

Palacký University Olomouc
Faculty of Science
Department of Cell Biology and Genetics



and

Institute of Experimental Botany of the Czech Academy of Sciences
Centre of Plant Structural and Functional Genomics



**Analysis of transcriptome dynamics in developing
barley seeds**

Ph.D. Thesis

Martin Kovačik

Olomouc 2024

Supervisor: Assoc. Prof. Dr. Aleš Pečinka
Co-supervisor: Ing. Anna Nowicka Ph.D.

Podklad pro zadání DISERTAČNÍ práce studenta

Jméno a příjmení: **Mgr. Martin KOVAČIK**
Osobní číslo: **R180056**
Adresa: **Břenkova 2966/15, Ostrava – Zábřeh, 70030 Ostrava 30, Česká republika**
Téma práce: **Analýza dynamiky transkriptomu v průběhu vývoje semene ječmene setého**
Téma práce anglicky: **Analysis of transcriptome dynamics of developing barley seeds**
Jazyk práce: **Angličtina**
Vedoucí práce: **doc. Mgr. Aleš Pečinka, Ph.D.**
Katedra buněčné biologie a genetiky

Zásady pro vypracování:

1. Vývoj protokolu pro separaci endospermu, embrya a osemení v průběhu vývoje semen ječmene setého (*Hordeum vulgare*).
2. Analýza transkriptomu separovaných pletiv v jednotlivých fázích vývoje endospermu.
3. Analýza aktivity PRC2 komplexu, depozice H3K27me3 a identifikace imprintovaných genů ve vývoji endospermu ječmene setého.

Seznam doporučené literatury:

JIANG, Hua; KÖHLER, Claudia. Evolution, function, and regulation of genomic imprinting in plant seed development. *Journal of experimental botany*, 2012, 63.13: 4713-4722.
ERILOVA, Aleksandra, et al. Imprinting of the polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis. *PLoS Genetics*, 2009, 5.9: e1000663.
SABELLI, Paolo A.; LARKINS, Brian A. The development of endosperm in grasses. *Plant physiology*, 2009, 149.1: 14-26.
MAYER, K. F., et al. A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 2012, 491: 711-716.
OLSEN, Odd-Arne. Endosperm development: cellularization and cell fate specification. *Annual review of plant biology*, 2001, 52.1: 233-267.

Podpis školitele:

Datum:

1. 2. 2024

Podpis vedoucího pracoviště:

Datum:

prof. RNDr. Zdeněk Dvořák, DrSc. et Ph.D.

Bibliographical identification

Author's name	Martin Kovačik
Title	Analysis of transcriptome dynamics of developing barley seeds
Type of thesis	Ph.D. Thesis
Department	Department of Cell Biology and Genetics, Palacký University Olomouc and Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany of the Czech Academy of Sciences
Supervisor	Assoc. Prof. Dr. Aleš Pečinka
Co-supervisor	Ing. Anna Nowicka, Ph.D.
The year of presentation	2024

Abstract

Cereal grains constitute an important source of food and feed, owing to their large endosperm. Seed development initiates with double-fertilization, wherein two sperm nuclei migrate into embryo sac; one fuses with the egg cell nucleus, and the second with the central cell nucleus. This intricate process results in the formation of the diploid embryo and triploid endosperm being surrounded by a diploid seed coat of maternal origin. Understanding the regulatory mechanisms governing seed development in cereals is essential for plant breeding and yield enhancement. Although extensive research has been conducted in model systems such as *Arabidopsis thaliana* and maize, comprehensive information on these processes in barley has remained incomplete.

To provide detailed spatiotemporal insights into gene expression in developing seeds of cultivated barley, we established a detailed protocol for the manual dissection of high-purity barley seed tissues and conducted a transcriptomic analysis of the embryo, endosperm, and seed maternal tissues dissected from grains 4 to 32 days after pollination.

Our analysis of differential gene expression and co-expression networks identified specific groups of genes and transcription factors pivotal for developmental transitions

and marker genes relevant for studying endosperm differentiation. Expression of histones and Polycomb repressive complex 2 subunits indicated that epigenetic processes are highly dynamic and play a significant role in the development of barley endosperm. The repressive modification H3K27me3 is globally reduced in endosperm tissues, particularly at genes associated with accumulation of storage compounds. Our comparative approach to identify conserved imprinted genes established a foundation for future studies on genomic imprinting in barley.

The transcriptome atlas we have developed provides a solid base for further biological investigations into the key factors influencing barley grain development.

Keywords plants, barley, *Hordeum vulgare*, seed, embryo, endosperm, seed maternal tissues, RNA-seq, transcriptome, epigenetics, genomic imprinting

Number of pages/appendices 108/VIII

Language English

Bibliografická identifikace

Jméno autora	Martin Kovačik
Název práce	Analýza dynamiky transkriptomu v průběhu vývoje semene ječmene setého
Typ práce	Disertační práce
Pracoviště	Katedra buněčné biologie a genetiky, Univerzita Palackého v Olomouci a Centrum strukturní a funkční genomiky rostlin, Ústav experimentální botaniky Akademie věd České republiky
Školitel	Assoc. Prof. Dr. Aleš Pečinka
Školitel specialista	Ing. Anna Nowicka, Ph.D.
Rok obhajoby práce	2024

Abstrakt

Obiloviny jsou důležitým zdrojem potravy a krmiva díky jejich velkému endospermu. Vývoj semene začíná dvojitým oplozením, při kterém dvě spermatická jádra migrují do zárodečného vaku; jedno splyne s jádrem vaječné buňky a druhé s jádrem centrální buňky. Tento složitý proces vede k vytvoření diploidního embrya a triploidního endospermu, které jsou obklopeny diploidním semenným obalem mateřského původu. Porozumění regulačním mechanismům, které řídí vývoj semene u obilovin, je nezbytné pro šlechtění rostlin a zvýšení výnosu. Ačkoliv byl proveden rozsáhlý výzkum na modelových systémech, jako jsou *Arabidopsis thaliana* a kukuřice, podrobné informace o těchto procesech u ječmene zůstaly neúplné.

Abychom poskytli detailní časoprostorové informace o genové expresi ve vyvíjejících se semenech ječmene, vyvinuli jsme podrobný protokol pro ruční disekci pletiv vysoké čistoty ze semen ječmene a provedli transkriptomickou analýzu embrya, endospermu a semenného obalu disektovaných ze zrn 4 až 32 dní po opylení.

Naše analýza diferenciální genové exprese a koexpresních sítí identifikovala specifické skupiny genů a transkripčních faktorů klíčových pro vývojové přechody a

markerové geny relevantní pro studium diferenciacce endospermu. Expresce histonů a podjednotek Polycomb represivního komplexu 2 ukázala, že epigenetické procesy jsou vysoce dynamické a hrají významnou roli ve vývoji endospermu ječmene. Množství represivní modifikace H3K27me3 je v endospermu globálně nižší, zejména u genů spojených s akumulací zásobních látek. Naše komparativní analýza vedoucí k identifikaci konzervovaných imprintovaných genů vytvořila základ pro budoucí studie genomového imprintingu u ječmene.

Transkriptomický atlas, který jsme vyvinuli, poskytuje pevnou základnu pro další studie klíčových faktorů ovlivňujících vývoj zrna ječmene.

Klíčová slova	rostliny, ječmen, <i>Horedum vulgare</i> , semínko, embryo, endosperm, obalové vrstvy, RNA-seq, transkriptom, epigenetika, genomický imprinting
Počet stran/příloh	108/VIII
Jazyk	Anglický

Declaration

I hereby declare that I have written this thesis independently under the supervision of Assoc. Prof. Dr. Aleš Pečinka using the information sources listed in the References.

In Olomouc

Acknowledgements

First, I would like to say thanks to my supervisor Assoc. Prof. Dr. Aleš Pečinka, who was willing to guide me, inspire me, and trust me with this interesting and challenging project. He supported my efforts every day during my Ph.D. studies.

My thanks also belong to my dear colleague Ing. Anna Nowicka, Ph.D., for her enormous support. She was always here for me, ready to help me, advise me, and listen to me. Her positive motivation, enthusiasm, and faith was the best motivation I could ever get.

None of this could happen without the former and current head of the laboratory Prof. Ing. Jaroslav Doležel, DrSc., and Mgr. Jan Bartoš, Ph.D, who allows me to work at the Centre of Plant Structural and Functional Genomics. I want to thanks to them and to all other colleagues for a friendly and inspirative environment.

My gratitude belongs also to the most important people in my life. I want to thank to my parents, who was always here ready to support me. I would like to thank my loving wife for her compassion and understanding. The knowledge that she is next to me all the time was a great support for me. At the very end, my sincere gratitude belongs to our two children, who make me enjoy life every day.

Experiments presented in this thesis were supported by Czech Science Foundation grants 18-12197S and 21-02929S and European Regional Development Fund project “Plants as a mean of sustainable development” (CZ.02.1.01/0.0/0.0/16_019/0000827). Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic under the Projects CESNET (Project No. LM2015042) and CERIT-Scientific Cloud (Project No. LM2015085) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Table of contents

List of abbreviations	x
1 Literature overview	1
1.1 Cereal crops and the importance of barley	1
1.2 Grain structures and development	2
1.2.1 Embryo development	4
1.2.2 Endosperm development	5
1.2.3 Seed maternal tissues	10
1.3 Organization and structure of barley genome	12
1.3.1 Genes	13
1.3.2 Cis-regulatory elements.....	13
1.3.3 Repetitive sequences	14
1.3.4 Transposable elements	15
1.4 Epigenetic regulation	16
1.4.1 DNA methylation	16
1.4.2 Histone variants and modifications	18
1.4.3 Genomic imprinting	20
1.5 Gene expression analysis	24
1.5.1 Global gene expression analysis	24
1.5.2 Targeted gene expression analysis	25
2 Aims of the thesis	28
2.1 Development of a protocol for dissection of endosperm, embryo and seed maternal tissues during seed development of barley (<i>Hordeum vulgare</i>)	28
2.2 Transcriptome analysis of barley seed development	28
2.3 Analysis of PRC2 complex expression, H3K27me3 distribution and identification of imprinted genes during barley seed development	29
3 Results	30
3.1 High purity tissue isolation from developing barley seeds.....	30
3.2 Transcriptome landscape of developing barley seeds.....	34
3.3 Implications of PRC2 complex expression in H3K27me3 epigenetic landscape and identification of conserved imprinted genes in developing barley seed	42
3.4 Original publications.....	47
3.4.1 Isolation of High Purity Tissues from Developing Barley Seeds	48

3.4.2	Transcriptome landscape of developing barley seeds	49
3.5	Co-author publications	50
3.5.1	Dynamics of endoreduplication in developing barley seeds	51
3.5.2	Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel..	52
3.5.3	Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues	53
3.5.4	Core promoterome of barley embryo	54
3.6	Published abstracts	55
3.6.1	Analysis of transcriptome landscape in developing barley seeds	56
3.6.2	Transcriptome landscape of endosperm in developing barley seeds	57
3.6.3	Developing an atlas of gene expression during barley grain development.....	58
4	Discussion	59
4.1.1	Isolation of High Purity Tissues from Developing Barley Seeds	59
4.1.2	Transcriptome landscape of developing barley seeds	60
4.1.3	Expression of PRC2 complex, H3K27me3 epigenetic landscape and conserved imprinted genes in seed development	62
5	Conclusions.....	64
6	References.....	65
7	List of appendices	87

List of abbreviations

5mC	5'-methyl cytosine
ABA	Abscisic acid
AL	Aleurone layer
BAC	Bacterial artificial chromosome
BAR	The Bio-Analytic Resource for Plant Biology
BETL	Basal endosperm transfer layer
cDNA	Complementary DNA
ChIP-seq	Chromatin immuno-precipitation sequencing
CRE	Cis-regulatory element
CRM	Cis-regulatory module
C _t	Threshold cycle
DAF	Days after flowering
DAP	Days after pollination
DEG	Differentially expressed gene
DMR	Differentially methylated region
DOP	Day of pollination
dsRNA	Double-stranded RNA
ESR	Embryo surrounding region
EST	Expressed sequence tag
ETC	Endosperm transfer cells
FANS	Fluorescence-activated nuclei sorting
FDR	False discovery rate
GFP	Green fluorescent protein
GO	Gene ontology
GUS	β-glucuronidase
LEA	Late embryogenesis abundant
LINE	Long interspersed nuclear element
LTR	Long terminal repeat
MC	Motif cluster
MEG	Maternally expressed gene
MITE	Miniature inverted-repeat transposable element

mRNA	Messenger RNA
PCA	Principal component analysis
PCD	Programmed cell death
PcG	Polycomb group
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PCR	Polymerase chain reaction
PEG	Paternally expressed gene
qPCR	Quantitative PCR
RdDM	RNA-directed DNA methylation
rDNA	Ribosomal DNA
RIN	RNA integrity number
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA
SE	Starchy endosperm
SINE	Short interspersed nuclear element
scRNA-seq	Single cell RNA sequencing
siRNA	Small interfering RNA
SMT	Seed maternal tissue
TE	Transposable element
TF	Transcription factor
TIR	Terminal inverted repeat
TPM	Transcripts per million reads
TSS	Transcription start site
WGCNA	Weighted gene co-expression network analysis

1 Literature overview

1.1 Cereal crops and the importance of barley

Cereals are species of the grass family *Poaceae* that are used as crops to produce edible seeds, also referred to as grains. There is a number of cereals that are grown in different climate zones, including rye, oats, barley, wheat, maize and sorghum (McKevith, 2004). Cereal seeds represent an important source of food, feed, and industrial raw materials for humans and domestic animals. These seeds contribute to over 60% of the global food resources, making them the most valuable product of plant production. Global cereal production was predicted to reach 2.7 billion tones in 2022. Within this context, barley stands out as the fourth most important temperate cereal crop, with over 145 million tones in production just after wheat, maize, and rice (Crop Prospects and Food Situation #1, March 2023, 2023).

