*, with the identification of 1,622 genes not expressed in pollen or seedlings but present in sperm cells (Russell et al. 2012). In *P. zeylanica*, distinct profiles of the two dimorphic sperm cells were described using Affymetrix (Gou et al. 2009). Subsequent studies employing RNA-seq have provided higher resolution, enabling the description of miRNA dynamics and detailed mechanisms of epigenetic regulation in developing pollen (Borges et al. 2011; Calarco et al. 2012; Slotkin et al. 2010). A recent study provided an in-depth analysis of transcriptome and alternative splicing in mature pollen sperm cells, identifying 2,091 potentially novel transcripts in sperm cells and 3,600 transcripts in vegetative nuclei (Misra et al. 2023). Further analyses of sperm cell transcriptomes were also conducted in other plant species including *O. sativa* and *S. lycopersicum* (Anderson et al. 2013; L. Liu et al. 2018). Recently, single-cell RNA-seq has been employed to elucidate the dynamics of pollen gene expression. The initial study established a protocol for acquiring a single-cell transcriptome of *A. thaliana* sperm cells from *semi in vivo* grown pollen tubes (Misra et al. 2019). A subsequent study analysed developing pollen at the single-cell level, identifying TE suppressors MBD5, MBD6, and SILENZIO as crucial factors for TE inhibition in the vegetative nucleus, and providing an interactive map of the single-cell transcriptomics of various pollen stages (Ichino et al. 2022).

Methodological and resolution advancements in genome-wide analysis have paralleled progress in pollen stage isolation. Until recently, transcriptomic analysis of developing pollen was hampered by the need for high amounts of RNA for RNA-seq experiments, a challenge when using Percoll gradient stage separation due to insufficient amount of



material. However, novel methods for specific pollen stage separation have emerged. Recent innovations include the isolation of pollen from a single anther, termed single anther RNA-seq (SA RNA-seq). This technique successfully yielded clear fractions of UNM, BCP, and TCP pollen from *A. thaliana*, and UNM and BCP stage kiwifruit (*Actinidia chinensis*), providing sufficient material for RNA-seq from a limited number of plants (Le Lievre et al. 2023).

Collectively, the methodological advancements in cell isolation, cell type separation, and genome-wide 'omics' analyses, including scRNA-seq, have opened an array of possibilities for further advancements in understanding the dynamics of gene expression within the male gametophyte.

### **Pollen desiccation, the third peak in microgametogenesis expression**

The published transcriptomic and proteomic data can provide a valuable source of information for further research beyond the scale of interest of the original study. In the study of gene expression dynamics in the developing pollen, the focus was aimed towards transcription and translation regulations, which resulted in a detailed functional analysis and a categorisation of gene cluster connected to these processes (Honys and Twell 2004; Klodová et al. 2023). However, the data provide an additional huge wealth of information. One such area is the stress tolerance mechanisms related to the dehydration processes that accompany pollen desiccation and subsequent reactivation.

The male gametophyte exhibits an intriguing paradox: the activation of stress response pathways in the absence of exogenous stressors. For instance, germinating pollen expresses molecular chaperones—"heat shock proteins" (HSPs)—even in the absence of heat stress (Qin et al. 2009; Wang et al. 2008). Furthermore, pollen tubes traversing the pistil at basal temperatures initiate the endoplasmic reticulum (ER) stress response, a reaction traditionally associated with stressed vegetative tissues (Poidevin et al. 2020; Rahmati Ishka et al. 2018). Mutations in numerous ER-stress response genes result in male gametophyte lethality or defects in pollen germination or pollen tube elongation, underscoring their essential role (Fragkostefanakis et al. 2016; Singh, Lohani, and Bhalla 2021). All together, these experiments suggest that a heat stress-like response is crucial for pollen development and pollen tube germination and growth.

The molecular mechanism behind this phenomenon was addressed in a data mining study, where the authors implemented results of available transcriptomes and experimental data to understand the processes behind pollen desiccation and activation (Sze et al. 2021). This analysis is also a great example of how -omics data can be further exploited to provide a wealth of new information. Based on their analysis, the authors propose a model illustrating the behaviour of mRNP granules and the regulation of molecular processes that occur as pollen undergoes dehydration during maturation and subsequent rehydration during germination. They also propose the outcomes of external heat stress on different developmental stages, hypothesising the reasons behind the extreme sensitivity of male gametophyte to abiotic stresses (Rieu et al. 2017).

In a follow up study (recently accepted in *Plant Physiology*, Sze *et al.*) we expanded upon the model proposed by Sze et al., 2021, incorporating both proteome and transcriptome data to

investigate the expression dynamics throughout the developmental stages of male gametophyte. We identified a new gene cluster associated with pollen dehydration, exhibiting peak expression between the bicellular pollen (BCP) and tricellular pollen (TCP) stages. This cluster includes numerous potential candidates and processes implicated in pollen desiccation and reactivation, as well as the dynamics of these transcripts in mRNA storage.

### **The online resources for the analysis of gene expression in male gametophyte**

Multomics analyses represent an essential source for the development of powerful online tools and databases. The databases present analysed data in a user-friendly format, enabling the scientific community to access and utilize the information efficiently. Currently, over 300,000 RNA-seq experiments from more than 100 plant species are available. Transcriptomic data can be utilized, for instance, to predict the functions of novel genes, to understand the functional evolution within gene families, to decipher spatiotemporal regulatory networks, or to explore the impact of various conditions on transcriptomes (Lim et al. 2022). These results are obtained through downstream analyses and comparisons of original data executed by publicly available programs. A comprehensive summary and categorization of such plant transcriptomics databases have been recently published (Lim et al. 2022).

Integration and interpretation of data from these databases can provide a nearly infinite amount of information in a relatively straightforward computational manner, circumventing the need for tedious or expensive laboratory experiments, which are also often constrained to a limited range of model species. Proper utilization of available data can thus conserve resources and time in research, offering a wealth of novel insights. However, in my opinion, the use of publicly available datasets is still vastly underappreciated.

Below, there are three recently developed online resources that, in my opinion, provide useful insights into pollen expression dynamics. Moreover, the information provided by these platforms extends beyond pollen transcripts, making them valuable for numerous other plant research areas. To illustrate their significance, the expression of NAC $\beta$ 1 subunit and DAZ3 were analysed by these platforms (Fig. 5).

#### **EVOREPRO Database**

<https://evorepro.sbs.ntu.edu.sg>

This database offers a visualization tool for gene expression and co-expression analyses across various developmental stages, encompassing both male and female gametophytes and sporophytes within 13 plant species, including bryophytes, vascular plants, gymnosperms, and flowering plants, thus covering all major phylogenetic branches. The database incorporates RNA-seq data from a comprehensive comparative analysis study of land plant transcriptomes, which aimed to identify conserved genetic programs fundamental to development and reproduction across plant species (Julca et al. 2021). Expression profiles across tissues and developmental stages of gene of interest can be retrieved, downloaded, and compared. Genes are organized into co-expression networks, which can facilitate the determination of their potential functional categorization. Thus, genes specific or enriched in the male gametophyte can be easily identified, and their conservation and specificity can be compared between different species. Additionally,

phylogenetic trees can be imported from OrthoFinder databases and created, displaying expression values for selected stages.

### **Single-Nucleus RNA-seq of Developing *A. thaliana* Male Gametophytes**

[https://singlecell.mcdb.ucla.edu/snRNAseq\\_pollen/](https://singlecell.mcdb.ucla.edu/snRNAseq_pollen/)

In this study, the authors analysed single-cell transcriptomes for each nucleus during the development of pollen grains, including UNM, BCP, and TPC stages (Ichino et al. 2022). They identified 13 distinct clusters using UMAP visualization to distinguish the VC, GC, and sperm cells (SC) at individual stages. They validated the clusters by the presence of marker genes. The study highlighted the critical role of three TE repressors MBD5, MBD6, and SILENZIO in the vegetative nucleus of developing pollen. The online tool allows for the rendering of plots for a gene of interest using their scRNA-seq data. As a result, expression profiles of the gene of interest across the developmental stages can be distinguished between VN and GC or SC, as well as its expression in both Col-0 WT and the mutant lines used in their study—*met1*, *ddm1*, *mom1*, *morc*, *met1* and *ddm1* mutants exhibit significant loss of methylation and high derepression of TEs, whereas *mom1* and *morc* mutations lead to less TE derepression and limited loss of methylation. Changes in the expression of the gene of interest across these mutants can provide insights into its epigenetic regulation.