Cultivated barley (*Hordeum vulgare* L. subsp. *vulgare*) has become an important model organism for scientific research, offering insights into organ development, genetic diversity and epigenetics (Baker et al., 2015; Jayakodi et al., 2020; Thiel et al., 2021; Hertig et al., 2023). Its diploid nature, and close evolutionary relationships to other cereal crops make it an ideal candidate for the study of fundamental biological processes. From taxonomical point of view, barley belongs to the globally and economically important tribe *Triticeae* L. within the large grass family (*Poaceae* Barnhart). This tribe is characterized by its large genomes, basic chromosome number of $x = 7$, and spike inflorescences. Besides barley, *Triticeae* includes major temperate zone cereals such as bread wheat (*Triticum aestivum* L.) and rye (*Secale cereale* L.), as well as forage grasses like Russian wildrye (*Psathyrostachys juncea* (Fisch.) Nevski). Barley grains are utilized predominantly for animal feed, malt, and human food. Cultivated across approximately 49 million hectares worldwide, barley prospers in a wide range of environmental conditions, ranging from Alaska and the Nordic countries of Finland and Norway to the borders of the Sahara Desert in Algeria (Akar Taner et al., 2004; Shewry and Ullrich, 2016; FAO, 2023). In several developing countries with arid and semi-arid climates, barley dominates other grains as the only cereal and primary food staple.

From a historical perspective, barley is documented as one of “Neolithic founder crops” and was one of the first cereals that became a pillar of food and feed for ancient

societies (Badr et al., 2000). It is estimated that genus *Hordeum* L. diverged approximately 13 million years ago and passed domestication about 8000 B.C. (Badr et al., 2000; Fetch et al., 2003; Matus et al., 2003). The primary gene pool for cultivated barley originates from the wild subspecies progenitor *H. vulgare* subsp. *spontaneum* (K.Koch) Asch. & Graebn. Currently, this subspecies is being increasingly explored as a source of genes for trait improvement in cultivated barley (Fetch et al., 2003; Matus et al., 2003). Additionally, progress in barley genomics, coupled with recent efforts in sequencing and annotation of barley genome and pangenome, significantly contributed to an in-depth understanding of gene functions (Sreenivasulu et al., 2008). So far, a broad spectrum of resources has been developed to facilitate systematic analysis of the barley genome. These resources include extensive collections of barley accessions, molecular markers, EST collections, BAC libraries, mutant collections, DNA arrays, and advanced technologies such as the double haploids and efficient transformation protocols (Sreenivasulu et al., 2008; CGIAR, 2023). The availability of well-characterized genetic resources and genomic tools facilitates the discovery of fundamental mechanisms related to growth and development, offering insights beneficial to both barley and other crop species.

1.2 Grain structures and development

From a botanical point of view, seeds are the result of ovule fertilization. The female gametophyte of gymnosperms usually contains one or several egg cells surrounded by multicellular nursing tissue, whereas in angiosperms, the female gametophyte is usually reduced to an eight-nucleated, seven-celled embryo sac containing the egg and central cell (Maheshwari, 1950; Erbasol Serbes et al., 2019). The fertilization is initiated when the pollen tube delivers two haploid sperm cells to the embryo sac (Peris et al., 2010). One sperm cell fertilizes the haploid egg cell, generating the diploid embryo, while the second sperm cell fuses with the diploid central cell, forming the triploid endosperm (West and Harada, 1993; Goldberg et al., 1994). After fertilization, the seed is composed of three main compartments: the embryo, the endosperm, and the seed maternal tissues (SMTs). The embryo, which develops into the future adult plant, contains all the essential elements and patterns necessary for post-germination growth. The endosperm, dedicated to being nutrient reservoir for the developing embryo, varies across *Angiosperms*. In species with a non-persistent (transient) endosperm, it is consumed before seed

maturation and nutrients are stored in the cotyledons. In other species, such as cereal grains, endosperm contains a persistent nutrient storage compartment (also known as permanent endosperm), being a major source of food and industrial value for human nutrition (Figure 1) (Cocucci, 2005; Lu and Magnani, 2018; Baroux and Grossniklaus, 2019). SMTs, deriving from the integuments of the ovule, provide protection against mechanical injury, predation, and drying out (Locascio et al., 2014; Zhang et al., 2016).

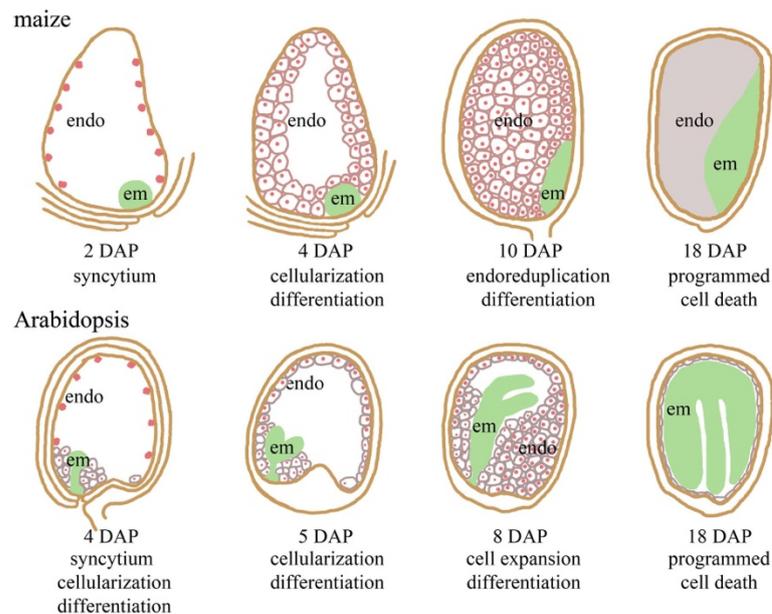


Figure 1: Distinct developmental stages of seeds with persistent (maize) and non-persistent endosperm (Arabidopsis). Initially, endosperm develops as a coenocyte, subsequently undergoes cellularization, followed by cell differentiation. In seeds with persistent endosperm, this compartment stays until seed maturation. In contrast, in seeds with non-persistent endosperm most of the tissue is absorbed by the embryo prior to seed maturation. endo – endosperm nuclei (red); em – embryo (green), DAP – day after pollination (inspired by Bai and Settles, 2015).

Overall, the embryo and endosperm share genetic identity but differ in their ploidy level and parental genome contributions. The embryo contains one maternal and one paternal genome (1m:1p), whereas in endosperm two maternal and one paternal genome (2m:1p) are present. Consequently, the embryo and endosperm follow distinct developmental trajectories within seed development in cereals (Figure 2) (Brown et al., 1999; Kiesselbach, 1999; Chandler et al., 2008; Peris et al., 2010). There are three major phases of growth. Stage I stands for early development, which is initiated by double fertilization followed by cell proliferation and a slight weight gain; stage II (middle) comprises differentiation of the main seed tissues and a large weight increase

accompanied by storage compounds accumulation; stage III, including late stages of seed development, corresponds to seed maturation, desiccation accompanied by weight reduction, and finally physiological maturation and dormancy (Sabelli and Larkins, 2009; Sreenivasulu et al., 2010; Dante et al., 2014).

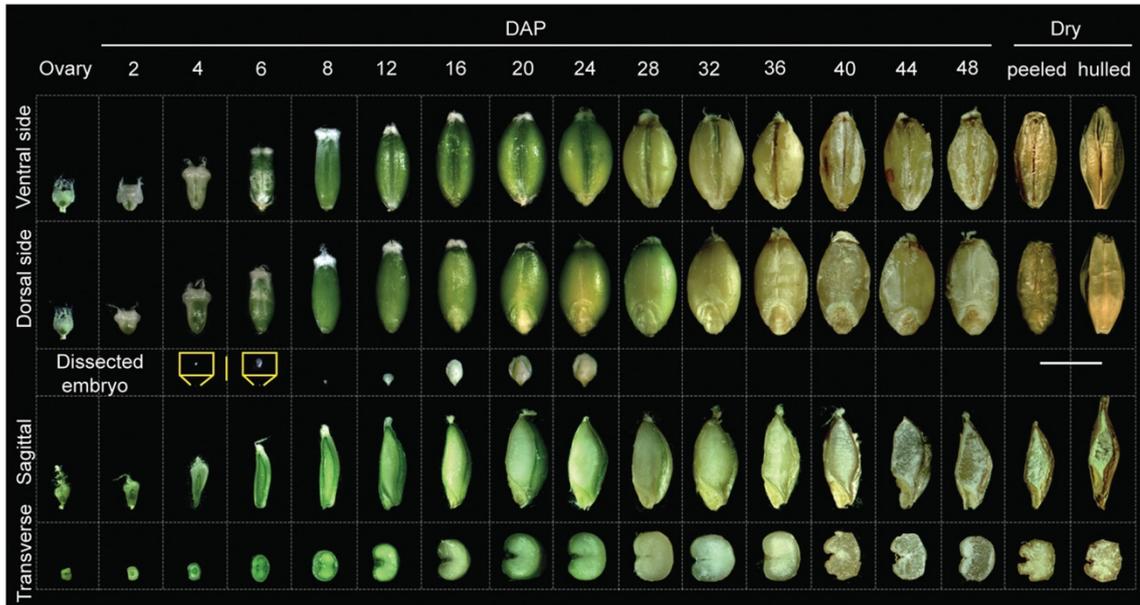


Figure 2: Phenotype of developing barley seeds of cv. Compana. (A) Developmental series from 0 (ovary) to 48 DAP and dry seeds. Seeds for sagittal and transverse sections were cut in half. The yellow insets show early-stage embryos. Scale bar = 5 mm, inset bar = 500 μ m. Figure adopted from (Nowicka et al., 2021a).

1.2.1 Embryo development

Embryogenesis in barley is a complex and well-coordinated developmental process, that defines the final architecture and functionality of the seedling. The embryo growth from a single cell to differentiated tissue is marked by a series of developmental milestones, each underpinned by complex cellular processes.

The initial stages of barley embryogenesis involve the establishing of the basic morphological plan. This process begins with an asymmetric cell division of the zygote, essential for forming an apical-basal axis (Chandler et al., 2008; Peris et al., 2010). This division produces a hypobasal cell, the precursor to the suspensor, and an epibasal cell, from which the embryo proper develops (Engell, 1989). The transition from a globular to a club-shaped proembryo involves organizing genetic programs that govern meristem

specification and lateral organ initiation (Chandler et al., 2008; De Smet and Beeckman, 2011; Wendrich and Weijers, 2013).

The localization of each cell is crucial for its differentiation, exposing it to a specific combination of mechanical and environmental stimuli, such as hormone gradients and signaling molecules (Dupuy et al., 2008; Mirabet et al., 2011; van Norman et al., 2011). These factors influence cell fate decisions, highlighting the importance of positional information over cellular lineage (Ingram, 2004).

Barley embryogenesis encompasses the differentiation of structures like the scutellum, a specialized tissue analogous to a cotyledon yet distinct in its molecular patterning and function (Werker, 1997; Tillich, 2007). The coleoptile and coleorhiza also develop, enclosing the shoot and root apical meristems, respectively (Merry, 1941; Shewry and Ullrich, 2016). A mature barley embryo is already highly differentiated, featuring adventitious root buds and leaf primordia, with the embryonic tiller bud forming in the axil of the first leaf primordium (Bonett, 1966; Luxová, 1986; Babb and Muehlbauer, 2003). The transition from radial to bilateral symmetry, marked by differentiation of the scutellum and cotyledon, and the establishment of a single bilateral axis, can be inhibited by auxin inhibitors, as observed in wheat (Fischer and Neuhaus, 1996). The scutellum and embryonic axis differentiation signify morphogenesis completion, typically around two weeks post-fertilization (Merry, 1941).

Post-morphogenetic development is characterized by growth predominantly through cell wall extension and storage reserve accumulation (Raghavan, 2006). During this phase, significant increases in fresh weight, embryo length, and protein content are observed (Duffus and Rosie, 1975; Nielsen, 1990). The barley embryo accumulates globulins, as the primary protein-storage components and lipids, mainly in the form of triacylglycerols within oil bodies in the scutellum (Price and Parsons, 1979; Bhatta and Rosnagel, 1980; Gubatz and Shewry, 2010).

1.2.2 Endosperm development

The development of persistent endosperm tissues of barley begins with double fertilization (refer to 1.2). Initially, the fertilized endosperm comprises a proximal mass of cytoplasm and a thin cytoplasmic layer surrounding a large central vacuole. Its development encompasses several not sharply delimited phases: coenocyte formation, cellularization, cellular differentiation, accumulation of storage compounds, and

maturation, which includes programmed cell death (PCD), maturation, and desiccation (Olsen, 2004; Sabelli and Larkins, 2009).