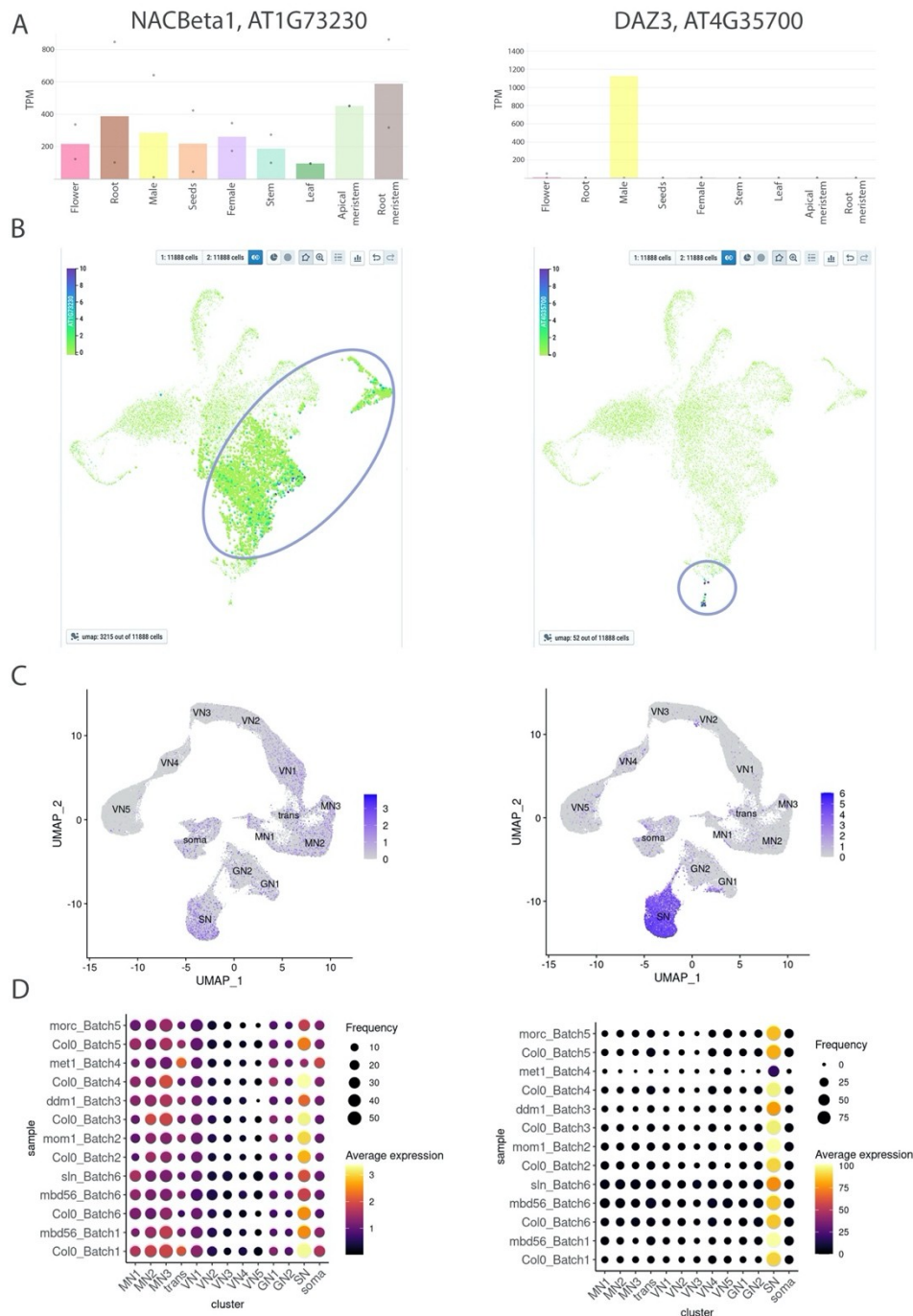
### **Arabidopsis Developmental Atlas**

<http://neomorph.salk.edu:9000>

In this study, a comprehensive "seed-to-seed" dataset of scRNA-seq data was created, spanning across 10 developmental stages including seeds, seedlings, rosettes, stems, flowers, and siliques. Unfortunately, the male gametophyte data are represented only as part of the flower transcriptome, which is divided into clusters including anther, epidermal cells, male meiocyte, pollen, tapetum, unknown (cluster of unassigned cells), and vascular tissues. However, this data is still valuable for understanding gene functions in the male gametophyte—for instance, deciphering the specificity between sporophytic influences (tapetum-expressed genes) or gametophytic influences (pollen/male meiocyte expressed genes). The gene of interest can be visualized at different stages using an easily understandable tool.

While these tools are based on transcriptomics, and the data for protein-coding genes are quite accessible, databases for protein expression, alternative splicing, or non-coding RNA data lag behind despite the nearly exponential increase in available datasets. This gap provides an area of interest for future enhancements.

Lastly, I would like to highlight the emerging role of artificial intelligence in data processing, interpretation, and visualization, which, in my opinion, cannot be overlooked and should be highly exploited and incorporated into future data research. To demonstrate how useful it is in bioinformatics, I asked to provide a pipeline for RNA-seq data analysis and an R script for differential gene expression including data visualisation. The results can be observed in Fig. 6.



**Figure 5** Demonstration how the EVOREPRO database, snRNA-seq pollen database, and Arabidopsis Developmental Atlas are used in the analysis of the expression of the pleiotropic NAC $\beta$ 1 (AT1G73230) and sperm cell-specific DAZ3 (AT4G35700). A. Expression profiles of tissue specificity retrieved from the EVOREPRO database show pleiotropic expression of NAC $\beta$ 1 and nearly exclusive expression of DAZ3 in male tissues. B. UMAP visualization of flower tissue from the Arabidopsis Developmental Atlas displays results from 11,888 cells distributed across 7 defined clusters. For NAC $\beta$ 1, expression in 3,215 epidermal cells is highlighted, whereas for DAZ3, the expression in 52 pollen cells is highlighted. C. UMAP visualization of snRNA-seq data from the analysis of three developmental pollen stages in individual cells, including vegetative cell nucleus (VN), sperm cell nucleus (SN), and generative cell nucleus (GN). The SN, where expression of DAZ3—a DUO1 target important for sperm cell differentiation—is most prominent, represents the sperm cell nucleus. Expression of NAC $\beta$ 1 is generally present in a lower number of cells and spans across SN, VN1 (vegetative nucleus of bicellular pollen), MN1, MN2, MN3 (nuclei of microspores), and the 'soma' fraction, which includes "some contaminating nuclei". The expression of NAC $\beta$ 1, encoding a cytoplasmic chaperone, corresponds with its expression data reported in Klodová et al., 2023, with predominant expression in the early stage of pollen development. D. Comparison of expression in different *A. thaliana* genotypes with their roles in epigenetic silencing, including met1, ddm1, mom1, and morc, across developmental pollen stages and nuclei. The highest influence on NAC $\beta$ 1 expression appears to be met1 silencing in SN. MET1 is also possibly involved in regulating the expression of DAZ3.

**Q:** provide a pipeline for stranded pair-end RNA-seq data of *Arabidopsis thaliana* bicellular pollen (3 biological replicas) and tricellular pollen (3 biological replicas) from the raw fasta files to differential gene expression including illustratory plots

**ChatGPT**

Creating a pipeline for stranded pair-end RNA-seq data involves several steps from the initial quality control to the final differential gene expression analysis and visualization. Below is a generalized pipeline that you can follow for the analysis of *Arabidopsis thaliana* bicellular and tricellular pollen with three biological replicates each:

- 1. Quality Control**
  - Use FastQC for initial quality checks on the raw FASTQ files.
  - Perform trimming of adapters and filtering of low-quality reads with Trimmomatic or Cutadapt.
- 2. Read Alignment**
  - Align the processed reads to the *Arabidopsis thaliana* reference genome using a splice-aware aligner like STAR or HISAT2.
  - Sort and index the resulting BAM files with Samtools.
- 3. Quantification of Gene Expression**
  - Use featureCounts (part of the Subread package) or HTSeq to count the number of reads mapping to each gene, taking into account the strand-specificity of the library.
- 4. Differential Expression Analysis**
  - Import read counts into R and use DESeq2 for differential expression analysis between bicellular and tricellular pollen samples.
  - Include normalization for library size and potential batch effects, especially if samples were sequenced in different runs.
- 5. Results Summarization and Visualization**
  - Use the DESeq2 results to summarize findings, such as a table of significantly differentially expressed genes with fold changes and p-values.
  - Generate illustrative plots such as:
    - MA plots showing the relationship between fold change and mean expression for each gene.
    - Volcano plots to visualize the significance versus the fold change of each gene.
    - Heatmaps of the top differentially expressed genes for an overall expression pattern comparison.
    - PCA plots to visualize the overall effect of treatment and potential batch effects.
- 6. Functional Enrichment Analysis (optional but recommended)**
  - Use tools like Goseq, DAVID, or clusterProfiler for functional enrichment analysis of differentially expressed genes to find over-represented biological processes or pathways.
- 7. Reporting**
  - Compile all findings, plots, and tables into a comprehensive report. This can be done using R Markdown for an integrated code and output format, which can then be exported to HTML or PDF.

```
r
library(DESeq2)
library(ggplot2)

# Load the count data (assuming it's in a CSV file)
counts <- read.csv("gene_counts.csv", row.names=1)

# Create a metadata dataframe indicating the conditions
metadata <- data.frame(
  condition = factor(c(rep("bicellular", 3), rep("tricellular", 3))),
  row.names = colnames(counts)
)

# DESeq2 Dataset setup
dds <- DESeqDataSetFromMatrix(countData = counts,
                              colData = metadata,
                              design = ~ condition)
dds <- DESeq(dds)

# Get the results
res <- results(dds)

# MA Plot
plotMA(res, main="MA Plot", ylim=c(-2,2))

# Volcano Plot
ggplot(as.data.frame(res), aes(x=log2FoldChange, y=-log10(pvalue))) +
  geom_point(aes(color = padj < 0.05)) +
  scale_color_manual(values = c("grey", "red")) +
  theme_classic() +
  labs(x = "Log2 Fold Change", y = "-Log10 P-value")

# Heatmap of top differentially expressed genes (requires pheatmap pack)
library(pheatmap)
select <- order(res$pvalue)[1:20]
mat <- assay(dds)[select,]
pheatmap(mat)

# PCA Plot
rld <- rlog(dds)
plotPCA(rld, intgroup="condition")
```

**Figure 6** Demonstration of AI platform ChatGPT 4.0 usage in the bioinformatic data analyses. The software was asked to provide a pipeline for RNA-seq data analyses and an R script for differential expression analysis.

## The role and functional conservation of the Nascent polypeptide associated complex in plants

One objective of the -omics studies is to provide lists of potential candidates linked to specific pathways or processes of interest. In a phosphoproteomic study of tobacco activated pollen, the family of nascent polypeptide-associated complex (NAC) emerged as a candidate with a potential role in pollen development or activation, as its  $\alpha$  subunits were specifically phosphorylated during pollen activation (Fíla et al. 2012). Consequently, they were chosen for further investigation in the context of male gametophyte development and fertilization. NAC appears to be a crucial protein family, present in all eukaryotes and archaea, maintaining high levels of sequence and structural conservation. In plants, several independent subunit duplications suggest potential subfunctionalization or neofunctionalization, although the core functions seem to be conserved, as evidenced by the ability of *Marchantia polymorpha* NAC $\beta$  subunits to complement mutation of those from *A. thaliana*. One possible explanation of emergence of the number of NAC subunits in

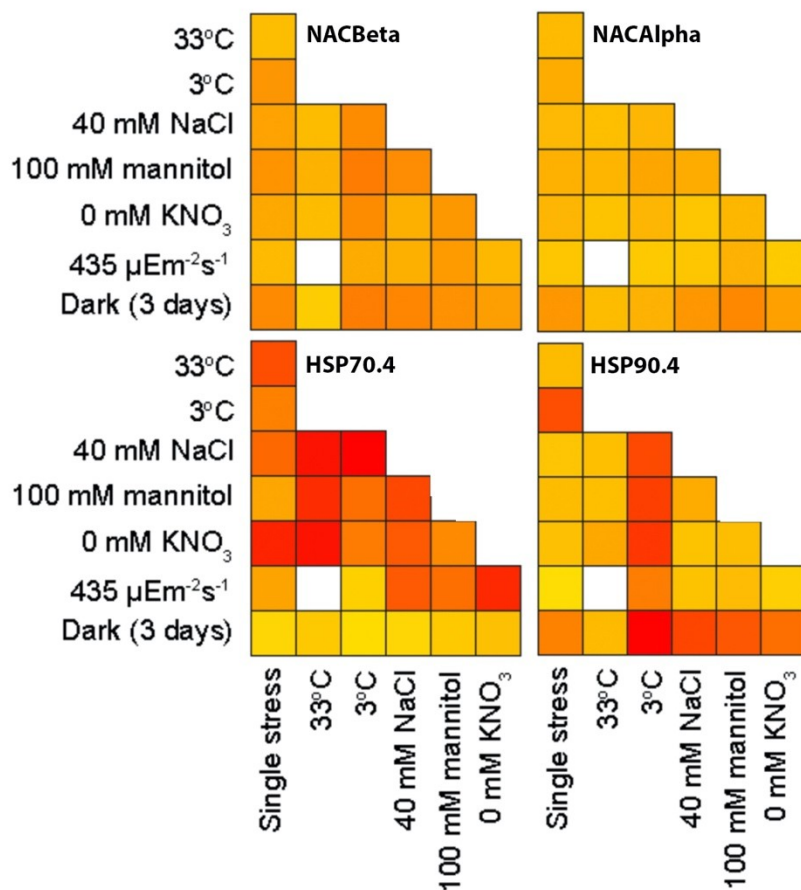
angiosperms could be higher level of ribosome heterogeneity. Ribosome heterogeneity was reported to play a role in developmental processes such as leaf formation or growth as well as in process involving auxin signalling (Horiguchi et al. 2012). Possibly, the different usage of ribosomal subunits could necessitate a further adaptation and specialisation of NAC subunits to accommodate to the usage of different ribosomal proteins. NAC functions span various processes, including stress response, development, chloroplast development, and plant-pathogen interactions. In human, NAC complex is associated with several diseases and is essential for survival, as its silencing in metazoans led to embryonic lethality (for instance in *Mus musculus*, *Drosophila melanogaster*, and *Caenorhabditis elegans*). We were not able to acquire an out of frame CRISPR mutant for neither NAC subunit of *M. polymorpha*. Furthermore, the previously reported *A. thaliana nacβ1/nacβ2* double homozygous mutant had retained expression of one NACβ subunit resulting only in knock-down. We thus suggest that complete silencing of NACβ subunits may also be lethal in plants.