During the coenocytic phase, the primary endosperm nuclei undergo multiple rounds of rapid, synchronized divisions without cell wall formation and cytokinesis. In contrast, cell division in the early maize embryo is significantly slower, occurring four to eight times slower than in the endosperm (Sabelli and Larkins, 2009). This difference may be attributed to the dissimilarities in cellular components synthesized during endosperm and embryonic cell proliferation; notably, endosperm development does not involve the creation of cytoplasm, cell membranes, and cell walls. The reduction in endosperm proliferation rate, coinciding with the cellularization of the coenocyte, supports this interpretation (Bennett et al., 1975). The coenocytic development of endosperm is possibly an evolutionary strategy, to rapidly populate the large cytoplasm of the central cell with a higher basal cell number, supporting the growth of this tissue and its nutritive function (Sabelli and Larkins, 2009). Initially, all endosperm nuclei divide synchronously. Subsequently, developmental gradients emerge, leading to a synchronized proliferation of neighboring nuclei. Within the first three days after pollination (DAP), up to 512 nuclei in maize, and up to 2,000 coenocytic nuclei in *Triticum* and *Hordeum* species, are generated (Walbot, 1994; Bennett et al., 1975). These nuclei become distributed along the periphery of the primary endosperm cell, due to the enlargement of its central vacuole. In barley, a 2-day hiatus follows the initial proliferation phase, coinciding with a dramatic cytoskeletal rearrangement and preparation for the subsequent cellularization of the initial layer of nuclei (Olsen, 2001, 2004; Sabelli and Larkins, 2009).

During cellularization, the coenocytic endosperm transforms through the formation of the internuclear radial microtubule systems (Figure 3A-C). Microtubules, radiating from the nuclear surface, define initial nuclear-cytoplasmic domains. Cell wall formation is initiated at the intersection of repolarizing microtubules through the deposition of an adventitious phragmoplast. This is followed by centripetal extension of the cell walls in an open-ended alveolation mechanism, proceeding from the periphery toward the central vacuole (Figure 3D-F). Nuclei divide synchronously and periclinally, immediately followed by cytokinesis, moving the alveolar layer inwards, overlying residual coenocyte cytoplasm (Figure 3G). By 3 to 6 DAP, the central cell cavity becomes completely cellularized (Brown et al., 1994, 1999; Olsen, 2004).

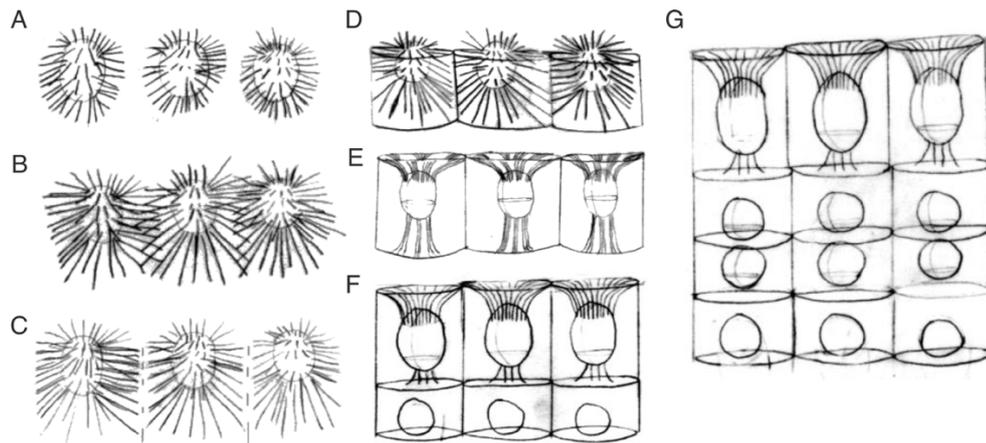


Figure 3: Process of endosperm cellularization Formation of radial microtubular systems (RMS) and anticlinal cell wall formation. RMS forms at the nuclear membrane (A), extends and initially overlaps (B), but soon cytoplasmic phragmoplasts that deposit anticlinal cell walls are formed (C). Partial anticlinal cell walls are formed around each nucleus (D), and walls extend centripetally toward the central vacuole (E). The second layer of alveoli develops (F) and continues in centripetal growth toward the central vacuole (G) (inspired by Olsen, 2001).

Cellularization and differentiation are closely integrated processes, both temporally and spatially. As cellularization progresses, cells begin to differentiate. Endosperm differentiation predominantly occurs between 4 to 6 DAP, resulting in the formation of the endosperm transfer cells (ETC), embryo surrounding region (ESR), and aleurone layer (AL), originating from peripheral cells, and starchy endosperm (SE) deriving from internal cells (Brown et al., 1994, 1999; Olsen, 2001; Becraft and Gutierrez-Marcos, 2012; Leroux et al., 2014).

The ESR comprises the cells lining the cavity of the endosperm where embryo development takes place. As the embryo grows, the ESR gradually reduces in size, and by the early to middle stages of endosperm development, only minimal fragments of the ESR are left at the base of the endosperm (Sabelli and Larkins, 2009). In maize, these cells are distinguished by their dense cytoplasm and cell-specific expression of *ESR* (*EMBRYO SURROUNDING REGION*) genes. While the precise function of the ESR as well as *ESR* genes remains elusive, they are hypothesized to serve as both a physical barrier and communication interface between the endosperm and the embryo (Olsen, 2001, 2004; Balandín et al., 2005).

The ETC, known in the maize as the basal endosperm transfer layer (BETL), develops above the main vascular tissue of the maternal plant, enabling nutrient transport into the endosperm. These cells feature significant secondary wall ingrowths and

extensive endomembrane systems in the early grain stage. Their location varies by species, being situated over the nucellar projection in barley and wheat, and above the chalazal pad in maize (Rost et al., 1984). Recent research has pinpointed several genes expressed exclusively in BETL, such as *MRP-1* (*MYB-RELATED PROTEIN-1*), a transcription factor (TF) known to activate BETL-specific genes (Gómez et al., 2002; Gutiérrez-Marcos et al., 2004; Gómez et al., 2009). In barley, a distinct transcript in ETC, *END1*, is noticeable in the endosperm coenocyte above the maternal vascular tissue, and differentiation of transfer cells occurs within a narrow temporal window (Doan et al., 1996). *END1* also persists in fully developed grains. In maize, *BETL* (*BASAL ENDOSPERM TRANSFER LAYER*), *BAP* (*BASAL LAYER ANTIFUNGAL PROTEIN*) and *EBE* (*EMBRYO-SAC BASAL-ENDOSPERM LAYER EMBRYO-SURROUNDING-REGION*) were identified (Magnard et al., 2003). While the exact functions of these genes are unclear, a possible link between *BETL1*, *BETL3* and plant defensins suggests a role in embryo protection (Olsen, 2001).

The AL in barley consists of three cell layers that cover the entire perimeter of the endosperm, except where transfer cells are located and the endosperm touches the embryo (Olsen, 2001; Sabelli and Larkins, 2009). These cells undergo endoreduplication (explained below), and feature dense, granular cytoplasm with aleurone grains and small vacuoles (Buttrose, 1963; Nowicka et al., 2021b). Mature aleurone cells are cuboidal in cross-section (Sabelli and Larkins, 2009). Aleurone cell differentiation starts with changes in the microtubule cytoskeletal cycle, differing from SE. Peripheral cells retain the complete cytoskeletal arrays of meristematic cells, including the pre-prophase band that orients cell wall deposition during division (Brown et al., 1994; Sabelli and Larkins, 2009). This division plane is crucial for the AL growth and surface area expansion (Olsen, 2004). Early molecular markers in barley, such as *LTP2* (*LIPID TRANSFER PROTEIN 2*), appear following initial phenotypic changes. Other significant markers include *LTP1*, *B22E*, *pZE40*, *OLE-1*, *OLE-2*, *PER-1*, *CHI33* and *C1* (Klemsdal et al., 1991; Madrid, 1991; Smith et al., 1992; Kalla et al., 1994; Leah et al., 1994; Stacy et al., 1999).

The largest part of cell mass in endosperm is represented by the SE, which contains at least three distinct areas: the central starchy endosperm, the conducting zone, and the subaleurone layer (Becraft, 2001; Olsen, 2001, 2004; Sabelli and Larkins, 2009). Its primary role is starch synthesis, facilitated by enzymes like ADP-glucose, starch synthase, and branching and debranching enzymes (Smith, 1999). Additionally, SE contains prolamins, proteins whose expression largely depends on the P-box promoter

motif (Mena et al., 1998). SE cells originate from the inner daughter cell of the first periclinal division, differing from aleurone cells in lacking a preprophase band. Consequently, SE cell walls are formed randomly, disrupting the uniformity and pattern of cell files after a few divisions, and resulting in non-uniform cells, especially at the periphery where they are smaller and vary in storage protein concentration (Olsen, 2004). Most of these cells are likely derived from aleurone cells redifferentiated to SE (Olsen, 2001). Research on barley's SE is advanced by the availability of *shrunkened endosperm xenia* (*sex*) and *shrunkened endosperm gene* (*seg*) mutants, although the causal genes for most of them have not been identified. However, the *sex3* phenotype was recently linked to a mutation in the P-box binding transcription factor (PBF), also conserved in wheat (Mena et al., 1998; Orman-Ligeza et al., 2020).

Two crucial processes occur in the second half of endosperm development in cereals, endopolyploidization and PCD. Endopolyploidization is characterized by the duplication of nuclear genetic material without concomitant cell division, leading to an increased DNA content, a phenomenon particularly prominent in cereal endosperm, or *A. thaliana* (Arabidopsis) epidermal pavement and hair cells (Chojecki et al., 1986; Melaragno et al., 1993; Roeder et al., 2010). This process is typically associated with larger cell sizes and enhanced transcriptional and biosynthetic capabilities (Larkins et al., 2001; Breuer et al., 2010). A higher number of DNA templates may support increased rates of transcription and translation, thereby enhancing gene expression, protein synthesis, and metabolic activity. The chromatin in endoreduplicated endosperm nuclei tends to be loosely condensed, which might allow greater accessibility for TFs and thus facilitate higher transcription rates (Sabelli, 2012). Additionally, endoreduplication optimizes the ratio of cell volume to cell walls, keeping it minimal. This is beneficial as cell wall biosynthesis is an energy-intensive process. By minimizing the need for extensive cell wall formation, endoreduplication allows for more efficient synthesis of storage products (Kowles, 2009).

Upon completion of storage compound accumulation, the SE cells in cereals undergo PCD, a strategically timed event (Sabelli, 2012). PCD is thought to facilitate the hydrolysis of nutrients and their subsequent uptake by the embryo during seed germination (Nguyen et al., 2007). In developing barley seeds, a distinct pattern of PCD has been observed. Cell death in the endosperm begins at 8 DAP and increases over time, except in the AL, which usually contains only few dead cells around 48 DAP (Nowicka et al., 2021a). In the SE of barley, PCD is characterized by selective proteolysis involving

the activation of the proteasomal complex, the ubiquitin pathway, and the F-box protein machinery. Although the SE is dead at maturity, most cellular components and cell walls remain intact until seed germination (Sreenivasulu et al., 2006).

As the grain matures, its water content decreases, leading to a desiccated state essential for survival during periods of low metabolism and preparing it for germination under favorable conditions (Kermode and Finch-Savage, 2002). This transition is marked by significant transcriptional and metabolic changes, enabling seeds to maintain viability over extended periods and respond appropriately to germination cues (Finkelstein et al., 2008; Angelovici et al., 2010). Abscisic acid (ABA) plays a central role in inducing and controlling desiccation. During barley grain maturation, ABA levels peak around 14 days after flowering (DAF), followed by a second increase from 35 DAF in isolated embryos, which become capable of enduring severe desiccation initiated from approximately 15 DAF (Bartels et al., 1988; Morris et al., 1991). ABA's influence on the expression of late-embryogenesis-abundant (LEA) proteins underscores its importance in promoting desiccation tolerance (Espelund et al., 1995; Shen et al., 2004). LEA proteins protect cells by remaining unfolded when hydrated and adopting a secondary structure upon desiccation, interacting with proteins, oligosaccharides, and membrane lipids to prevent protein denaturation. This interaction facilitates the formation of a glasslike state in the dehydrated cytoplasm, crucial for cell survival under desiccation (Bernal-Lugo and Leopold, 1995; Gilles et al., 2007; Shih et al., 2008; Hand et al., 2011). Genes encoding LEA proteins are regulated by ABA-responsive elements, highlighting ABA's role in the desiccation response (Shen et al., 2004). Additionally, as seed maturation concludes, aleurone cells undergo a specialized developmental program that imparts desiccation tolerance, enabling their survival during the drying process. This protective capacity is common among many plants and is vital for the survival of barley grains during desiccation (Leprince and Buitink, 2010). Desiccated grains enter a quiescent state known as dormancy (Bewley, 1997).