The molecular functions of NAC complex have been most thoroughly described in metazoans and yeast. It was first identified as a ribosome-bound chaperone that shields nascent polypeptides from other cytoplasmic complexes, which may assist in their folding. We discovered that NAC is likely the part of the complex plant chaperone system, having confirmed its direct interaction with cytoplasmic chaperones—behavior also observed in Archaea, where 4 cytoplasmic chaperones were co-immunoprecipitated as potential NACα interactome (Singhal et al. 2020). Additionally, the reduced expression of NACβ subunits results in the upregulation of other chaperones at both the transcriptomic and proteomic levels. This suggests the integration of NAC into the large plant chaperone pathway. However, chaperone assays are needed to assess NAC's folding capacity. We have started such analyses following the previously described methodology (Haslbeck and Buchner 2015). However, we have experienced difficulties to purify enough NAC proteins from *A. thaliana* leaves to perform the assay (Unpublished data). As a next step, we will most likely try to isolate the recombinant proteins from *E.coli* or *S. cerevisiae*.

NAC association with ribosomes by the binding sites close to the ribosomal exit tunnel, busy with targets for various ribosome-associated complexes like RAC or SRP, have been computationally established. A recent mathematical model outlined a NAC regulatory role in protein import specificity to the endoplasmic reticulum by steric inhibition of the SRP binding site. Since many ribosomal proteins appeared on the list, our pull-down data provide indirect evidence that NAC association with ribosomes is maintained also in plants. We further propose that NAC complex is involved in protein sorting to ER, chloroplast and mitochondria, even though the pathway may be compensated by alternative mechanisms in case of NAC absence and thus a direct influence on the photosynthetic performance and protein sorting to chloroplasts in the *A. thaliana nacβ1/nacβ2* double homozygous mutant was not observed.

Future research directions might include investigating the NAC function in plant stress tolerance, where it was proposed to play an important role in various plant species under both biotic and abiotic stresses. For instance, NAC regulation of OST1 during cold stress via phosphorylation was reported (Yang et al. 2018). When studying the *nacβ1/nacβ2* knock-down mutant of *Arabidopsis thaliana*, the results were somewhat perplexing, as the mutant exhibited only mildly delayed germination under osmotic and salt stress. Furthermore, the

stressed plants did not show almost any upregulation of NAC subunits since in most cases, the expression did not change or was even lower than under the normal conditions (Unpublished data). This contrasts with typical stress response proteins, which can increase their abundance 10-100 fold upon induction. Similar observations were made in *M. polymorpha* NAC $\beta$  at the BAR expression database ([http://bar.utoronto.ca/efp\\_marchantia/cgi-bin/efpWeb.cgi](http://bar.utoronto.ca/efp_marchantia/cgi-bin/efpWeb.cgi)), which included experiments carried out under various stress conditions, which did not influence expression of any NAC subunit (Fig. 7). Given that chaperones are involved in protein refolding and regulation of protein aggregation under stress, NAC silencing may lead to proteostatic stress, necessitating the upregulation of other complexes to compensate for NAC roles in sorting and chaperoning. Interestingly, similar hypothesis is the one explaining why male gametophyte is the most stress- susceptible developmental stage. It claims that the same processes that stress normally triggers are also essential for pollen development, making it difficult for the gametophyte to cope with additional environmental stress (Sze et al. 2021).



**Figure 7** Expression profile of selected genes in *M. polymorpha* under various stress conditions obtained from the BAR expression database ([http://bar.utoronto.ca/efp\\_marchantia/cgi-bin/efpWeb.cgi](http://bar.utoronto.ca/efp_marchantia/cgi-bin/efpWeb.cgi)). The NAC subunits do not exhibit a significant change in expression under any of the stress conditions tested, whereas the *M. polymorpha* homologues of *A. thaliana* heat shock proteins, HSP70.4 and HSP90.4, show a response in expression to several stress stimuli.

Another area potentially interesting for exploration concerning NAC function in flowering plants is in meristem development. Evorepro transcriptome data indicate that NAC subunits

are notably expressed in meristematic tissues. In the *A. thaliana nacβ1/nacβ2* mutants, there is a distortion in the number of floral organs and an enlargement of the entire meristem, reminiscent of the *clavata3* phenotype. Our preliminary analyses, which included searches in available transcriptome and proteome datasets for links to meristem regulators, did not yield reports of NAC involvement. Furthermore, our quantitative PCR (qPCR) analysis of selected flower meristem regulators did not reveal any significant changes in expression in the mutant. However, meristem regulation appears to depend more significantly on proteomic rather than transcriptomic alterations thus further investigations would be required to delineate NAC role in meristem regulation.

Overall, our goal was to analyse NAC molecular function in plants. Initially, from the perspective of a laboratory specializing in pollen biology, we were interested in the potential role of NAC complex in male gametophyte development or fertilization. Yet, our data suggest that NAC does not likely possess a specialized function in these processes. Instead, it appears to be a pleiotropic, housekeeping protein involved in multiple stages of plant development across various tissues. Consequently, our research expanded to sporophytic tissues, and with the support of omics data and literature, we shifted our focus to NAC involvement in translation and protein sorting. This shift presented significant challenges; the pleiotropic functions of NAC complex together with its interplay with other chaperones and translation-associated complexes make it a difficult candidate to study. It is challenging to pinpoint the direct function or effect of NAC silencing due to the fact that our experiments showed usually indirect or subtly manifesting results.



## Conclusion

This thesis addresses the regulation of gene expression within the developing male gametophyte. The complex world of gene expression regulation begins with the initiation of transcription, driven by the binding of transcription factors to specific DNA motifs. To assist in this area of study, we created an online tool named GOLEM for the straightforward visualization and analysis of motif distribution across various plant species. Gene expression is intricately controlled in a temporal and spatial manner to precisely adjust growth, development, and responses to environmental changes. GOLEM uses RNA-seq data from various tissues to enable the selection of genes by their expression levels. This facilitates the examination of how motif distribution is specific to certain tissues and developmental stages. The tool includes data from various plant species also allowing comparison of motif distributions across evolution of various plant lineages.

Through multi-omics analysis including transcripts and proteins, we have enriched the understanding of the dynamics of gene expression in developing pollen with unprecedented resolution. We identified significant trends and functionally categorized families and groups of genes based on their expression patterns. The dataset offers a valuable repository of candidates for future functional studies. By integrating transcriptome and proteome data, we have suggested potential trends in mRNAs storage and dynamics, which can be further substantiated through methods such as translomics, analysing polysome- and monosome-associated mRNAs across different pollen stages. Current research in the Laboratory of Pollen Biology is complementing these findings, aiming to provide even more detailed and comprehensive regulatory map of mRNA fate in developing pollen. Extending our focus to protein regulation, we reviewed the dynamics of the phosphoproteome in the male gametophyte and analysed similarities in regulatory patterns from data published on various plant species. We identified potential regulatory commonalities essential for the polar growth of pollen tubes and root tips.

The nascent polypeptide-associated complex (NAC) protein family emerged as a promising candidate for studying their regulatory role in male gametophyte development due to their precise post-translational modifications during pollen activation. In the experimental benchwork phase of my thesis, we aimed to functionally analyse this family of ribosome-bound proteins, initially focusing on the male gametophyte. Given their housekeeping functions in both gametophyte and sporophyte tissues, we expanded the study to explore their broader molecular functions in plants. Our analysis revealed phenotypic defects in the *A. thaliana* plants with knocked-down NAC $\beta$ . Furthermore, the additional data suggest that NAC $\beta$  knock-out could be lethal, underscoring the necessity of at least one NAC $\beta$  subunit for plant viability. Given NAC multifaceted roles, unravelling their molecular function presents a significant challenge.

NAC involvement in pollen development, particularly in pollen tube growth, likely arises from its essential role in translation. During pollen tube germination and rapid growth, effective translation is critical, and the silencing of NAC $\beta$  subunits results in defective, slow-growing pollen tubes. Expression data from developing pollen reinforces the hypothesis of NAC role within the translation machinery, as its expression at both transcriptional and

translational levels parallels that of genes involved in translation and protein processing. Our findings suggest the association of NAC complex with the translation machinery, as previously described in animal systems. We hypothesize that NAC participates in protein transport to chloroplasts following its role as a cytoplasmic chaperone and that it may be part of a comprehensive cytoplasmic chaperone network, displaying direct interactions with several other chaperones and showing upregulation of other chaperones in the NAC $\beta$  knock-down plants. Nevertheless, additional experiments are required to ascertain the precise mechanisms of the functions of NAC subunits.

Given the high sequential and structural conservation of the NAC family in plants, we hypothesized that its function may be conserved in different plant species. Addressing this, we carried out functional complementation between angiosperm (represented by *Arabidopsis thaliana*) and liverwort (represented by *Marchantia polymorpha*) NAC $\beta$  subunits, demonstrating that a single NAC $\beta$  subunit under the *A. thaliana* promoter can rescue the phenotypic defects of the *A. thaliana* *nac $\beta$ 1nac $\beta$ 2* plants. Moreover, the NAC $\beta$  subunit from *M. polymorpha* was capable of dimerizing with NAC $\alpha$  paralogues from *A. thaliana*. Annotations of *M. polymorpha* NAC subunits mirrored the behaviour of *A. thaliana* NAC subunits in terms of cellular localization and expression dynamics. Regrettably, we were unable to obtain NAC subunit knock-outs in *Marchantia polymorpha*, potentially due to the lethality of the mutants.

In conclusion, this thesis sheds light on the complexities of gene expression regulation in developing pollen at transcriptional, post-transcriptional, translational, and post-translational levels. It provides a detailed annotation of the NAC protein family across two plant species, with the emphasis on their role in translation and male gametophyte development.

## Bibliography

- Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., & Tasaka, M. (1997). *Genes Involved in Organ Separation in Arabidopsis : . An Analysis of the cup-shaped cotyledon Mutant*. 9(June), 841–857.
- Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., Dreyfuss, G., Eddy, S. R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., & Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA (New York, N.Y.)*, 9(3), 277–279. <https://doi.org/10.1261/rna.2183803>
- Anderson, S. N., Johnson, C. S., Jones, D. S., Conrad, L. J., Gou, X., Russell, S. D., & Sundaresan, V. (2013). Transcriptomes of isolated *Oryza sativa* gametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization. *The Plant Journal : For Cell and Molecular Biology*, 76(5), 729–741. <https://doi.org/10.1111/tpj.12336>
- Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., & Crespi, M. (2014). Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Molecular Cell*, 55(3), 383–396. <https://doi.org/10.1016/j.molcel.2014.06.011>
- Arsova, B., Watt, M., & Usadel, B. (2018). Monitoring of plant protein post-translational modifications using targeted proteomics. *Frontiers in Plant Science*, 9(August), 1–9. <https://doi.org/10.3389/fpls.2018.01168>
- Aukerman, M. J., & Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *The Plant Cell*, 15(11), 2730–2741. <https://doi.org/10.1105/tpc.016238>
- Baerenfaller, K., Grossmann, J., Grobei, M. A., Hull, R., Hirsch-Hoffmann, M., Yalovsky, S., Zimmermann, P., Grossniklaus, U., Gruissem, W., & Baginsky, S. (2008). Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science (New York, N.Y.)*, 320(5878), 938–941. <https://doi.org/10.1126/science.1157956>
- Bai, B., Van Der Horst, S., Cordewener, J. H. G., America, T. A. H. P., Hanson, J., & Bentsink, L. (2020). Seed-stored mRNAs that are specifically associated to monosomes are translationally regulated during germination. *Plant Physiology*, 182(1), 378–392. <https://doi.org/10.1104/pp.19.00644>
- Balasubramanian, S., Sureshkumar, S., Lempe, J., & Weigel, D. (2006). Potent Induction of *Arabidopsis thaliana* Flowering by Elevated Growth Temperature. *PLOS Genetics*, 2(7), 1–10. <https://doi.org/10.1371/journal.pgen.0020106>
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1), 4–9. <https://doi.org/10.1186/s13100-015-0041-9>
- Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzergue, S., Brown, J. W. S., & Crespi, M. (2014). Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. *Developmental Cell*, 30(2), 166–176. <https://doi.org/10.1016/j.devcel.2014.06.017>
- Barnabas, B., & Fridvalszky, L. (1984). Adhesion and germination of differently treated maize pollen grains on the stigma. *Acta Botanica Hungarica*, 30(3–4), 329–332. <https://eurekamag.com/research/001/293/001293517.php>
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*,

- 143(3), 470–484. <https://doi.org/10.1016/j.cell.2010.10.012>
- Bate, N., & Twell, D. (1998). Functional architecture of a late pollen promoter: Pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Molecular Biology*, 37(5), 859–869. <https://doi.org/10.1023/A:1006095023050>
- Bazin, J., Baerenfaller, K., Gosai, S. J., Gregory, B. D., Crespi, M., & Bailey-Serres, J. (2017). Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 114(46), E10018–E10027. <https://doi.org/10.1073/pnas.1708433114>
- Beatrix, B., Sakai, H., & Wiedmann, M. (2000). The  $\alpha$  and  $\beta$  subunit of the nascent polypeptide-associated complex have distinct functions. *Journal of Biological Chemistry*, 275(48), 37838–37845. <https://doi.org/10.1074/jbc.M006368200>
- Berger, F., & Twell, D. (2011). Germline specification and function in plants. *Annual Review of Plant Biology*, 62, 461–484. <https://doi.org/10.1146/annurev-arplant-042110-103824>
- Beuder, S., Dorchak, A., Bhide, A., Moeller, S. R., Petersen, B. L., & MacAlister, C. A. (2020). Exocyst mutants suppress pollen tube growth and cell wall structural defects of hydroxyproline O-arabinosyltransferase mutants. *Plant Journal*, 103(4), 1399–1419. <https://doi.org/10.1111/tpj.14808>
- Bhat, S. S., Bielewicz, D., Gulanicz, T., Bodi, Z., Yu, X., Anderson, S. J., Szewc, L., Bajczyk, M., Dolata, J., Grzelak, N., Smolinski, D. J., Gregory, B. D., Fray, R. G., Jarmolowski, A., & Szweykowska-Kulinska, Z. (2020). mRNA adenosine methylase (MTA) deposits m6A on pri-miRNAs to modulate miRNA biogenesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 117(35), 21785–21795. <https://doi.org/10.1073/pnas.2003733117>
- Bhogireddy, S., Mangrauthia, S. K., Kumar, R., Pandey, A. K., Singh, S., Jain, A., Budak, H., Varshney, R. K., & Kudapa, H. (2021). Regulatory non-coding RNAs: a new frontier in regulation of plant biology. *Functional and Integrative Genomics*, 21(3–4), 313–330. <https://doi.org/10.1007/s10142-021-00787-8>
- Bloss, T. A., Witze, E. S., & Rothman, J. H. (2003). Suppression of CED-3-independent apoptosis by mitochondrial  $\beta$ NAC in *Caenorhabditis elegans*. *Nature*, 424(6952), 1066–1071. <https://doi.org/10.1038/nature01920>
- Bokszczanin, K. L., & Fragkostefanakis, S. (2013). Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Frontiers in Plant Science*, 4, 315. <https://doi.org/10.3389/fpls.2013.00315>
- Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., & Twella, D. (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *Plant Cell*, 23(2), 534–549. <https://doi.org/10.1105/tpc.110.081059>
- Borg, M., Papareddy, R. K., Dombey, R., Axelsson, E., Nodine, M. D., Twell, D., & Berger, F. (2021). Epigenetic reprogramming rewires transcription during the alternation of generations in *Arabidopsis*. *ELife*, 10, 1–66. <https://doi.org/10.7554/eLife.61894>
- Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J. A., & Becker, J. D. (2008). Comparative transcriptomics of *Arabidopsis* sperm cells. *Plant Physiology*, 148(2), 1168–1181. <https://doi.org/10.1104/pp.108.125229>
- Borges, F., Pereira, P. A., Slotkin, R. K., Martienssen, R. A., & Becker, J. D. (2011). MicroRNA activity in the *Arabidopsis* male germline. *Journal of Experimental Botany*, 62(5), 1611–

1620. <https://doi.org/10.1093/jxb/erq452>
- Budak, H., Kaya, S. B., & Cagirici, H. B. (2020). Long Non-coding RNA in Plants in the Era of Reference Sequences. *Frontiers in Plant Science*, 11(March), 1–10. <https://doi.org/10.3389/fpls.2020.00276>
- Calarco, J. P., Borges, F., Donoghue, M. T. A., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijó, J. A., Becker, J. D., & Martienssen, R. A. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*, 151(1), 194–205. <https://doi.org/10.1016/j.cell.2012.09.001>
- Čapková-Balatková, V., Hrabětová, E., & Tupý, J. (1980). Effects of cycloheximide on, pollen of *Nicotiana tabacum* in Culture. *Biochemie Und Physiologie Der Pflanzen*, 175(5), 412–420. [https://doi.org/https://doi.org/10.1016/S0015-3796\(80\)80027-8](https://doi.org/https://doi.org/10.1016/S0015-3796(80)80027-8)
- Chen, L. L., DeCerbo, J. N., & Carmichael, G. G. (2008). Alu element-mediated gene silencing. *EMBO Journal*, 27(12), 1694–1705. <https://doi.org/10.1038/emboj.2008.94>
- Chen, Y., Liu, M., & Dong, Z. (2021). Preferential ribosome loading on the stress-upregulated mrna pool shapes the selective translation under stress conditions. *Plants*, 10(2), 1–15. <https://doi.org/10.3390/plants10020304>
- Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant Journal*, 89(4), 789–804. <https://doi.org/10.1111/tpj.13415>
- Chinnusamy, V., & Zhu, J.-K. (2009). Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology*, 12(2), 133–139. <https://doi.org/10.1016/j.pbi.2008.12.006>
- Conn, V. M., Hugouvieux, V., Nayak, A., Conos, S. A., Capovilla, G., Cildir, G., Jourdain, A., Tergaonkar, V., Schmid, M., Zubieta, C., & Conn, S. J. (2017). A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nature Plants*, 3, 17053. <https://doi.org/10.1038/nplants.2017.53>
- Csorba, T., Questa, J. I., Sun, Q., & Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45), 16160–16165. <https://doi.org/10.1073/pnas.1419030111>
- Cuellar, T. L., & McManus, M. T. (2005). MicroRNAs and endocrine biology. *Journal of Endocrinology*, 187(3), 327–332. <https://doi.org/10.1677/joe.1.06426>
- Daniela Cordeiro, S., Canhoto, J., & Correia, S. (2022). Regulatory non-coding RNAs: Emerging roles during plant cell reprogramming and in vitro regeneration. *Frontiers in Plant Science*, 13(November), 1–10. <https://doi.org/10.3389/fpls.2022.1049631>
- Das, S., & Bansal, M. (2019). Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. In *PLoS ONE* (Vol. 14, Issue 3). <https://doi.org/10.1371/journal.pone.0212678>
- Datta, R., & Paul, S. (2019). Long non-coding RNAs: Fine-tuning the developmental responses in plants. *Journal of Biosciences*, 44(4), 1–11. <https://doi.org/10.1007/s12038-019-9910-6>
- David, R., Burgess, A., Parker, B., Li, J., Pulsford, K., Sibbritt, T., Preiss, T., & Searle, I. R. (2017). Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. *The Plant Cell*, 29(3), 445–460. <https://doi.org/10.1105/tpc.16.00751>
- De Storme, N., & Geelen, D. (2013). Cytokinesis in plant male meiosis. *Plant Signaling and Behavior*, 8(3). <https://doi.org/10.4161/psb.23394>