1.2.3 Seed maternal tissues

The outermost layer of the caryopsis includes the pericarp and the seed coats, which consists of testa and nucellar epidermis, both with an outer cuticle (Simpson, 1990; Kent, 1994; Bewley et al., 2006). Endosperm and embryo are surrounded by the nucellar epidermis with a cuticle on the outer side in most cereals. Subsequent layer of the testa

consists of one or two cellular layers, with exception of some sorghum varieties where the testa is absent. The pericarp, a multilayered tissue of the maternal origin, develops from the ovary wall post-fertilization. Comprising an outer epidermis, parenchymatic cells, chlorenchyma, and inner epidermis, it plays a crucial role during the early stages of grain development by serving as the primary nutrient sink (Cochrane and Duffus, 1979, 1983; Freeman and Palmer, 1984; Gubatz and Shewry, 2010; Shewry and Ullrich, 2016). Nutrients are supplied to the pericarp through the main adaxial vascular bundle, which differentiates acropetally and contains predominantly phloem in its distal part, and through lateral bundles extending into the styles (Lingle and Chevalier, 1985). From anthesis to endosperm cellularization, the pericarp maintains stable conditions, transports assimilates, protects the embryo and endosperm, and serves as the first source of starch during early grain development (Weschke et al., 2003; Radchuk et al., 2009; Sreenivasulu et al., 2010). However, during this phase, the transfer of nutrients from the pericarp to the filial grain parts is limited due to the non-functional nucellar projection, leading to nutrient accumulation around the minor vascular bundles and near the upper part of the nucellar projection (Weschke et al., 2000; Radchuk et al., 2009). As grain development progresses, stored nutrients are remobilized, and cellular disintegration occurs. This process is marked in barley by the expression of *VACUOLAR PROCESSING ENZYME 4 (VPE4)*, which is associated with PCD in the pericarp parenchyma cells (Radchuk et al., 2009, 2011). Despite its significant role, the contribution of the pericarp's assimilates to the developing endosperm is relatively minor, and comprehensive analyses of different sources' contributions to the endosperm are still lacking (Shewry and Ullrich, 2016).

Beyond its nutritive function, the pericarp also serves as a protective hull for the grain. The role of the pericarp's chlorenchyma in influencing the energy state of the developing endosperm and potentially regulating the grain's ABA content has been speculated but remains under-explored (Rolletschek et al., 2004). Additionally, seed maternal tissues play a major role in the grain dormancy (Rodríguez et al., 2015). The external structures of cereal grains, including testa and pericarp, accumulate various phenolic compounds, such as phenolic acids, coumarins, flavonoids and tannins, some of which were proposed to contribute to coat-based dormancy in wheat (Rathjen et al., 2009).

In conclusion, the pericarp is essential in the early development of grains, acting as a nutrient sink, contributing to the grain's protective structure, and possibly influencing regulatory mechanisms that affect seed yield.

1.3 Organization and structure of barley genome

Barley has a relatively large diploid genome, estimated to be around 4.88 billion base pairs, divided to seven chromosomes (Doležel et al., 2018). During interphase, the genome assumes a unique spatial arrangement known as the Rabl conformation (Nowicka et al., 2023). In this configuration, chromosomal regions display a non-random distribution within the nucleus, with centromeres clustered at one pole and telomeres at the opposite pole (Rabl, 1885). This spatial organization was long thought to be uniform for the whole organism, however recent studies suggest tissue-specific variation in chromosomal organization (Idziak et al., 2015; Němečková et al., 2020; Shan et al., 2021). In barley, Rabl organization tends to diminish with the increased level of endoreduplication observed during seed development (Nowicka et al., 2023). Recent advancements in sequencing technology and genomics allowed assembling barley reference genome and opened the way toward further explorations. These developments have shed light on the arrangement of genes, regulatory elements, and structural variations (Mascher et al., 2017, 2021). The genomic architecture of barley is comprising a combination of protein-coding genes, non-coding regions, and repetitive sequences (Figure 4).

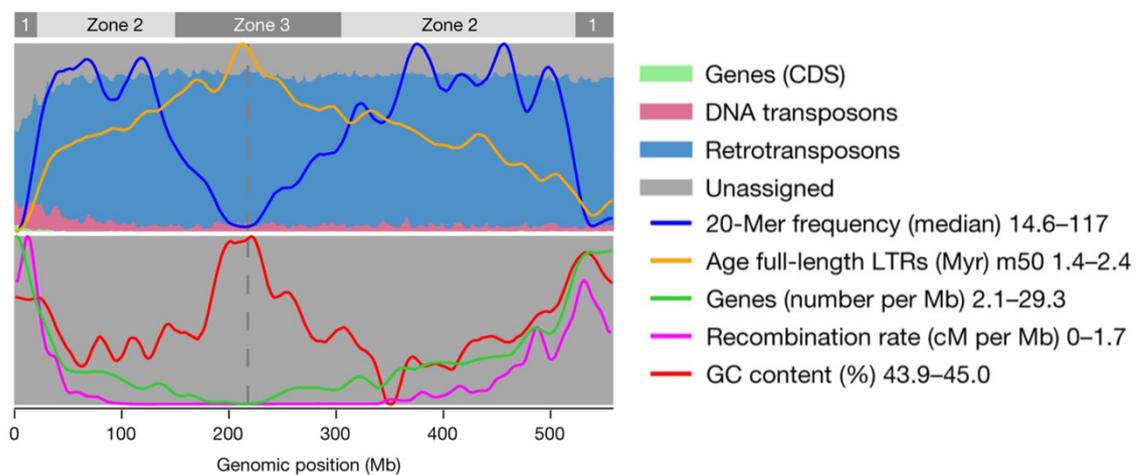


Figure 4: Characteristics of genomic features in barley on the example of chromosome 1. Chromosomes were partitioned into three zones using median 20-mer frequencies. Zone 1 is enriched for low-copy regions, high gene content and meiotic recombination events. Zone 2 had the highest 20-mer frequencies and intermediate gene density. Zone 3 is dominated by older mobile elements. CDS – predicted coding sequences; cM – centi Morgans. Vertical dashed line shows approximate position of centromeric region. According to (Mascher et al., 2017).

1.3.1 Genes

In the third version of barley genome annotation (Morex V3), a total of 81,687 genes comprising 83,990 transcripts were identified. Among these, 35,827 have been classified as high-confidence genes and the rest as low-confidence genes. Barley's genome encompasses a diverse array of protein-coding genes that regulate a wide range of biological processes. A notable difference in gene density exists between the distal and proximal regions of chromosomes, with higher gene concentration observed in the distal parts. This distribution correlates with the Rab1 intranuclear chromatin organization (Mascher et al., 2017). Gene family analyses have uncovered lineage-specific duplications of genes that are crucial for nutrient transport and carbohydrate mobilization in developing grains (Mascher et al., 2021).

1.3.2 Cis-regulatory elements

Cis-regulatory elements (CREs) are pivotal in the regulation of gene expression. These specific DNA sequences, typically 5–15 bp in length, can be bound by TFs and are located near target genes, significantly influencing their expression. Often clustered together in *cis*-regulatory modules (CRMs), CREs are crucial for fine-tuning gene expression in response to diverse environmental cues, developmental stages, and physiological conditions (Yocca and Edger, 2022; Marand et al., 2023).

Variety of CREs is vast, however, they can be categorized into core promoters, enhancers, and silencers based on their functional role. Core promoters, situated upstream of a gene's coding sequence and around the gene transcription start site (TSS), contain binding sites for RNA polymerase II and various TFs. They play a role in assembling components of the transcription preinitiation complex (Hampsey, 1998; Haberle and Stark, 2018). Core promoters are further classified as either narrow or broad, or based on distinct combinations of motifs such as the TATA-box, Y-patch and CpG motifs (Yamamoto et al., 2007; Morton et al., 2014). In the developing embryo of barley, narrow domains typically associate with TATA-box-like motifs, while broader domains are linked with non-TATA promoters (Pavlu et al., 2024). Enhancers and silencers, in contrast, do not possess the capacity for autonomous transcription initiation (Marand et al., 2023). Enhancers can be located in intergenic regions, or within the target genes, increase gene expression (Li et al., 2019; Peng et al., 2019; Concia et al., 2020). Both enhancers and silencers function by interacting with TFs and modifying chromatin

structure, thus affecting gene accessibility to transcriptional machinery. Silencers may recruit for example the Polycomb Repressive Complex 2 (PRC2) for the deposition of histone H3 lysine 27 tri-methylation (H3K27me3) (Marand et al., 2023). However, the specifics regarding which TFs act as regulators, and their precise timing and locations, remain largely unexplored, particularly in crop species (Liew et al., 2020).

1.3.3 Repetitive sequences

Repetitive sequences are segments of DNA present in multiple copies throughout the genome, and may account for up to 90% of eukaryotic genomes (Mehrotra and Goyal, 2014). The repetitive DNA elements fall into two main categories: tandemly repeated sequences and dispersed repetitive DNA elements. Tandem repeats consist of (nearly) identical sequences placed consecutively, ranging from two to thousands of copies. Dispersed repetitive elements, on the other hand, are scattered throughout the genome (Heslop-Harrison, 2000).

Tandem repeats are further classified based on the size of the monomer unit. These include microsatellites with 2-5 bp repeats, minisatellites with 6-100 bp repeats, and satellites with monomer lengths of 150-400 bp (Mehrotra and Goyal, 2014). These repeats are often enriched in specific chromosomal regions, including pericentromeric, subtelomeric, telomeric, and intercalary regions. The DNA elements that are organized in tandem arrays encompass various satellite DNAs, centromeric and telomeric repeats, and ribosomal DNA (rDNA), all of which are involved in maintenance of chromosomal integrity and production of protein non-coding RNAs (Meagher and Vassiliadis, 2005; Mehrotra and Goyal, 2014). Tandem repeats in centromeric and telomeric heterochromatin were found in almost all animal and plant genomes (Melters et al., 2013). These repeats are often genus-specific, like *HvRT* in barley subtelomeric regions, or *pAL* centromeric repeat of Arabidopsis. Centromeric repetitive sequences are integral to chromosome movement and pairing, centromeric condensation, chromosome recombination, interactions with chromatin proteins, and histone binding (Mehrotra and Goyal, 2014).

Dispersed DNA sequences, in contrast, include mobile elements such as DNA transposons, retrotransposons, and *Helitrons* along with their truncated copies or leftover sequences (Meagher and Vassiliadis, 2005; Schnable et al., 2009).

1.3.4 Transposable elements

Transposable elements (TEs), also known as "jumping genes", are mobile DNA sequences capable of moving or duplicating within a genome. TEs are key drivers of genome evolution and have a significant impact on its function (Feschotte et al., 2002). In barley, a diverse array of TEs constitutes about 80% of its genome (Wicker et al., 2017). TEs are broadly classified into two main categories: Class I retrotransposons, and Class II DNA transposons, each further subdivided into various orders and superfamilies (Wicker et al., 2007). While some TEs can disrupt gene function when inserted into coding regions, others contribute to genetic diversity and adaptation (Feschotte et al., 2002).

Retrotransposons represent class I elements that replicate via a "copy and paste" mechanism, involving transcription into RNA, reverse transcription into DNA, and subsequent reintegration into the genome. Retrotransposons are typically more abundant than DNA transposons. They are categorized into long terminal repeat (LTR) retrotransposons, which resemble retroviral elements with LTRs flanking internal coding regions, and non-LTR retrotransposons, which lack LTRs and are further divided into LINE and SINE families. In plants with large genomes, LTR retrotransposons belonging either to the *Gypsy* or *Copia* superfamilies, contribute at least 50% of the TE fraction (Vicent et al., 1999; Middleton et al., 2013; Schnable et al., 2009). Some of these retroelements are found at higher densities in specific genomic regions. For instance, a *Ty3/Gypsy*-like retroelement called *Cereba* (*Centromeric Retroelement of Barley*) is predominantly located in the centromeres of barley (Hudakova et al., 2001).

DNA transposons represent class II elements and move through a "cut and paste" mechanism. They are excised from one genomic location and inserted into another. A special type of DNA transposons, typically found near genes, are miniature inverted-repeat transposable elements (MITEs) (Bureau and Wessler, 1994a, 1994b). They are extremely diverse and contribute less to the total genomic DNA. DNA transposons are classified into various superfamilies based on their transposase enzyme domains. Notable superfamilies in barley include *CACTA*, characterized by terminal inverted repeats (TIRs) and transposases-encoding genes, and the *En/Spm* superfamily, known for elements that can transpose and influence gene expression (Wicker et al., 2017).

1.4 Epigenetic regulation

Nuclear DNA occurs in the form of chromatin, i.e. in complexes with DNA-associated proteins. The three-dimensional organization of chromatin, comprising DNA and its associated proteins, is pivotal in the regulation of gene expression. Histone-dependent DNA packaging into nucleosomes represents a fundamental level of compaction and protection for genetic material. Chromatin modifications, including cytosine (DNA) methylation and histone modifications, play a significant role in shaping the chromatin landscape and influencing gene activity. These modifications can determine whether a genomic region is open and accessible for TFs and the transcriptional machinery to bind and initiate gene expression, or closed and compacted, leading to effective gene silencing (Zhang et al., 2018). The chromatin modifications together with protein non-coding RNAs are the basis of epigenetic regulation, i.e. heritable information that is not stored directly in the nucleotide sequence.

1.4.1 DNA methylation

Conserved epigenetic modification at the 5' position of cytosine (5mC), commonly referred to as DNA methylation, plays a critical role in regulating gene expression, maintaining genome integrity, and gene imprinting (Gallego-Bartolomé, 2020). In plants, DNA methylation occurs in symmetrical CG and CHG, as well as asymmetrical CHH sequence contexts (H – A, C, or T). The level of DNA methylation is a result of the dynamic interplay of its establishment, maintenance, and active removal (Zhang et al., 2018). Catalyzed by several different DNA methyltransferases, *de novo* DNA methylation at all sequence contexts is facilitated by DOMAINS REARRANGED METHYLASE 2 (DRM2), in addition symmetric CG methylation is maintained by METHYLTRANSFERASE 1 (MET1), CHG methylation is preserved mainly by plant-specific CHROMOMETHYLASE 3 (CMT3) and CHH methylation is maintained by CHROMOMETHYLASE 2 (CMT2) (He et al., 2011; Zhang et al., 2018; Gallego-Bartolomé, 2020). Disruptions in the DNA methylation pathways often lead to developmental abnormalities in both plants and mammals (Robertson, 2005; Lang et al., 2017).