- Del Duca, S., Serafini-Fracassini, D., Bonner, P. L. R., Cresti, M., & Cai, G. (2009). Effects of post-translational modifications catalysed by pollen transglutaminase on the functional properties of microtubules and actin filaments. *Biochemical Journal*, 418(3), 651–664. <https://doi.org/10.1042/BJ20081781>
- Deng, J. M., & Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Research*, 4(4), 264–269. <https://doi.org/10.1007/BF01969120>
- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., & Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400(6745), 693–696. <https://doi.org/10.1038/23301>
- Dieci, G., Conti, A., Pagano, A., & Carnevali, D. (2013). Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochimica et Biophysica Acta*, 1829(3–4), 296–305. <https://doi.org/10.1016/j.bbagr.2012.09.010>
- Dinkova, T. D., Zepeda, H., Martínez-Salas, E., Martínez, L. M., Nieto-Sotelo, J., & Sánchez De Jiménez, E. (2005). Cap-independent translation of maize Hsp101. *Plant Journal*, 41(5), 722–731. <https://doi.org/10.1111/j.1365-313X.2005.02333.x>
- Doench, J. G., & Sharp, P. A. (2004). Specificity of microRNA target selection in translational repression. *Genes & Development*, 18(5), 504–511. <https://doi.org/10.1101/gad.1184404>
- Dong, Q., Hu, B., & Zhang, C. (2022). microRNAs and Their Roles in Plant Development. *Frontiers in Plant Science*, 13(February). <https://doi.org/10.3389/fpls.2022.824240>
- Eaton, J. D., & West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! *Trends in Genetics*, 36(9), 664–675. <https://doi.org/10.1016/j.tig.2020.05.008>
- Edlund, A. F., Swanson, R., & Preuss, D. (2004). Pollen and stigma structure and function: The role of diversity in pollination. *Plant Cell*, 16(SUPPL.), 84–97. <https://doi.org/10.1105/tpc.015800>
- Estrada, A. D., Freese, N. H., Blakley, I. C., & Loraine, A. E. (2015). Analysis of pollen-specific alternative splicing in *Arabidopsis thaliana* via semi-quantitative PCR. *PeerJ*, 2015(3). <https://doi.org/10.7717/peerj.919>
- Fíla, J., Matros, A., Radau, S., Zahedi, R. P., Čapková, V., Mock, H. P., & Honys, D. (2012). Revealing phosphoproteins playing role in tobacco pollen activated in vitro. *Proteomics*, 12(21), 3229–3250. <https://doi.org/10.1002/pmic.201100318>
- Fíla, J., Radau, S., Matros, A., Hartmann, A., Scholz, U., Feciková, J., Mock, H.-P., Čapková, V., Zahedi, R. P., & Honys, D. (2016). Phosphoproteomics Profiling of Tobacco Mature Pollen and Pollen Activated in vitro. *Molecular & Cellular Proteomics*, 15(4), 1338–1350. <https://doi.org/10.1074/mcp.m115.051672>
- Fíla, J., Závěská Drábková, L., Gibalová, A., & Honys, D. (2017). When Simple Meets Complex: Pollen and the -Omics. In G. Obermeyer & J. Feijó (Eds.), *Pollen Tip Growth: From Biophysical Aspects to Systems Biology* (pp. 247–292). Springer International Publishing. [https://doi.org/10.1007/978-3-319-56645-0\\_10](https://doi.org/10.1007/978-3-319-56645-0_10)
- Flores-Tornero, M., & Becker, J. D. (2023). Fifty years of sperm cell isolations: from structural to omic studies. *Journal of Experimental Botany*, 74(12), 3449–3461. <https://doi.org/10.1093/jxb/erad117>
- Footitt, S., & Cohn, M. A. (2001). Developmental arrest: from sea urchins to seeds. *Seed Science Research*, 11(1), 3–16. <https://doi.org/10.1079/SSR200055>
- Fort, V., Khelifi, G., & Hussein, S. M. I. (2021). Long non-coding RNAs and transposable elements: A functional relationship. *Biochimica et Biophysica Acta - Molecular Cell*

- Research*, 1868(1), 118837. <https://doi.org/10.1016/j.bbamcr.2020.118837>
- Fragkostefanakis, S., Mesihovic, A., Hu, Y., & Schleiff, E. (2016). Unfolded protein response in pollen development and heat stress tolerance. *Plant Reproduction*, 29(1-2), 81-91. <https://doi.org/10.1007/s00497-016-0276-8>
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J. A., & Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics*, 39(8), 1033-1037. <https://doi.org/10.1038/ng2079>
- Friso, G., & Van Wijk, K. J. (2015). Posttranslational protein modifications in plant metabolism. *Plant Physiology*, 169(3), 1469-1487. <https://doi.org/10.1104/pp.15.01378>
- Ganie, S. A. (2020). RNA chaperones: Potential candidates for engineering salt tolerance in rice. *Crop Science*, 60(2), 530-540. <https://doi.org/10.1002/csc2.20134>
- Giorno, F., Wolters-Arts, M., Mariani, C., & Rieu, I. (2013). Ensuring reproduction at high temperatures: The heat stress response during anther and pollen development. *Plants*, 2(3), 489-506. <https://doi.org/10.3390/plants2030489>
- Gisselbrecht, S. S., Palagi, A., Kurland, J. V., Rogers, J. M., Ozadam, H., Zhan, Y., Dekker, J., & Bulyk, M. L. (2020). Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Molecular Cell*, 77(2), 324-337.e8. <https://doi.org/10.1016/j.molcel.2019.10.004>
- Golicz, A. A., Allu, A. D., Li, W., Lohani, N., Singh, M. B., & Bhalla, P. L. (2021). A dynamic intron retention program regulates the expression of several hundred genes during pollen meiosis. *Plant Reproduction*, 0123456789. <https://doi.org/10.1007/s00497-021-00411-6>
- Gou, X., Yuan, T., Wei, X., & Russell, S. D. (2009). Gene expression in the dimorphic sperm cells of *Plumbago zeylanica*: Transcript profiling, diversity, and relationship to cell type. *Plant Journal*, 60(1), 33-47. <https://doi.org/10.1111/j.1365-313X.2009.03934.x>
- Guan, Y., Meng, X., Khanna, R., LaMontagne, E., Liu, Y., & Zhang, S. (2014). Phosphorylation of a WRKY Transcription Factor by MAPKs Is Required for Pollen Development and Function in Arabidopsis. *PLoS Genetics*, 10(5). <https://doi.org/10.1371/journal.pgen.1004384>
- Gupta, P., Rangan, L., Ramesh, T. V., & Gupta, M. (2016). Comparative analysis of contextual bias around the translation initiation sites in plant genomes. *Journal of Theoretical Biology*, 404, 303-311. <https://doi.org/10.1016/j.jtbi.2016.06.015>
- Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, 19(10), 621-637. <https://doi.org/10.1038/s41580-018-0028-8>
- Haerizadeh, F., Singh, M. B., & Bhalla, P. L. (2006). Transcriptional repression distinguishes somatic from germ cell lineages in a plant. *Science*, 313(5786), 496-499. <https://doi.org/10.1126/science.1125526>
- Hafidh, S., & Honys, D. (2021). Reproduction Multitasking: The Male Gametophyte. *Annual Review of Plant Biology*, 72(1), 581-614. <https://doi.org/10.1146/annurev-arplant-080620-021907>
- Hafidh, S., Potěšil, D., Müller, K., Fíla, J., Michailidis, C., Herrmannová, A., Feciková, J., Ischebeck, T., Valášek, L. S., Zdráhal, Z., & Honys, D. (2018). Dynamics of the pollen sequestrome defined by subcellular coupled omics. *Plant Physiology*, 178(1), 258-282. <https://doi.org/10.1104/pp.18.00648>

- Hamamura, Y., Nagahara, S., & Higashiyama, T. (2012). Double fertilization on the move. *Current Opinion in Plant Biology*, 15(1), 70–77. <https://doi.org/10.1016/j.pbi.2011.11.001>
- Hamilton, A., Voinnet, O., Chappell, L., & Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. *The EMBO Journal*, 21(17), 4671–4679. <https://doi.org/10.1093/emboj/cdf464>
- Hammoudi, V., Beerens, B., Jonker, M. J., Helderma, T. A., Vlachakis, G., Giesbers, M., Kwaaitaal, M., & van den Burg, H. A. (2021). The protein modifier SUMO is critical for integrity of the Arabidopsis shoot apex at warm ambient temperatures. *Journal of Experimental Botany*. <https://doi.org/10.1093/jxb/erab262>
- Hao, W., Liu, G., Wang, W., Shen, W., Zhao, Y., Sun, J., Yang, Q., Zhang, Y., Fan, W., Pei, S., Chen, Z., Xu, D., & Qin, T. (2021). RNA Editing and Its Roles in Plant Organelles. *Frontiers in Genetics*, 12(September). <https://doi.org/10.3389/fgene.2021.757109>
- Haslbeck, M., & Buchner, J. (2015). Assays to characterize molecular chaperone function in vitro. *Methods in Molecular Biology (Clifton, N.J.)*, 1292, 39–51. [https://doi.org/10.1007/978-1-4939-2522-3\\_3](https://doi.org/10.1007/978-1-4939-2522-3_3)
- Herschlag, D. (1995). RNA chaperones and the RNA folding problem. *Journal of Biological Chemistry*, 270(36), 20871–20874. <https://doi.org/10.1074/jbc.270.36.20871>
- Heslop-Harrison, J. (1987). Pollen Germination and Pollen-Tube Growth. In G. H. Bourne, K. W. Jeon, & M. Friedlander (Eds.), *Pollen: Cytology and Development* (Vol. 107, pp. 1–78). Academic Press. [https://doi.org/10.1016/S0074-7696\(08\)61072-4](https://doi.org/10.1016/S0074-7696(08)61072-4)
- Hoffmann, R. D., Olsen, L. I., Husum, J. O., Nicolet, J. S., Thøfner, J. F. B., Wätjen, A. P., Ezike, C. V., & Palmgren, M. (2017). A cis-Regulatory Sequence Acts as a Repressor in the Arabidopsis thaliana Sporophyte but as an Activator in Pollen. *Molecular Plant*, 10(5), 775–778. <https://doi.org/10.1016/j.molp.2016.12.010>
- Honys, D., Renak, D., Fecikova, J., Jedelsky, P. L., Nebesarova, J., Dobrev, P., & Capkova, V. (2009). Cytoskeleton-associated large RNP complexes in tobacco male gametophyte (EPPs) are associated with ribosomes and are involved in protein synthesis, processing, and localization. *Journal of Proteome Research*, 8(4), 2015–2031. <https://doi.org/10.1021/pr8009897>
- Honys, D., & Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. *Genome Biology*, 5(11). <https://doi.org/10.1186/gb-2004-5-11-r85>
- Horiguchi, G., Van Lijsebettens, M., Candela, H., Micol, J. L., & Tsukaya, H. (2012). Ribosomes and translation in plant developmental control. *Plant Science*, 191–192, 24–34. <https://doi.org/10.1016/j.plantsci.2012.04.008>
- Hsieh, H. H., Lee, J. H., Chandrasekar, S., & Shan, S. (2020). A ribosome-associated chaperone enables substrate triage in a cotranslational protein targeting complex. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-19548-5>
- Hu, L., Liang, W., Yin, C., Cui, X., Zong, J., Wang, X., Hu, J., & Zhang, D. (2011). Rice MADS3 regulates ROS homeostasis during late anther development. *Plant Cell*, 23(2), 515–533. <https://doi.org/10.1105/tpc.110.074369>
- Ichino, L., Picard, C. L., Yun, J., Chotai, M., Wang, S., Lin, E. K., Papareddy, R. K., Xue, Y., & Jacobsen, S. E. (2022). Single-nucleus RNA-seq reveals that MBD5, MBD6, and SILENZIO maintain silencing in the vegetative cell of developing pollen. *Cell Reports*, 41(8), 111699. <https://doi.org/10.1016/j.celrep.2022.111699>
- Ivanyi-Nagy, R., Lavergne, J. P., Gabus, C., Ficheux, D., & Darlix, J. L. (2008). RNA chaperoning



- and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic Acids Research*, 36(3), 712–725. <https://doi.org/10.1093/nar/gkm1051>
- Jabnourne, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., & Poirier, Y. (2013). A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *The Plant Cell*, 25(10), 4166–4182. <https://doi.org/10.1105/tpc.113.116251>
- Jiao, Y., & Meyerowitz, E. M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Molecular Systems Biology*, 6(419). <https://doi.org/10.1038/msb.2010.76>
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 45(D1), D1040–D1045. <https://doi.org/10.1093/nar/gkw982>
- Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics (Oxford, England)*, 31(22), 3593–3599. <https://doi.org/10.1093/bioinformatics/btv422>
- Julca, I., Ferrari, C., Flores-Tornero, M., Proost, S., Lindner, A.-C., Hackenberg, D., Steinbachová, L., Michaelidis, C., Gomes Pereira, S., Misra, C. S., Kawashima, T., Borg, M., Berger, F., Goldberg, J., Johnson, M., Honys, D., Twell, D., Sprunck, S., Dresselhaus, T., ... Mutwil, M. (2021). Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. *Nature Plants*, 7(8), 1143–1159. <https://doi.org/10.1038/s41477-021-00958-2>
- Julca, I., Flores, M., Proost, S., Lindner, A.-C., Hackenberg, D., Steinbachová, L., Michaelidis, C., Gomes Pereira, S., Misra, C. S., Kawashima, T., Borg, M., Berger, F., Goldberg, J., Johnson, M., Honys, D., Twell, D., Sprunck, S., Dresselhaus, T., Becker, J., & Mutwil, M. (2020). Comparative transcriptomic analysis reveals conserved transcriptional programs underpinning organogenesis and reproduction in land plants. *BioRxiv*, 2020.10.29.361501. <http://biorxiv.org/content/early/2020/10/30/2020.10.29.361501.abstract>
- Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L. A., Bourque, G., Yandell, M., & Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genetics*, 9(4). <https://doi.org/10.1371/journal.pgen.1003470>
- Kato, M., Watanabe, H., & Hoshino, Y. (2022). Evaluation of pollen tube growth ability in *Petunia* species having different style lengths. 92, 85–92. <https://doi.org/10.5511/plantbiotechnology.21.1113a>
- Keller, M., Simm, S., Bokszczanin, K. L., Bostan, H., Bovy, A., Chaturvedi, P., Chen, Y., Chiusano, M. L., Firon, N., Fragkostefanakis, S., Iannacone, R., Jegadeesan, S., Li, H., Mariani, C., Marko, D., Mesihovic, A., Müller, F., Paul, P., Paupiere, M., ... Vriezen, W. (2018). The coupling of transcriptome and proteome adaptation during development and heat stress response of tomato pollen. *BMC Genomics*, 19(1), 1–20. <https://doi.org/10.1186/s12864-018-4824-5>
- Kelley, D., & Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology*, 13(11), R107. <https://doi.org/10.1186/gb-2012-13-11-r107>
- Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database.

- Scientific Reports*, 1, 1–5. <https://doi.org/10.1038/srep00090>
- Kim, M. H., Sato, S., Sasaki, K., Saburi, W., Matsui, H., & Imai, R. (2013). COLD SHOCK DOMAIN PROTEIN 3 is involved in salt and drought stress tolerance in Arabidopsis. *FEBS Open Bio*, 3, 438–442. <https://doi.org/10.1016/j.fob.2013.10.003>
- Klodová, B., Potěšil, D., Steinbachová, L., Michailidis, C., Lindner, A.-C., Hackenberg, D., Becker, J. D., Zdráhal, Z., Twell, D., & Honys, D. (2023). Regulatory dynamics of gene expression in the developing male gametophyte of Arabidopsis. *Plant Reproduction*, 36(3), 213–241. <https://doi.org/10.1007/s00497-022-00452-5>
- Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, 15(20), 8125–8148. <https://doi.org/10.1093/nar/15.20.8125>
- Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, 299(1–2), 1–34. [https://doi.org/10.1016/s0378-1119\(02\)01056-9](https://doi.org/10.1016/s0378-1119(02)01056-9)
- Kurihara, Y., Makita, Y., Kawashima, M., Fujita, T., Iwasaki, S., & Matsui, M. (2018). Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 115(30), 7831–7836. <https://doi.org/10.1073/pnas.1804971115>
- Kwak, K. J., Park, S. J., Han, J. H., Kim, M. K., Oh, S. H., Han, Y. S., & Kang, H. (2011). Structural determinants crucial to the RNA chaperone activity of glycine-rich RNA-binding proteins 4 and 7 in Arabidopsis thaliana during the cold adaptation process. *Journal of Experimental Botany*, 62(11), 4003–4011. <https://doi.org/10.1093/jxb/err101>
- Lageix, S., Lanet, E., Pouch-Pélissier, M.-N., Espagnol, M.-C., Robaglia, C., Deragon, J.-M., & Pélissier, T. (2008). Arabidopsis eIF2alpha kinase GCN2 is essential for growth in stress conditions and is activated by wounding. *BMC Plant Biology*, 8, 134. <https://doi.org/10.1186/1471-2229-8-134>
- Le Lievre, L., Chakkatu, S. P., Varghese, S., Day, R. C., Pilkington, S. M., & Brownfield, L. (2023). RNA-seq analysis of synchronized developing pollen isolated from a single anther. *Frontiers in Plant Science*, 14(April), 1–14. <https://doi.org/10.3389/fpls.2023.1121570>
- Lee, D. H., Lal, N. K., Lin, Z. J. D., Ma, S., Liu, J., Castro, B., Toruño, T., Dinesh-Kumar, S. P., & Coaker, G. (2020). Regulation of reactive oxygen species during plant immunity through phosphorylation and ubiquitination of RBOHD. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-15601-5>
- Lim, P. K., Zheng, X., Goh, J. C., & Mutwil, M. (2022). Exploiting plant transcriptomic databases: Resources, tools, and approaches. *Plant Communications*, 3(4), 100323. <https://doi.org/10.1016/j.xplc.2022.100323>
- Lin, S. Y., Chen, P. W., Chuang, M. H., Juntawong, P., Bailey-Serres, J., & Jauh, G. Y. (2014). Profiling of transcriptomes of in vivo-grown pollen tubes reveals genes with roles in micropylar guidance during pollination in Arabidopsis. *Plant Cell*, 26(2), 602–618. <https://doi.org/10.1105/tpc.113.121335>
- Lindner, H., Kessler, S. A., Müller, L. M., Shimamoto-Asano, H., Boisson-Dernier, A., & Grossniklaus, U. (2015). TURAN and EVAN Mediate Pollen Tube Reception in Arabidopsis Synergids through Protein Glycosylation. *PLoS Biology*, 13(4), 1–22. <https://doi.org/10.1371/journal.pbio.1002139>
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., & Chua, N.-H.

- (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *The Plant Cell*, 24(11), 4333–4345.  
<https://doi.org/10.1105/tpc.112.102855>
- Liu, L., Lu, Y., Wei, L., Yu, H., Cao, Y., Li, Y., Yang, N., Song, Y., Liang, C., & Wang, T. (2018). Transcriptomics analyses reveal the molecular roadmap and long non-coding RNA landscape of sperm cell lineage development. *Plant Journal*, 96(2), 421–437.  
<https://doi.org/10.1111/tpj.14041>
- Lohani, N., Golicz, A. A., Allu, A. D., Bhalla, P. L., & Singh, M. B. (2023). Genome-wide analysis reveals the crucial role of lncRNAs in regulating the expression of genes controlling pollen development. *Plant Cell Reports*, 42(2), 337–354.  
<https://doi.org/10.1007/s00299-022-02960-0>
- Loraine, A. E., McCormick, S., Estrada, A., Patel, K., & Qin, P. (2013). RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiology*, 162(2), 1092–1109. <https://doi.org/10.1104/pp.112.211441>
- Lubelsky, Y., & Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*, 555(7694), 107–111.  
<https://doi.org/10.1038/nature25757>
- MacAlister, C. A., Ortiz-Ramírez, C., Becker, J. D., Feijó, J. A., & Lippman, Z. B. (2016). Hydroxyproline O-arabinosyltransferase mutants oppositely alter tip growth in *Arabidopsis thaliana* and *Physcomitrella patens*. *Plant Journal*, 85(2), 193–208.  
<https://doi.org/10.1111/tpj.13079>
- Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters*, 583(24), 3966–3973.  
<https://doi.org/10.1016/j.febslet.2009.10.036>
- Maldonado-Bonilla, L. D. (2014). Composition and function of P bodies in *Arabidopsis thaliana*. *Frontiers in Plant Science*, 5, 201. <https://doi.org/10.3389/fpls.2014.00201>
- Markesich, D. C., Gajewski, K. M., Nazimiec, M. E., & Beckingham, K. (2000). bicaudal encodes the *Drosophila* beta NAC homolog, a component of the ribosomal translational machinery\*. *Development (Cambridge, England)*, 127(3), 559–572.  
<http://www.ncbi.nlm.nih.gov/pubmed/10631177>
- Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q., & Dean, C. (2014). Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. *Molecular Cell*, 54(1), 156–165. <https://doi.org/10.1016/j.molcel.2014.03.026>
- Mascarenhas, J. P. (1993). Molecular Mechanisms of Pollen Tube Growth and Differentiation. *The Plant Cell*, 5(10), 1303. <https://doi.org/10.2307/3869783>
- Mayer, O., Rajkowitsch, L., Lorenz, C., Konrat, R., & Schroeder, R. (2007). RNA chaperone activity and RNA-binding properties of the *E. coli* protein StpA. *Nucleic Acids Research*, 35(4), 1257–1269. <https://doi.org/10.1093/nar/gkl1143>
- McCormick, S. (2004). Control of male gametophyte development. *Plant Cell*, 16(SUPPL.), 142–153. <https://doi.org/10.1105/tpc.016659>
- Mejía-Guerra, M. K., Li, W., Galeano, N. F., Vidal, M., Gray, J., Doseff, A. I., & Grotewold, E. (2015). Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. *Plant Cell*, 27(12), 3309–3320.  
<https://doi.org/10.1105/tpc.15.00630>
- Melencion, S. M. B., Chi, Y. H., Pham, T. T., Paeng, S. K., Wi, S. D., Lee, C., Ryu, S. W., Koo, S. S., & Lee, S. Y. (2017). RNA chaperone function of a universal stress protein in *Arabidopsis* confers enhanced cold stress tolerance in plants. *International Journal of*