The RNA-directed DNA methylation (RdDM) pathway is crucial for *de novo* methylation in any cytosine sequence context in plants. It involves transcription by Pol IV, the production of double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA

POLYMERASE 2 (RDR2), which is subsequently processed into 24-nt siRNAs by DICER-LIKE PROTEIN 3 (DCL3) (Xie et al., 2004). These siRNAs are loaded onto ARGONAUTE 4 and 6 (AGO4 and 6) and interact with the Pol V-DRM2 complex, catalyzing *de novo* DNA methylation in a sequence-specific manner. Besides the canonical POL IV-RDR2-DCL3 pathway leading to DNA methylation, paralogues of these proteins can also generate siRNA and promote RdDM through the non-canonical POL II-RDR6-DCL2/4 pathway that cleaves the target transcripts (He et al., 2011; Zhang et al., 2018).

The maintenance of DNA methylation in plants varies based on the cytosine sequence context. MET1 maintains methylation at CG sites by recognizing hemimethylated CG dinucleotides after DNA replication and copying the information to the newly synthesized strand. CMT3, and to a lesser extent CMT2, catalyze CHG methylation. CMT3-associated DNA methylation often occurs in tandem with histone 3 lysine 9 di-methylation (H3K9me2), reinforcing each other through a regulatory feedback loop. CHH methylation is established by DRM2 in euchromatic chromosome arms and by CMT2 in histone H1-containing heterochromatin (Lindroth et al., 2001; Law and Jacobsen, 2010; He et al., 2011; Zhang et al., 2018).

Active DNA demethylation also requires multiple enzymes, including DNA demethylases (Zhang et al., 2018). This process begins with 5-mC DNA glycosylase initiating base excision repair (Gong et al., 2002; Gehring et al., 2006; Ortega-Galisteo et al., 2008). In Arabidopsis, four bifunctional 5-mC DNA glycosylases have been identified: *REPRESSOR OF SILENCING 1 (ROS1)* and *DEMETER-LIKE PROTEIN 2 and 3 (DML2 and 3)*, expressed in vegetative tissues, and *TRANSCRIPTIONAL ACTIVATOR DEMETER (DME)* in the central cell of the megagametophyte (Gong et al., 2002; Penterman et al., 2007; Huh et al., 2008; Ortega-Galisteo et al., 2008). These demethylases exhibit target specificity depending on chromatin characteristics. For instance, DME and ROS1 target euchromatic transposons. siRNAs produced from demethylated transposons in the vegetative cell travel to the sperm cells, reinforcing RdDM. Additionally, sperm cell-accumulated transposons can reinforce transposon silencing post-fertilization of egg cells (Slotkin et al., 2009; Ibarra et al., 2012; Martínez et al., 2016; Tang et al., 2016; Ingouff et al., 2017; Zhang et al., 2018).

1.4.2 Histone variants and modifications

Chromatin is composed of nucleosomes, each consisting of an octamer of histones H2A, H2B, H3, and H4, around which 147 bp of DNA is wrapped. This assembly is facilitated by the linker histone H1 (Luger et al., 1997; Filipescu et al., 2014). The accessibility of chromatin is regulated by various mechanisms, including the exchange of histone variants and post-translational histone modifications (Zhao et al., 2019).

Individual paralogous genes within each histone family encode related, yet functionally distinct, proteins known as histone variants (Talbert et al., 2012; Talbert and Henikoff, 2016). H2A variants include the canonical H2A, H2A.Z and H2A.X, with the latter two predominantly associated with transcription and DNA repair, respectively (Turinetto and Giachino, 2015; Giaimo et al., 2019). Additionally, plant genomes contain H2A.W, which is involved in heterochromatin organization (Yelagandula et al., 2014). An expanded set of H2B variants has been identified in Arabidopsis, including seed-specific histone H2B.S variant, which accumulates in mature embryos and may enhance nucleosome stability and DNA binding (Jiang et al., 2020). Some H3 variants, such as the replicative histones H3.1/H3.2, and the replacement histone H3.3, vary by cell cycle phase, with H3.3 functioning throughout the cell cycle during transcription (Stroud et al., 2012; Wollmann et al., 2012; Jiang and Berger, 2017). A notably divergent H3 variant, CenH3/CENP-A, is specifically incorporated within centromeric regions (Fukagawa and Earnshaw, 2014).

Histone modifications act as a “histone code”, specifying chromatin functions and transcriptional activities. These modifications, including methylation, acetylation, ubiquitination, and phosphorylation, are primarily located at the N-terminal tails of histones or less often within the nucleosome core regions.

Histones undergo methylation at lysine or arginine residues, with varying numbers of methyl groups added. Each type of histone methylation has distinct biological functions and is typically associated with specific genomic regions (Zhao et al., 2019). For example, TEs and repetitive sequences in heterochromatin regions are marked by H3K9me1 and H3K9me2, maintaining a constantly repressive state (Bernatavichute et al., 2008). Heterochromatin marked by H3K27me1, deposited by ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 and 6 (ATRX5 and ATRX6), can be reversibly decondensed through histone demethylation, leading to the release of TE silencing (Jacob et al., 2009). In contrast, the repressive H3K27me3, deposited by the PRC2 in

euchromatin regions, has a different role. In *Drosophila*, PRC2 comprises four core subunits, which have expanded in plants (Figure 5) (Zhao et al., 2019). In *Arabidopsis*, distinct sets of PRC2 subunits are combined at different developmental stages (Mozgova and Hennig, 2015). The removal of H3K27me3 is catalyzed by JmjC domain-containing histone demethylases such as RELATIVE OF EARLY FLOWERING (REF6), EARLY FLOWERING 6 (ELF6) and JUMONJI 13 (JMJ13) (Lu et al., 2011; Yan et al., 2018; Zheng et al., 2019).

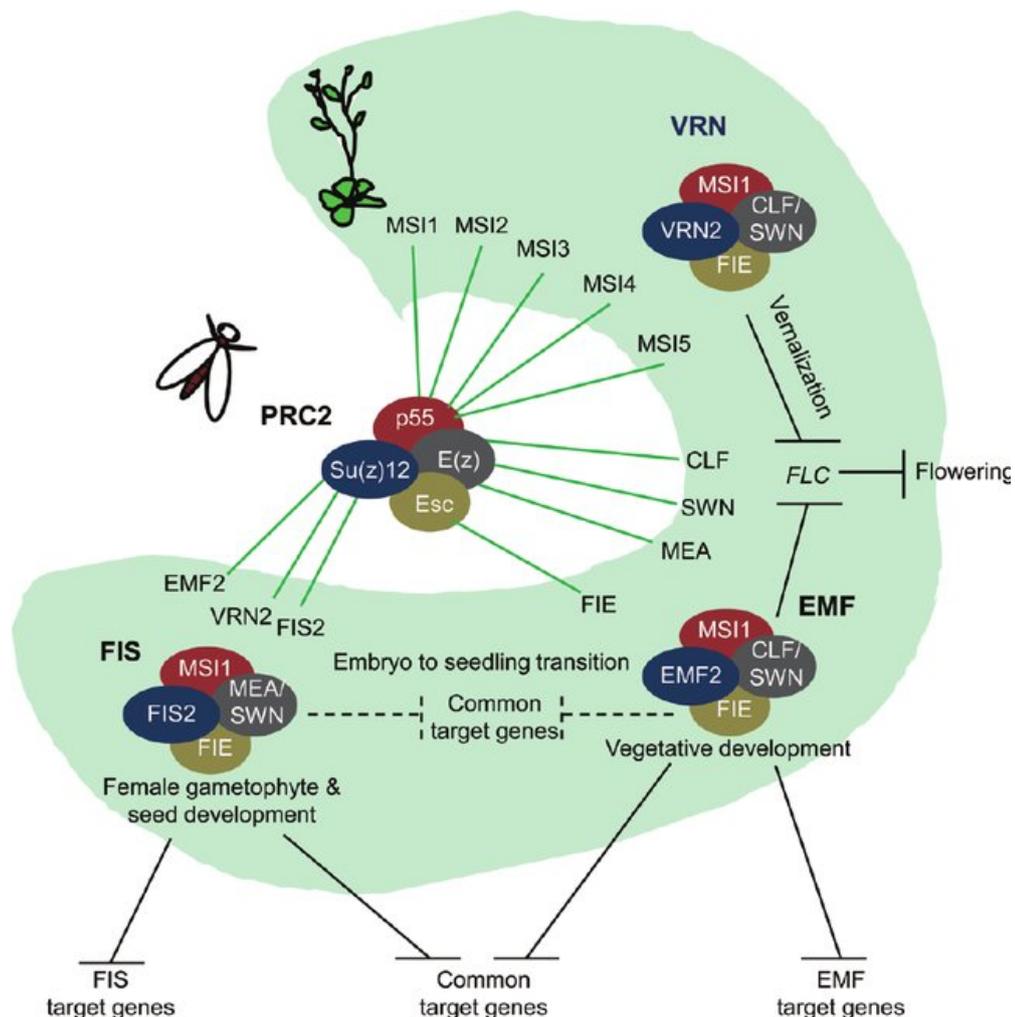


Figure 5: Overview of *Arabidopsis* PRC2 variants and their roles at different developmental stages. *Drosophila* orthologs are depicted in the middle. EMBRYONIC FLOWER (EMF) complex regulates vegetative plant development and transition to flowering, VERNALIZATION (VRN) complex active during vernalization response and FERTILIZATION INDEPENDENT SEED (FIS) complex is important for female gametophyte and seed development. Figure adopted from (Derkacheva and Hennig, 2014)

H3K4me3 is typically associated with the TSS of actively transcribed regions in euchromatin. However, H3K4me1 and H3K4me2 are present in euchromatin but do not correlate directly with active transcription (Li et al., 2008; Zhang et al., 2009). In plants, H3K36me3 is associated with the 5' region of active genes (Liu et al., 2019).

Histone acetylation generally leads to chromatin relaxation and transcriptional activation, whereas deacetylation results in transcriptional repression (Shahbazian and Grunstein, 2007). The interplay between these two states is modulated by histone acetyltransferases and histone deacetylases (Marmorstein and Zhou, 2014). Acetylation can occur at numerous lysine residues of H2A, H2B, H3 and H4 (Zhang et al., 2007).

In addition to acetylation, histones H2A and H2B can undergo monoubiquitination, each having different effects on gene transcription. H2B monoubiquitination (H2Bub1) is linked with active genes, H2A monoubiquitination (H2Aub1) is associated with transcriptional repression. Nevertheless, both modifications are found in euchromatic regions (Roudier et al., 2011; Zhao et al., 2019).

In summary, the diverse types of histone modifications collectively impact chromatin accessibility and transcriptional state. The dynamics of histone modifications arise from complex mechanisms, including the modulation of the activity and recruitment of various histone modifiers in response to developmental and environmental cues. Alongside DNA methylation and the deposition of non-histone proteins, histone modifications play crucial roles in regulating gene expression, transposon silencing, chromosome interactions, and trait inheritance (Zhang et al., 2018; Zhao et al., 2019).

1.4.3 Genomic imprinting

Genomic imprinting is an epigenetic phenomenon leading to differential expression of alleles based on their parental origin (Figure 6). As a result, imprinted genes can be either maternally expressed gene (MEG), or paternally expressed gene (PEG). This occurrence is limited to flowering plants, therian mammals, and some insects, suggesting convergent evolution of this phenomenon at least three times independently (Pires and Grossniklaus, 2014). In flowering plants, imprinting primarily occurs in the endosperm and to a lesser extent in the embryo (Gehring, 2004; Gehring et al., 2011; Hsieh et al., 2011; Waters et al., 2011). The most widely accepted theory explaining the evolution of imprinting is the theory of parental conflict, which posits a conflict of interest between maternal and paternal genomes in organisms where females can bear offspring from multiple males.

This conflict leads to evolutionary selection of MEGs and PEGs with distinct functions and developmental effects (Haig and Westoby, 1989). In plants, it is supported by parent-of-origin-specific seed phenotypes, particularly when the 2m:1p genome ratio in endosperm is disturbed in favor of one parent (Scott et al., 1998; Stoute et al., 2012; Sekine et al., 2013; Rebernig et al., 2015; Roth et al., 2019). The rescue of PEG expression in paternal excess crosses, preventing seed unviability, underscores the significant impact of imprinted genes on endosperm development (Wolff et al., 2015; Huang et al., 2017; Wang et al., 2018).