- Molecular Sciences*, 18(12), 1–16. <https://doi.org/10.3390/ijms18122546>
- Meyer, K., Koester, T., & Staiger, D. (2015). Pre-mRNA Splicing in Plants: In Vivo Functions of RNA-Binding Proteins Implicated in the Splicing Process. *Biomolecules*, 5(3), 1717–1740. <https://doi.org/10.3390/biom5031717>
- Meyers, B. C., Axtell, M. J., Bartel, B., Bartel, D. P., Baulcombe, D., Bowman, J. L., Cao, X., Carrington, J. C., Chen, X., Green, P. J., Griffiths-Jones, S., Jacobsen, S. E., Mallory, A. C., Martienssen, R. A., Poethig, R. S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., ... Zhu, J.-K. (2008). Criteria for annotation of plant MicroRNAs. *The Plant Cell*, 20(12), 3186–3190. <https://doi.org/10.1105/tpc.108.064311>
- Miller, W. A., Wang, Z., & Treder, K. (2007). The amazing diversity of cap-independent translation elements in the 3'-untranslated regions of plant viral RNAs. *Biochemical Society Transactions*, 35(6), 1629–1633. <https://doi.org/10.1042/BST0351629>
- Misra, C. S., Santos, M. R., Rafael-Fernandes, M., Martins, N. P., Monteiro, M., & Becker, J. D. (2019). Transcriptomics of Arabidopsis sperm cells at single-cell resolution. *Plant Reproduction*, 32(1), 29–38. <https://doi.org/10.1007/s00497-018-00355-4>
- Misra, C. S., Sousa, A. G. G., Barros, P. M., Kermanov, A., & Becker, J. D. (2023). Cell-type-specific alternative splicing in the Arabidopsis germline. *Plant Physiology*, 192(1), 85–101. <https://doi.org/10.1093/plphys/kiac574>
- Missra, A., Ernest, B., Lohoff, T., Jia, Q., Satterlee, J., Ke, K., & von Arnim, A. G. (2015). The Circadian Clock Modulates Global Daily Cycles of mRNA Ribosome Loading. *The Plant Cell*, 27(9), 2582–2599. <https://doi.org/10.1105/tpc.15.00546>
- Moore, M. J., & Proudfoot, N. J. (2009). Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell*, 136(4), 688–700. <https://doi.org/10.1016/j.cell.2009.02.001>
- Mustroph, A., & Bailey-Serres, J. (2010). The Arabidopsis transcriptome cell-specific mRNA atlas: Mining suberin and cutin lipid monomer biosynthesis genes as an example for data application. *Plant Signaling and Behavior*, 5(3), 320–324. <https://doi.org/10.4161/psb.5.3.11187>
- Ngan, C. Y., Wong, C. H., Tjong, H., Wang, W., Goldfeder, R. L., Choi, C., He, H., Gong, L., Lin, J., Urban, B., Chow, J., Li, M., Lim, J., Philip, V., Murray, S. A., Wang, H., & Wei, C.-L. (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nature Genetics*, 52(3), 264–272. <https://doi.org/10.1038/s41588-020-0581-x>
- Nguyen, C. C., Nakaminami, K., Matsui, A., Kobayashi, S., Kurihara, Y., Toyooka, K., Tanaka, M., & Seki, M. (2016). Oligouridylylate Binding Protein 1b Plays an Integral Role in Plant Heat Stress Tolerance. *Frontiers in Plant Science*, 7, 853. <https://doi.org/10.3389/fpls.2016.00853>
- Nguyen, T. M., Kabotyanski, E. B., Reineke, L. C., Shao, J., Xiong, F., Lee, J. H., Dubrulle, J., Johnson, H., Stossi, F., Tsoi, P. S., Choi, K. J., Ellis, A. G., Zhao, N., Cao, J., Adewunmi, O., Ferreon, J. C., Ferreon, A. C. M., Neilson, J. R., Mancini, M. A., ... Rosen, J. M. (2020). The SINEB1 element in the long non-coding RNA Malat1 is necessary for TDP-43 proteostasis. *Nucleic Acids Research*, 48(5), 2621–2642. <https://doi.org/10.1093/nar/gkz1176>
- Nie, H., Cheng, C., Kong, J., Li, H., & Hua, J. (2023). Plant non-coding RNAs function in pollen development and male sterility. *Frontiers in Plant Science*, 14(February), 1–12. <https://doi.org/10.3389/fpls.2023.1109941>
- Ozadam, H., Zhan, Y., Dekker, J., & Bulyk, M. L. (2020). *Transcriptional Enhancers in*

- Alternate Cellular Contexts*. 77(2), 324–337.  
<https://doi.org/10.1016/j.molcel.2019.10.004>. Transcriptional
- Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J. T., Jin, J., Mecchia, M. A., Debernardi, J. M., Palatnik, J. F., Balazadeh, S., Arif, M., Ó'Maoiléidigh, D. S., Wellmer, F., Krajewski, P., Riechmann, J. L., Angenent, G. C., & Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology*, 15(3). <https://doi.org/10.1186/gb-2014-15-3-r41>
- Pal, S. K., Liput, M., Piques, M., Ishihara, H., Obata, T., Martins, M. C. M., Sulpice, R., van Dongen, J. T., Fernie, A. R., Yadav, U. P., Lunn, J. E., Usadel, B., & Stitt, M. (2013). Diurnal changes of polysome loading track sucrose content in the rosette of wild-type arabidopsis and the starchless pgm mutant. *Plant Physiology*, 162(3), 1246–1265. <https://doi.org/10.1104/pp.112.212258>
- Palos, K., Yu, L., Railey, C. E., Nelson Dittrich, A. C., & Nelson, A. D. L. (2023). Linking discoveries, mechanisms, and technologies to develop a clearer perspective on plant long noncoding RNAs. *Plant Cell*, 35(6), 1762–1786. <https://doi.org/10.1093/plcell/koad027>
- Pang, B., & Snyder, M. P. (2020). Systematic identification of silencers in human cells. *Nature Genetics*, 52(3), 254–263. <https://doi.org/10.1038/s41588-020-0578-5>
- Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L., & Poethig, R. S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes and Development*, 18(19), 2368–2379. <https://doi.org/10.1101/gad.1231804>
- Peters, B., Casey, J., Aidley, J., Zohrab, S., Borg, M., Twell, D., & Brownfield, L. (2017). A conserved cis-regulatory module determines germline fate through activation of the transcription factor duo1 promoter. *Plant Physiology*, 173(1), 280–293. <https://doi.org/10.1104/pp.16.01192>
- Peters, L., & Meister, G. (2007). Argonaute Proteins: Mediators of RNA Silencing. *Molecular Cell*, 26(5), 611–623. <https://doi.org/10.1016/j.molcel.2007.05.001>
- Podlaha, O., & Zhang, J. (2009). Processed pseudogenes: the “fossilized footprints” of past gene expression. *Trends in Genetics*, 25(10), 429–434. <https://doi.org/10.1016/j.tig.2009.09.002>
- Poidevin, L., Forment, J., Unal, D., & Ferrando, A. (2020). Transcriptome and translome changes in germinated pollen under heat stress uncover roles of transporter genes involved in pollen tube growth. <https://doi.org/10.1101/2020.05.29.122937>
- Pral, W., Sharma, B., & Gregory, B. D. (2019). Transcription Is Just the Beginning of Gene Expression Regulation: The Functional Significance of RNA-Binding Proteins to Post-transcriptional Processes in Plants. *Plant and Cell Physiology*, 60(9), 1939–1952. <https://doi.org/10.1093/pcp/pcz067>
- Qin, P., Loraine, A. E., & McCormick, S. (2018). Cell-specific cis-natural antisense transcripts (cis-NATs) in the sperm and the pollen vegetative cells of Arabidopsis thaliana [version 1; referees: 2 approved]. *F1000Research*, 7(May), 1–10. <https://doi.org/10.12688/f1000research.13311.1>
- Qin, Y., Leydon, A. R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M. A., & Palanivelu, R. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genetics*, 5(8). <https://doi.org/10.1371/journal.pgen.1000621>
- Rabbani, N., Al-Motawa, M., & Thornalley, P. J. (2020). Protein glycation in plants—an

- under-researched field with much still to discover. *International Journal of Molecular Sciences*, 21(11). <https://doi.org/10.3390/ijms21113942>
- Rahmati Ishka, M., Brown, E., Weigand, C., Tillett, R. L., Schlauch, K. A., Miller, G., & Harper, J. F. (2018). A comparison of heat-stress transcriptome changes between wild-type *Arabidopsis* pollen and a heat-sensitive mutant harboring a knockout of cyclic nucleotide-gated cation channel 16 (*cngc16*). *BMC Genomics*, 19(1), 1–19. <https://doi.org/10.1186/s12864-018-4930-4>
- Rajkowitzsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M. F., Konrat, R., Bläsi, U., & Schroeder, R. (2007). RNA chaperones, RNA annealers and RNA helicases. *RNA Biology*, 4(3), 118–130. <https://doi.org/10.4161/rna.4.3.5445>
- Rieu, I., Twell, D., & Firon, N. (2017). Pollen development at high temperature: From acclimation to collapse. *Plant Physiology*, 173(4), 1967–1976. <https://doi.org/10.1104/pp.16.01644>
- Roy, B., & Arnim, A. G. von. (2013). Translational Regulation of Cytoplasmic mRNAs. *The Arabidopsis Book*, 11(11), e0165. <https://doi.org/10.1199/tab.0165>
- Ruiz-Ferrer, V., & Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology*, 60, 485–510. <https://doi.org/10.1146/annurev.arplant.043008.092111>
- Russell, J., & Zomerdijk, J. C. B. M. (2006). The RNA polymerase I transcription machinery. *Biochemical Society Symposium*, 73(73), 203–216. <https://doi.org/10.1042/bss0730203>
- Russell, S. D., Gou, X., Wong, C. E., Wang, X., Yuan, T., Wei, X., Bhalla, P. L., & Singh, M. B. (2012). Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. *New Phytologist*, 195(3), 560–573. <https://doi.org/10.1111/j.1469-8137.2012.04199.x>
- Russell, S. D., & Jones, D. S. (2015). The male germline of angiosperms: Repertoire of an inconspicuous but important cell lineage. *Frontiers in Plant Science*, 6(MAR), 1–10. <https://doi.org/10.3389/fpls.2015.00173>
- Rutley, N., Poidevin, L., Doniger, T., Tillett, R., Rath, A., Forment, J., Luria, G., Schlauch, K., Ferrando, A., Harper, J., & Miller, G. (2020). Characterization of novel pollen-expressed transcripts reveals their potential roles in pollen heat stress response in *Arabidopsis thaliana*. *BioRxiv*, 2020.08.20.258764. <https://www.biorxiv.org/content/10.1101/2020.08.20.258764v1>
- Schmitz, R. J., Grotewold, E., & Stam, M. (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *Plant Cell*, 34(2), 718–741. <https://doi.org/10.1093/plcell/koab281>
- Schulze, W. X. (2010). Proteomics approaches to understand protein phosphorylation in pathway modulation. *Current Opinion in Plant Biology*, 13(3), 279–286. <https://doi.org/10.1016/j.pbi.2009.12.008>
- Scutenaire, J., Deragon, J. M., Jean, V., Benhamed, M., Raynaud, C., Favory, J. J., Merret, R., & Bousquet-Antonelli, C. (2018). The YTH domain protein ECT2 is an m6A reader required for normal trichome branching in *Arabidopsis*. *Plant Cell*, 30(5), 986–1005. <https://doi.org/10.1105/tpc.17.00854>
- Seo, J. S., Sun, H.-X., Park, B. S., Huang, C.-H., Yeh, S.-D., Jung, C., & Chua, N.-H. (2017). ELF18-INDUCED LONG-NONCODING RNA Associates with Mediator to Enhance Expression of Innate Immune Response Genes in *Arabidopsis*. *The Plant Cell*, 29(5), 1024–1038. <https://doi.org/10.1105/tpc.16.00886>
- Sesma, A., Castresana, C., & Castellano, M. M. (2017). Regulation of translation by TOR,

- eIF4E and eIF2 $\alpha$  in plants: Current knowledge, challenges and future perspectives. *Frontiers in Plant Science*, 8(April), 2–8. <https://doi.org/10.3389/fpls.2017.00644>
- Shang, X., Cao, Y., & Ma, L. (2017). Alternative splicing in plant genes: A means of regulating the environmental fitness of plants. *International Journal of Molecular Sciences*, 18(2). <https://doi.org/10.3390/ijms18020432>
- Sharma, N., Russell, S. D., Bhalla, P. L., & Singh, M. B. (2011). Putative cis-regulatory elements in genes highly expressed in rice sperm cells. *BMC Research Notes*, 4, 1–10. <https://doi.org/10.1186/1756-0500-4-319>
- Singh, M. B., Lohani, N., & Bhalla, P. L. (2021). The Role of Endoplasmic Reticulum Stress Response in Pollen Development and Heat Stress Tolerance. *Frontiers in Plant Science*, 12(April), 1–12. <https://doi.org/10.3389/fpls.2021.661062>
- Singhal, N., Sharma, A., Kumari, S., Garg, A., Rai, R., Singh, N., Kumar, M., & Goel, M. (2020). Biophysical and Biochemical Characterization of Nascent Polypeptide-Associated Complex of *Picrophilus torridus* and Elucidation of Its Interacting Partners. *Frontiers in Microbiology*, 11(May), 1–10. <https://doi.org/10.3389/fmicb.2020.00915>
- Slotkin, R. K., Vaughn, M., Tanurdžic, M., Borges, F., Becker, J. D., Feijó, A., & Martienssen, R. A. (2010). *Epigenetic reprogramming and small RNA silencing of*. 136(3), 461–472. <https://doi.org/10.1016/j.cell.2008.12.038>. Epigenetic
- Sorenson, R., & Bailey-Serres, J. (2014). Selective mRNA sequestration by OLIGOURIDYLATEBINDING PROTEIN 1 contributes to translational control during hypoxia in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), 2373–2378. <https://doi.org/10.1073/pnas.1314851111>
- Sormani, R., Masclaux-Daubresse, C., Daniele-Vedele, F., & Chardon, F. (2011). Transcriptional regulation of ribosome components are determined by stress according to cellular compartments in *Arabidopsis thaliana*. *PLoS ONE*, 6(12). <https://doi.org/10.1371/journal.pone.0028070>
- Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), 613–626. <https://doi.org/10.1038/nrg3207>
- Sridevi Nidumukkala Lavanya Tayi, R. K. C. D. R. V., & Khareedu, V. R. (2019). DEAD box helicases as promising molecular tools for engineering abiotic stress tolerance in plants. *Critical Reviews in Biotechnology*, 39(3), 395–407. <https://doi.org/10.1080/07388551.2019.1566204>
- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765), 41–45. <https://doi.org/10.1038/47412>
- Swiezewski, S., Liu, F., Magusin, A., & Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature*, 462(7274), 799–802. <https://doi.org/10.1038/nature08618>
- Sze, H., Palanivelu, R., Harper, J. F., & Johnson, M. A. (2021). Holistic insights from pollen omics: Co-opting stress-responsive genes and ER-mediated proteostasis for male fertility. *Plant Physiology*, 187(4), 2361–2380. <https://doi.org/10.1093/plphys/kiab463>
- Tompa, P., & Kovacs, D. (2010). Intrinsically disordered chaperones in plants and animals. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, 88(2), 167–174. <https://doi.org/10.1139/o09-163>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.

- Nature Biotechnology*, 28(5), 511–515. <https://doi.org/10.1038/nbt.1621>
- Twell, D., Yamaguchi, J., Wing, R. A., Ushiba, J., & McCormick, S. (1991). Promoter analysis of genes that are coordinately expressed during pollen development reveals pollen-specific enhancer sequences and shared regulatory elements. *Genes and Development*, 5(3), 496–507. <https://doi.org/10.1101/gad.5.3.496>
- Twell, D., Klein, T. M., Fromm, M. E., & McCormick, S. (1989). Transient expression of chimeric genes delivered into pollen by microprojectile bombardment. *Plant Physiology*, 91(4), 1270–1274. <https://doi.org/10.1104/pp.91.4.1270>
- Twell, David, Oh, S. A., & Honys, D. (2006). Pollen development, a genetic and transcriptomic view. *Plant Cell Monographs*, 3(February), 15–45. [https://doi.org/10.1007/7089\\_042](https://doi.org/10.1007/7089_042)
- Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A. C., Hilbert, J.-L., Bartel, D. P., & Crété, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Molecular Cell*, 16(1), 69–79. <https://doi.org/10.1016/j.molcel.2004.09.028>
- Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4), 227–232. <https://doi.org/10.1038/nrg3185>
- Vogler, F., Konrad, S. S. A., & Sprunck, S. (2015). Knockin' on pollen's door: Live cell imaging of early polarization events in germinating arabidopsis pollen. *Frontiers in Plant Science*, 6(APR), 1–17. <https://doi.org/10.3389/fpls.2015.00246>
- Wang, H., Li, S., Li, Y., Xu, Y., Wang, Y., Zhang, R., Sun, W., Chen, Q., Wang, X. jie, Li, C., & Zhao, J. (2019). MED25 connects enhancer–promoter looping and MYC2-dependent activation of jasmonate signalling. *Nature Plants*, 5(6), 616–625. <https://doi.org/10.1038/s41477-019-0441-9>
- Wang, H. L. V., & Chekanova, J. A. (2017). Long noncoding RNAs in plants. *Advances in Experimental Medicine and Biology*, 1008(1), 133–154. [https://doi.org/10.1007/978-981-10-5203-3\\_5](https://doi.org/10.1007/978-981-10-5203-3_5)
- Wang, W., Vignani, R., Scali, M., Sensi, E., & Cresti, M. (2004). Post-translational modifications of  $\alpha$ -tubulin in *Zea mays* L. are highly tissue specific. *Planta*, 218(3), 460–465. <https://doi.org/10.1007/s00425-003-1122-4>
- Wang, Yi, Zhang, W. Z., Song, L. F., Zou, J. J., Su, Z., & Wu, W. H. (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in arabidopsis. *Plant Physiology*, 148(3), 1201–1211. <https://doi.org/10.1104/pp.108.126375>
- Wang, Ying, Luo, X., Sun, F., Hu, J., Zha, X., Su, W., & Yang, J. (2018). Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nature Communications*, 9(1), 1–9. <https://doi.org/10.1038/s41467-018-05829-7>
- Wang, Yuqiu, Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D., & Deng, X. W. (2014). Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28), 10359–10364. <https://doi.org/10.1073/pnas.1409457111>
- Wang, Z. W., Wu, Z., Raitskin, O., Sun, Q., & Dean, C. (2014). Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), 7468–7473. <https://doi.org/10.1073/pnas.1406635111>
- Waskiewicz, A. J., Flynn, A., Proud, C. G., & Cooper, J. A. (1997). Mitogen-activated protein



- kinases activate the serine/threonine kinases Mnk1 and Mnk2. *The EMBO Journal*, 16(8), 1909–1920. <https://doi.org/10.1093/emboj/16.8.1909>
- Watanabe, T., Cheng, E. C., Zhong, M., & Lin, H. (2015). Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Research*, 25(3), 368–380. <https://doi.org/10.1101/gr.180802.114>
- Wiedmann, B., Sakai, H., Davis, T. A., & Wiedmann, M. (1994). A protein complex required for signal-sequence-specific sorting and translocation. *Nature*, 370(6489), 434–440. <https://doi.org/10.1038/370434a0>
- Wierzbicki, A. T., Haag, J. R., & Pikaard, C. S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell*, 135(4), 635–648. <https://doi.org/10.1016/j.cell.2008.09.035>
- Woodson, S. A. (2010). Taming free energy landscapes with RNA chaperones. *RNA Biology*, 7(6), 677–686. <https://doi.org/10.4161/rna.7.6.13615>
- Wu, L., Fan, J., & Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11), 4034–4039. <https://doi.org/10.1073/pnas.0510928103>
- Wu, Z., Jankowski, V., & Jankowski, J. (2022). Irreversible post-translational modifications – Emerging cardiovascular risk factors. *Molecular Aspects of Medicine*, 86(August 2021), 101010. <https://doi.org/10.1016/j.mam.2021.101010>
- Xing, S., & Zachgo, S. (2008). ROXY1 and ROXY2, two Arabidopsis glutaredoxin genes, are required for anther development. *Plant Journal*, 53(5), 790–801. <https://doi.org/10.1111/j.1365-3113X.2007.03375.x>
- Yan, S., Bhawal, R., Yin, Z., Thannhauser, T. W., & Zhang, S. (2022). Recent advances in proteomics and metabolomics in plants. In *Molecular Horticulture* (Vol. 2, Issue 1). <https://doi.org/10.1186/s43897-022-00038-9>
- Yang, D. L., Huang, K., Deng, D., Zeng, Y., Wang, Z., & Zhang, Y. (2023). DNA-dependent RNA polymerases in plants. *Plant Cell*, 35(10), 3641–3661. <https://doi.org/10.1093/plcell/koad195>
- Yang, S., Song, C., Zhang, X., Jia, Y., Shi, Y., Ding, Y., & Gong, Z. (2018). OST1-mediated BTF3L phosphorylation positively regulates CBFs during plant cold responses. *The EMBO Journal*, e98228. <https://doi.org/10.15252/embj.201798228>
- Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, 25(21), 2227–2241. <https://doi.org/10.1101/gad.176826.111>
- Zehrmann, A., Verbitskiy, D., Van Der Merwe, J. A., Brennicke, A., & Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. *Plant Cell*, 21(2), 558–567. <https://doi.org/10.1105/tpc.108.064535>
- Zeng, Y., Wagner, E. J., & Cullen, B. R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Molecular Cell*, 9(6), 1327–1333. [https://doi.org/10.1016/s1097-2765\(02\)00541-5](https://doi.org/10.1016/s1097-2765(02)00541-5)
- Zhang, H., Zhong, H., Zhang, S., Shao, X., Ni, M., Cai, Z., Chen, X., & Xia, Y. (2019). NAD tagSeq reveals that NAD(+)-capped RNAs are mostly produced from a large number of protein-coding genes in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), 12072–12077. <https://doi.org/10.1073/pnas.1903683116>
- Zhang, Y., Wang, Y., Kanyuka, K., Parry, M. A. J., Powers, S. J., & Halford, N. G. (2008). GCN2-

- dependent phosphorylation of eukaryotic translation initiation factor-2alpha in Arabidopsis. *Journal of Experimental Botany*, 59(11), 3131–3141.  
<https://doi.org/10.1093/jxb/ern169>
- Zhang, Z., Hu, M., Feng, X., Gong, A., Cheng, L., & Yuan, H. (2017). Proteomes and Phosphoproteomes of Anther and Pollen: Availability and Progress. *Proteomics*, 17(20), 1–12. <https://doi.org/10.1002/pmic.201600458>
- Zhu, D., Wen, Y., Yao, W., Zheng, H., Zhou, S., Zhang, Q., Qu, L. J., Chen, X., & Wu, Z. (2023). Distinct chromatin signatures in the Arabidopsis male gametophyte. *Nature Genetics*, 55(4), 706–720. <https://doi.org/10.1038/s41588-023-01329-7>
- Zhu, H., Wang, G., & Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nature Reviews. Genetics*, 17(9), 551–565.  
<https://doi.org/10.1038/nrg.2016.83>
- Zilberman, D., Cao, X., & Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science (New York, N.Y.)*, 299(5607), 716–719. <https://doi.org/10.1126/science.1079695>
- Zulawski, M., Schulze, G., Braginets, R., Hartmann, S., & Schulze, W. X. (2014). The Arabidopsis Kinome: Phylogeny and evolutionary insights into functional diversification. *BMC Genomics*, 15(1). <https://doi.org/10.1186/1471-2164-15-548>