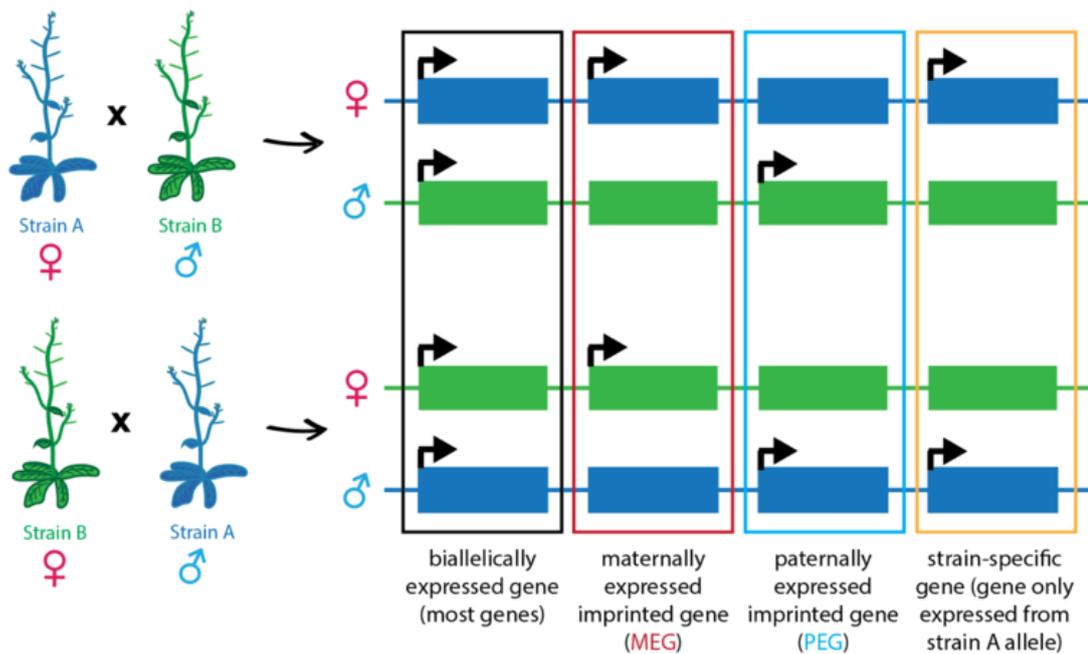


Figure 6: Gene expression patterns for biallelically-expressed genes, imprinted genes, and strain-specific genes. Most genes are expressed equally from both the maternal and paternal alleles. Maternally expressed genes (MEGs) show preferential expression from the maternal allele, whereas paternally expressed genes (PEGs) are expressed preferentially from the paternal allele. In the case of strain-biased genes, allele of one strain is more expressed than the other, independent of parent-of-origin effect (Gehring Lab Plant Imprinting Database, 2020).

The specific functions of many imprinted genes are still unknown, but several have been implicated in key developmental processes such as nutrient transfer, endosperm proliferation, and control of seed size (Gutiérrez-Marcos et al., 2004; Costa et al., 2012; Figueiredo et al., 2015; Yuan et al., 2017). Imprinting has also been observed in the PRC2 complex subunit *MEDEA* (*MEA*), an epigenetic regulator involved in imprinting establishment (Gehring et al., 2006). Dysregulation of these genes leads to altered

imprinting of many loci and seed abortion phenotypes, indicating that imprinting is essential for seed development (Grossniklaus et al., 1998; Kinoshita et al., 1999; Luo et al., 2000; Hsieh et al., 2011; Wolff et al., 2011; Hornslien et al., 2019).

In mammals, biased parental expression of imprinted genes results from epigenetic modifications established during gametogenesis, creating differential marks on parental genomes (Barlow, 1995). These imprints can be primary, such as DNA methylation established during gametogenesis and inherited stably, or secondary, arising as a result of primary imprints and further enforcing parent-specific gene expression (Barlow, 1994). Due to differing epigenetic pathways in maternal and paternal gametes, the same DNA sequence can carry distinct epigenetic marks (Batista and Köhler, 2020). In plants, the significant difference in DNA methylation between female and male genomes, particularly in the CG context, is primarily due to extensive DNA demethylation in the maternal genome in the central cell, compared to the sperm cell (Gehring et al., 2009; Hsieh et al., 2009; Zhang et al., 2014). Demethylation is performed mainly by DNA glycosylase DEMETER, active in the central cell of the female gametophyte and in the vegetative cell of pollen before fertilization. This leads to parental asymmetry in DNA methylation and maternal-specific expression of some genes, such as *MEA* and *FIS*, which are involved in H3K27me3 deposition (Choi et al., 2002; Gehring et al., 2006; Jullien et al., 2006; Gehring et al., 2009; Hsieh et al., 2011; Schoft et al., 2011). However, this phenomenon is also proposed as a mechanism of TE silencing in egg and sperm cells (Gehring et al., 2009; Hsieh et al., 2009; Calarco et al., 2012; Ibarra et al., 2012). The expression of demethylated TEs leads to siRNA production, which initiate DNA methylation via noncanonical RdDM pathway. These siRNAs, produced in the central cell and vegetative nucleus, are hypothesized to migrate to adjacent cells (egg and sperm cells) and promote DNA methylation of TEs (Slotkin et al., 2009; Calarco et al., 2012; Ibarra et al., 2012; Cuerda-Gil and Slotkin, 2016; Martínez et al., 2016; Zhang et al., 2018). Due to impact of TE demethylation on the expression of nearby genes, the primary role of DME and other DNA glycosylases might not be to generate imprinted genes, but imprinted genes could be a byproduct of DME activity on TEs (McDonald et al., 2005; Gehring et al., 2009; Hsieh et al., 2009). Different dynamics of DNA methylation in male and female gametes results in the differentially methylated regions (DMRs), where the maternal allele is often hypomethylated (Gehring et al., 2009; Zhang et al., 2011; Ibarra et al., 2012; Zhang et al., 2014; Park et al., 2016; Yuan et al., 2017). Nevertheless, only 26% of MEGs in Arabidopsis are related with hypomethylated DMRs, and many

imprinted genes are not associated with any of the epigenetic marks assessed so far. This is suggesting that there are additional mechanisms controlling imprinting in plants (Batista and Köhler, 2020).

Paternally biased expression of imprinted genes is facilitated by the Polycomb group (PcG) proteins, forming PRC1 and PRC2. The FERTILISATION INDEPENDENT (FIS) PRC2 complex, comprising FIS2, MEA, FIE and MSI1 subunits, is active during female gametogenesis and endosperm development (Gehring et al., 2006; Jullien et al., 2006; Hsieh et al., 2011; Wolff et al., 2011; Mozgova and Hennig, 2015; Moreno-Romero et al., 2016). DNA demethylation at some loci, coupled with H3K27me3 deposition, leads to the silencing of the maternal alleles of PEGs (Moreno-Romero et al., 2016). PEGs are marked by repressive H3K27me3 on the DNA-hypomethylated maternal allele, and by permissive H3K36me3 on the DNA-hypermethylated paternal allele (Dong et al., 2017). This model posits DNA demethylation as the primary imprint and H3K27me3 as a secondary imprint (Weinhofer et al., 2010; Moreno-Romero et al., 2016).

Approximately 36% of PEGs exhibit hypomethylated DMR and maternal H3K27me3 accumulation, whereas about 55% of PEGs are marked with maternal H3K27me3 unassociated with hypomethylated DMRs in Arabidopsis, indicating H3K27me3 as a primary factor in controlling imprinting of a significant part of PEGs (Batista and Köhler, 2020). Activation of maternal alleles of some PEGs in *dme* mutants suggest that DNA methylation in these mutants prevents H3K27me3 deposition on maternal alleles (Hsieh et al., 2011). However, as *MEA* and *FIS2* demethylation is necessary for their activation, the function of the entire FIS PRC2 is likely compromised in *dme* mutants (Choi et al., 2002; Gehring et al., 2006; Jullien et al., 2006; Hsieh et al., 2011). Additionally, the presence of H3K27me3 in demethylated regions does not exclude the possibility that FIS PRC2 also targets constitutively unmethylated regions, suggesting that FIS PRC2 activity does not strictly require *DME* DNA demethylation and that H3K27me3 can act as a primary imprint (Batista and Köhler, 2020). Several imprinted genes are not specifically expressed in the endosperm. This led to proposition of association between imprinting modes and genes expression in sporophytic tissues (Zhang et al., 2014). Suggesting that the epigenetic environment prior to gamete formation varies among genes based on their expression profile in sporophytic tissues. Consequently, this may influence the establishment of imprinted expression for each gene (Batista and Köhler, 2020).

1.5 Gene expression analysis

The transcriptome encompasses the entire collection of transcripts, and their quantities present in a cell at a specific developmental stage or under a particular physiological condition. The transcriptome of seeds exhibits a pronounced spatial and temporal pattern (Liew et al., 2020). Gene expression analysis is a common and powerful method widely used for understanding the transcriptional dynamics of biological systems and interpreting the functional elements of the genome (Lovén et al., 2012). Gene expression analysis can be categorized based on its scope into global and targeted.

1.5.1 Global gene expression analysis

Global gene expression analysis aims to quantify the RNA molecule population in cells and tissues. This approach is a remarkably powerful tool in molecular biology, utilized for exploring basic biology, discovering variability, studying development, and generating databases containing information about living processes. It ranks among the most popular methods in modern biology (Lovén et al., 2012). Consequently, numerous techniques have been developed over time, including expression microarrays and RNA sequencing (Lockhart and Winzeler, 2000; Heller, 2002; Wang et al., 2009).

The principle of the microarray technology is based on the complementarity between the two nucleic acid strands, enabling the detection of specific sequences through hybridization. In essence, the hybridization of a sample to a microarray is a parallel search by each molecule for a matching partner on the array (Lockhart and Winzeler, 2000). DNA microarrays, commonly used for gene expression analysis, consist of distinct oligonucleotide probes, typically representative sequences of known RNA species, attached to a solid surface like glass, silicon, or plastic substrates (Heller, 2002; Lovén et al., 2012). Various DNA microarrays, chip devices and systems have been developed and commercialized. Broadly, there are two types of microarrays: expression microarrays for detecting specific RNA sequences and genotyping microarrays for identification of specific DNA sequences (Heller, 2002).

Conversely, cDNA sequencing-based approaches, such as Sanger sequencing, directly determine the cDNA sequence. However, this method is relatively low throughput, costly, and not quantitative (Wang et al., 2009). More recently developed massively parallel sequencing technologies quantify RNA molecules through sequencing of RNA-derived cDNA populations (Lovén et al., 2012). Typically, a total population of

RNA molecules, or a fraction thereof (like polyA), is converted into a cDNA library with adaptors attached to one or both ends. This cDNA library is then sequenced in a high-throughput manner, either from one end (single-end sequencing) or both ends (pair-end sequencing). The typical read length is 30–400 bp, depending on the sequencing technology (Holt and Jones, 2008; Wang et al., 2009). Sequenced reads are, then aligned to a reference genome, or in the absence of such reference, to reference transcripts or assembled *de novo*. Finally, this process yields both the transcriptional structure and expression level for each gene (Wang et al., 2009). Furthermore, advanced sequencing technologies such as single cell RNA sequencing (scRNA-seq) facilitate the analysis of cellular heterogeneity, identification of cell types, and construction of cell lineages (Shahan et al., 2021; Shaw et al., 2021; Mo and Jiao, 2022). Nevertheless, the absence of spatial information has driven the development of spatial transcriptomics, which allows for the linking of gene expression and cellular localization (Yin et al., 2023).

1.5.2 Targeted gene expression analysis

Targeted gene expression analysis comprises a range of low-throughput techniques designed to precisely determine the presence, absence, or quantity of specific RNA or proteins. These methods can also serve to validate findings from global gene expression analyses by investigating selected genes. Techniques include quantitative PCR (qPCR), Northern blotting, RNA *in situ* hybridization, and gene reporter systems.

The qPCR allows for the quantification of input material (DNA, cDNA, or RNA). Fluorophores, either specific or nonspecific, are introduced into the PCR reaction, and fluorescence is measured in each cycle (Segundo-Val et al., 2016). The fluorescent signal, proportional to the amount of PCR product, is recorded in each cycle. The threshold cycle (C_t), at which fluorescence exceeds the background level, is measured. A lower C_t indicates a higher cDNA amount in the sample (Segundo-Val et al., 2016). Quantification can be either absolute (number of copies obtained) or relative (compared to a reference gene).

Northern blotting is a relatively simple and inexpensive technique that provides information on the size and abundance of a specific RNA in a complex sample. RNA is fractionated by size using gel electrophoresis, transferred (blotted) onto a membrane, and then analyzed by hybridization with one or more labeled RNA probes. It is often used to

demonstrate the presence or relative quantification of specific RNA in a sample (Walker, 2011).

In situ hybridization, in contrast to Northern blotting, combines molecular biology techniques with histology, offering precise localization of RNA or DNA in specific tissues or cells (Jin et al., 1997). Widely used in plants, it has been instrumental in examining the spatial expression of genes regulating floral, seed, and leaf development (Thiel et al., 2021; Zöllner et al., 2021; Hertig et al., 2023). The process involves tissue fixation, embedding in paraffin, wax, or cryostat medium, and then sectioning (Figure 7). Tissue sections are prepared for hybridization, and labeled nucleic acid probes are hybridized to mRNA in tissues. Unhybridized probes are removed through a series of wash steps, and detection can be performed via radiography or by enzyme-linked colorimetric reactions (Duck, 1994; Zöllner et al., 2021).

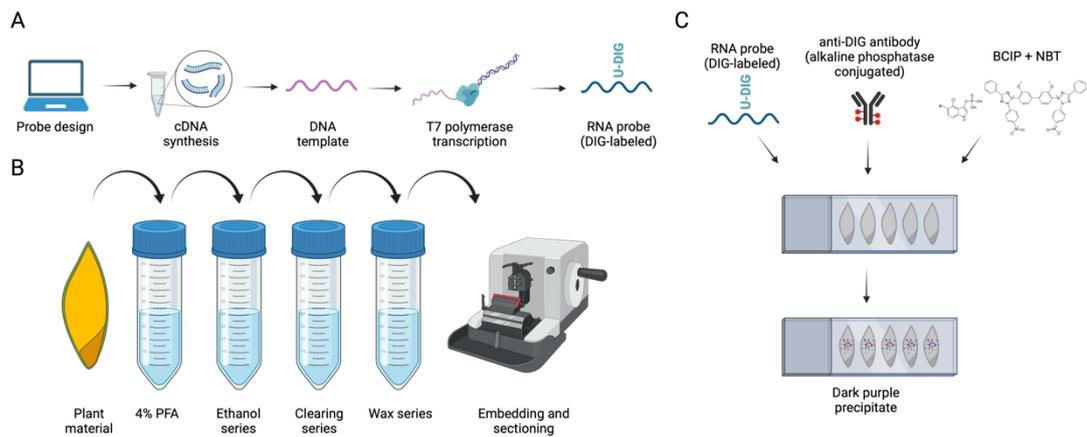


Figure 7: Schematic principle of RNA *in situ* hybridization (A) Preparation of RNA probes begins with probe design, followed by cDNA synthesis. DNA probes are synthesized from cDNA and by T7 polymerase transcribed into RNA probe. (B) Sample preparation starts with the fixation of plant material, followed by dehydration, a series of clearing and finally sample is embedded in the wax. Sections are prepared using microtome and fixed on microscopic slides. (C) The principle of RNA *in situ* hybridization is based on the hybridization of an RNA probe followed by the binding of antibodies and incubation with substrate for alkaline phosphatase. This results in dark purple precipitate in locations where target mRNA is present.

Genetic reporter systems investigate transcriptional activity within single cells. A reporter gene or cDNA is fused with a promoter in an expression vector and transferred into cells. The presence of the reporter is detected directly by measuring the amount of reporter mRNA, protein, or enzymatic activity. Reporter genes such as β -galactosidase,

β -glucuronidase (GUS), alkaline phosphatase, and green fluorescent protein (GFP) or its variants have been developed for studying promoter sequences, TFs, and protein subcellular localization. They can be analyzed directly by their enzymatic activity or spectrophotometric characteristics, or indirectly via antibody-based assays (Schenborn and Groskreutz, 1999). The fluorescence signal of the marker line is visible directly under the epifluorescence or a confocal microscope.

2 Aims of the thesis

2.1 Development of a protocol for dissection of endosperm, embryo and seed maternal tissues during seed development of barley (*Hordeum vulgare*)

The first aim of the thesis was to develop a protocol for the isolation of tissues from developing seeds. The reproducibility of tissue dissection is crucial for subsequent analyses and depends on two factors: accurate definition of the developmental stage and high purity of the dissected tissues. Manual pollination provides a precise definition of the seed development stage, although it typically results in lower success rate of pollination compared to self-pollination. The first task involved determining the day of pollination (DOP) in self-pollinated spikelets. Dissecting high-purity seed compartments poses a challenge due to their small size, diverse characteristics, compact structure, and tight adhesion of the seed tissues. Existing protocols either require specialized equipment, isolate only nuclei without cytoplasmatic content, or lacks the control of tissue purity. The second task was to manually isolate embryo, endosperm and seed maternal tissues in high purity.

2.2 Transcriptome analysis of barley seed development

The second aim of the thesis was to perform a transcriptomic analysis of developing barley seeds using high-throughput sequencing. The initial step in the bioinformatic pipeline involved trimming sequencing adaptors, removing duplicates, and aligning high-quality reads to the barley reference genome of the cultivar Morex. The subsequent step was determining transcripts per million reads (TPM), identifying significantly differentially expressed genes (DEGs) within the tissues, and classifying them according to their expression profile into clusters corresponding to early, middle, or late stages of seed development. The final step was focused on identifying regulatory motifs important for different tissues and developmental stages, thereby suggesting the involvement of specific groups of TFs.

2.3 Analysis of PRC2 complex expression, H3K27me3 distribution and identification of imprinted genes during barley seed development

The third aim of this thesis was to estimate the role of the PRC2 complex involved in the establishment of genomic imprinting and to analyze the distribution of the H3K27me3 modification across the genome in endosperm tissues. An integral part of this investigation was to compare transcriptomic and epigenomic data from different tissues to assess the direct impact of H3K27me3 modification on gene expression. Notably, H3K27me3 is crucial in the epigenetic regulation of genomic imprinting in Arabidopsis. The last task was to identify potential imprinted genes through a comparative search for evolutionary conserved imprinted genes among cereals, followed by their validation using Sanger sequencing.

3 Results

3.1 High purity tissue isolation from developing barley seeds

Establishing a protocol for high-purity tissue isolation was essential for the successful analysis of seed transcriptome. Our protocol, which involved manual tissue dissection using fine-pointed tools and a binocular microscope, focused on three main seed tissues: the embryo, endosperm and seed maternal tissues (SMTs). We selected four developmental stages ranging from 4 to 24 DAP to cover the main developmental events of the endosperm, including coenocyte formation, cellularization, differentiation, and storage compound accumulation (Figure 8A). Our goal was to develop a universal protocol applicable to seeds of any barley cultivar at any developmental stage.

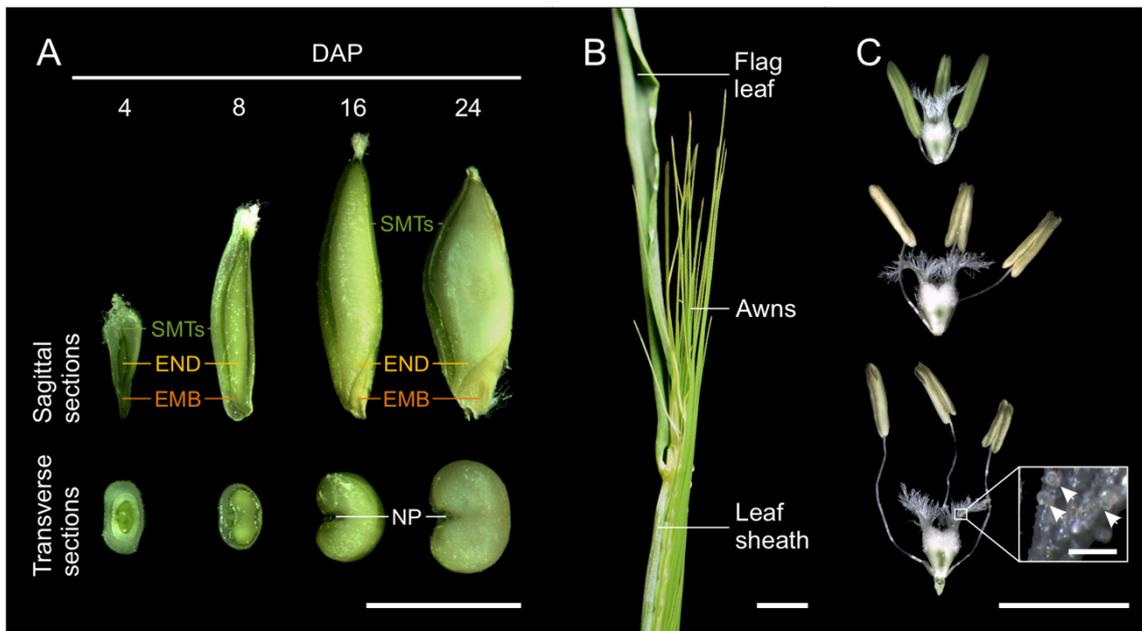


Figure 8: Developing barley seed and morphological characteristics of barley spike and anthers (A) Sagittal and transverse sections of 4, 8, 16 and 24 DAP seeds. EMB – embryo; END – endosperm; SMTs – seed maternal tissues; NP – nucellar projection; scale bar = 5 mm. (B) Morphology of barley spike close to anthesis. Scale bar = 10 mm. (C) Morphology of stigma and anthers at the Waddington stages W8.5, W9 and W10, corresponding to stages before, during and after pollination, respectively. Inset shows detail of the stigma with pollen grains (arrowheads). Scale bar = 5 mm, inset bar = 200 μ m.

Initially, we needed an accurate estimation of developmental progression. Therefore, we used self-pollinated spikes and determined the DOP based on a set of simple morphological parameters according to the Waddington scale (Waddington et al., 1983) (Figure 8B-C). The emergence of awns from the leaf sheath was the first indicator of the upcoming DOP. By examining anthers and stigmas of three to four spikelets in the central part of individual spikes, we identified spikes at DOP. We then removed the flag leaf and the top third of the hulls to facilitate the air drying of the anthers and the release of pollen. From this point, we began counting DAP to precisely define the developmental stage.

The process of tissue isolation for embryo, endosperm and SMTs was then applied. Dissection varied over time as the size, nature, and structure of the tissues changed (Figure 9). Generally, tissue dissection became simpler over time; however, increased tissue adherence during seed desiccation presented challenges. A critical part during the dissection of the tissues from 4 DAP seeds was to avoid collapsing the soft coenocyte endosperm (Figure 9A). To prevent any contamination, we removed SMTs, washed the rest of the seed in PBS buffer, and extracted the liquid endosperm using capillary needle. During dissection of 8 DAP grain, the careful removal of the nucellar projection (a part of SMTs) was crucial to avoid contamination of endosperm tissue. Conversely, it is an essential part of SMTs and should be included during isolation of this tissue. SMTs became more cohesive at later stages (16 and 24 DAP) and could be dissected into strips after moisturizing with PBS. A properly dissected embryo should have a clear, round shape of the scutellum, which is normally attached to the endosperm (Figure 9B).

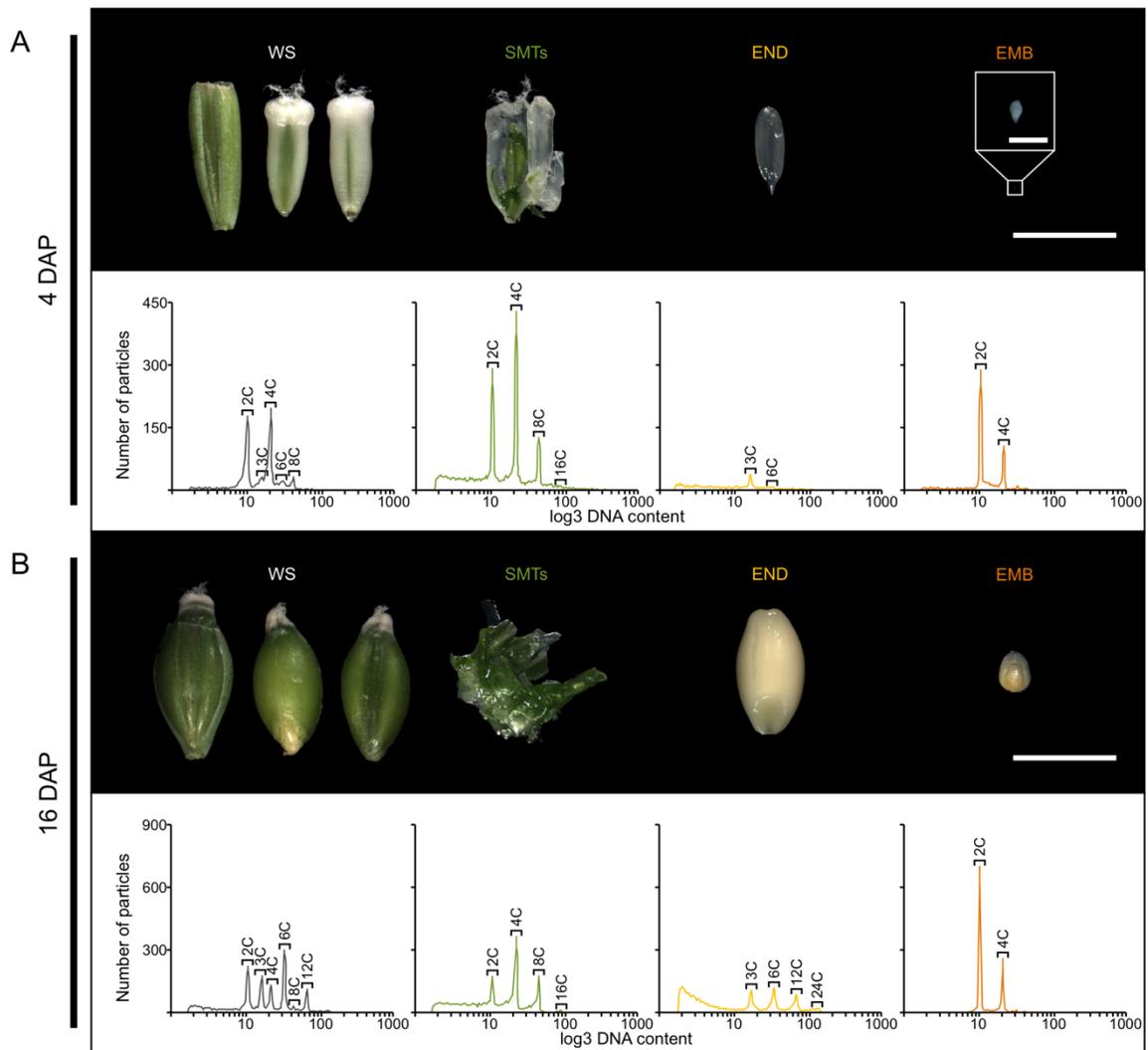


Figure 9: Morphology and flow cytometric estimation of the purity of dissected seed tissues (A) Dissected seed tissues and its representative histograms of nuclear DNA content from 4 DAP seeds. (B) Morphology and flow cytometric profiles of dissected tissues from 16 DAP seeds. WS – whole seeds (left – hulled seed, middle and right – ventral and dorsal side of peeled seed, respectively); SMTs – seed maternal tissues; END – endosperm; EMB – embryo. C-value peaks are marked for diploid embryo and/or seed maternal tissues (2C, 4C, 8C) and/or triploid endosperm (3C, 6C, 12C). The x-axis shows DNA content and the y-axis the number of measured particles. Scale bar = 5 mm.

Finally, we evaluated the purity of the isolated tissues. We used flow cytometry to estimate the nuclear DNA content, noting that the whole seeds contained a mix of diploid (2C, 4C, 8C and 16C) nuclei from embryo and SMTs, and triploid (3C, 6C, 12C and 24C) nuclei from the endosperm. Flow cytometric histograms of correctly dissected tissues showed C-value peaks corresponding to either diploid or triploid tissues exclusively (Figure 9).

To check if the samples are suitable for downstream analysis, we isolated RNA from the separated seed tissues. The use of TRIzol reagent resulted in insufficient RNA quality and high level of protein contamination in samples with high starch content. Therefore, we employed a commercial column-based kit for RNA isolation from problematic tissues, along with on-column *DNase I* treatment to remove residual DNA contamination, a prerequisite for RNA sequencing. The quality of isolated RNA was assessed using a bioanalyzer, and samples with an RNA integrity number (RIN) higher than 7 were considered of sufficient quality for further analysis, including RNA-sequencing. This allowed us to also verify the purity of isolated tissues at the molecular level in RNA-sequencing data. The analysis of transcript abundance for several well-known marker genes for each tissue confirmed minimal to no contamination from surrounding tissues (Figure 10).

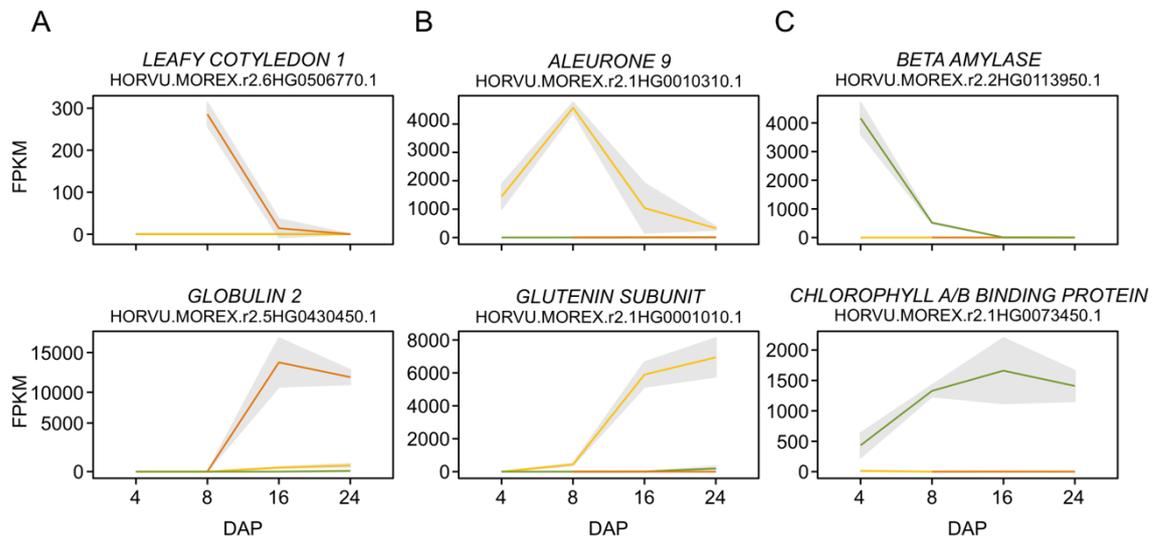


Figure 10: Example of expression from tissue marker genes. The graphs show an average expression level of tissue marker genes in embryo (orange), endosperm (yellow) and seed maternal tissues (green) at different stages of seed development. Standard deviation is indicated by the gray field. Two examples of early (top) and late (bottom) marker genes are shown for (A) embryo, (B) endosperm, (C) and seed maternal tissues. FPKM – fragments per kilobase per million reads; DAP – days after pollination.

3.2 Transcriptome landscape of developing barley seeds

To elucidate the transcriptional changes in the three primary tissues of developing barley grain, we performed RNA-sequencing on dissected embryos, endosperm and SMTs of the Morex cultivar at 4, 8, 16, 24 and 32 DAP (Figure 11A-B). Hierarchical clustering and principal component analysis (PCA) of the samples revealed distinct groupings based on developmental timepoints (PC1) and tissue types (PC2) (Figure 11C-D). Notably, samples from the early stages were more distant, indicating extensive transcriptional changes, while samples from later stages were more closely grouped, suggesting fewer transcriptional alterations. When compared with the transcriptome of other barley tissues (The International Barley Genome Sequencing Consortium, 2012), the overall distribution was defined by the seed samples. The caryopsis 5 and 15 days after anthesis grouped together with the 4 and 16 DAP endosperm, respectively, suggesting that the endosperm dominates the caryopsis transcriptome profile. To make our data easily accessible, we integrated our dataset into the Barley ePlant browser available on The Bio-Analytic Resource for Plant Biology (BAR) platform, which can be accessed at https://bar.utoronto.ca/eplant_barley/ (Figure 11E). A comparative analysis of gene expression levels across individual genes showed that median expression in the endosperm was two-fold lower than in other seed tissues (Figure 11F). Among genes with low expression (TPM 0 – 1; n = 31,571), 67.8% were expressed in the endosperm (Figure 11G). Interestingly, the endosperm also harbored the majority of genes with expression levels exceeding 1,000 TPM (n = 382) (Figure 11H). Genes highly expressed in the endosperm were significantly enriched in processes such as negative regulation of proteolysis, defense response, response to wounding, lipid transport, and cell wall macromolecule catabolic process, representing key processes critical for various stages of endosperm development.

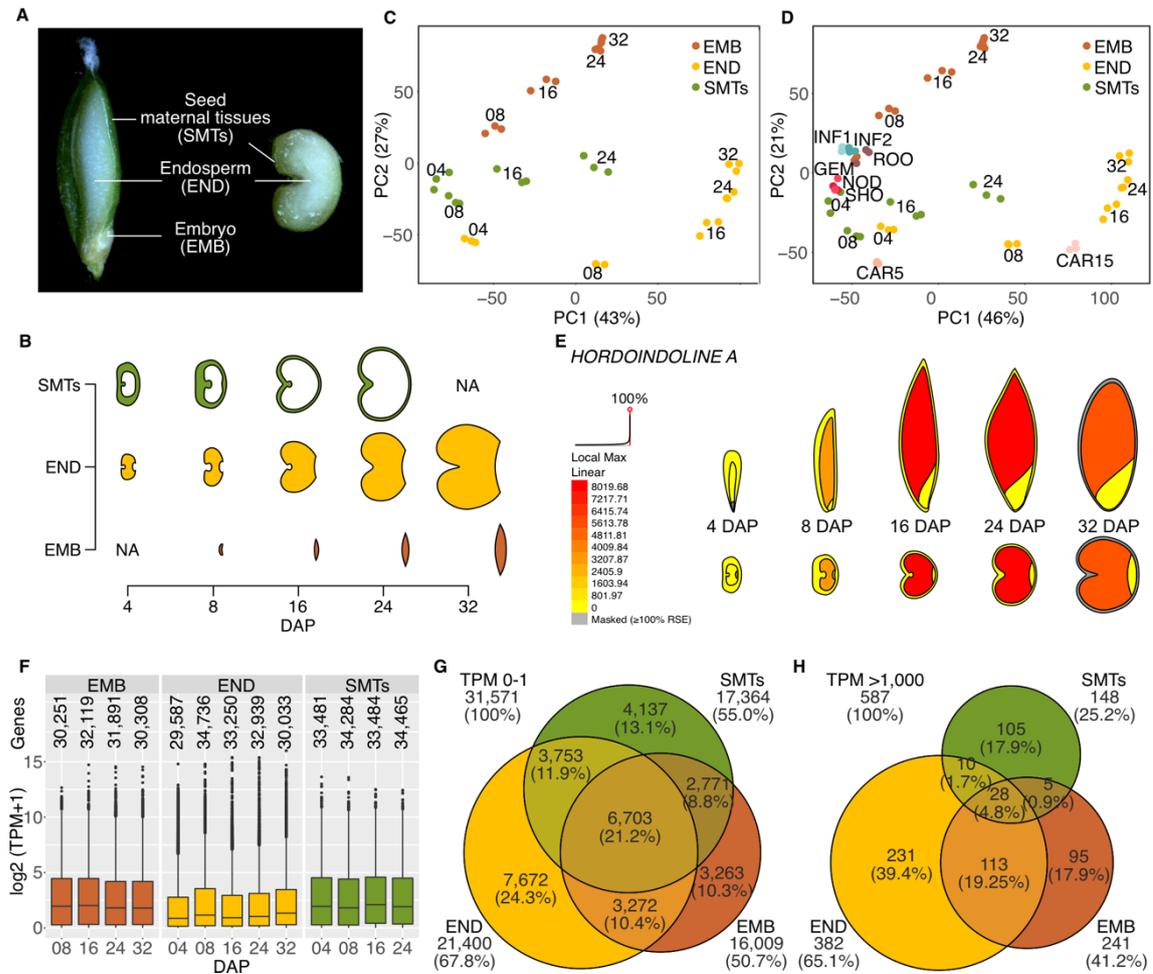


Figure 11: Transcriptomes of developing barley grain. (A-B) Schematic overview of the analyzed tissues and timepoints used for transcriptomic analysis. DAP – days after pollination; NA – not analyzed. (C) The variance of the embryo (orange), endosperm (yellow) and seed maternal tissues (green) represented by principal component analysis (PCA). The numbers within the graph indicate DAP and three close spots represent biological triplicates. The number next to PC indicates variance. (D) PCA of seed tissues used in (C) in combination with other barley tissues (The International Barley Genome Sequencing Consortium, 2012). ROO – root; GEM – germinating embryo; NOD – nodule; SHO – shoot; INF1 and INF2 – developing inflorescence of 5- and 10-mm length; CAR5 and CAR15 – caryopses 5 and 15 DAP. (E) Visualization of interactive heatmap at ePlant Barley showing TPM expression levels of *HORDOINDOLINE A* in different tissues of developing barley grain. (F) Boxplots of expression for genes with non-zero expression. (G-H) Venn diagrams showing numbers of lowly (0-1 TPM) and highly (>1,000 TPM) expressed genes, respectively.

To investigate tissue and stage-specific gene expression, we identified DEGs unique to one or shared across multiple time points within each tissue (Figure 12A). In the embryo, major transcriptional changes occurred between 8 and 24 DAP, with 8,952 DEGs observed between 8 and 16 DAP and 10,340 DEGs between 16 and 24 DAP (FDR-adjusted $P < 0.05$). The endosperm displayed significant transcriptional changes at early stage of development, i.e. from 4 to 8 DAP (15,990 DEGs). Similarly, substantial changes were noted in SMTs from 4 to 16 DAP (5,604 DEGs between 4 and 8 DAP; 8,264 DEGs between 8 and 16 DAP). The number of DEGs gradually decreased towards the later stages of development across all seed tissues. We then associated transitions between individual DAPs with over- or under-representation of gene ontology (GO) terms (Figure 12B). Early embryo development involved cell division at 8 DAP, which diminished by 16 DAP. Notable enrichment in ribosome biogenesis and cell wall synthesis was observed in the 16 DAP embryo. However, these GO terms are typical for many actively growing tissues and were also shared to a large extent with endosperm and SMTs. An enrichment of mRNA splicing via spliceosome observed after 16 DAP, increasing after 24 DAP, and was specific to the embryo, corresponding to the accumulation and subsequent translation of mRNAs during seed germination. In the endosperm, numerous tissue-specific processes related to storage compounds were enriched (lipid, starch, glucan, and glycoprotein biosynthesis and metabolism), highlighting its role as the primary nutritive tissue in grains. In SMTs, photosynthesis and cell wall development were upregulated at early stage, peaking at 8 DAP. From 4 to 16 DAP, we noticed upregulation of isoprenoid biosynthesis, which often serves as defense mechanisms against pathogens and herbivores (Kirby and Keasling, 2009). To pinpoint specific genes serving as molecular markers for individual tissues and time-points, we employed K-means clustering of DEGs (Figure 12C). DEGs from different tissues were divided into 12 – 14 co-expression clusters, with four to five clusters in each tissue exhibiting peak expression at single consecutive experimental timepoints. Tissue-specific genes with less than 5% TPM in other tissues were notably present in 4 and 8 DAP-specific clusters in both the embryo and endosperm, including several ethylene- and auxin-response factors in the embryo, and a prominent group of invertase inhibitors, pectinase inhibitors, protease inhibitors and *MEDIATOR OF RNA POLYMERASE II TRANSCRIPTION SUBUNIT 12 (MED12)*.

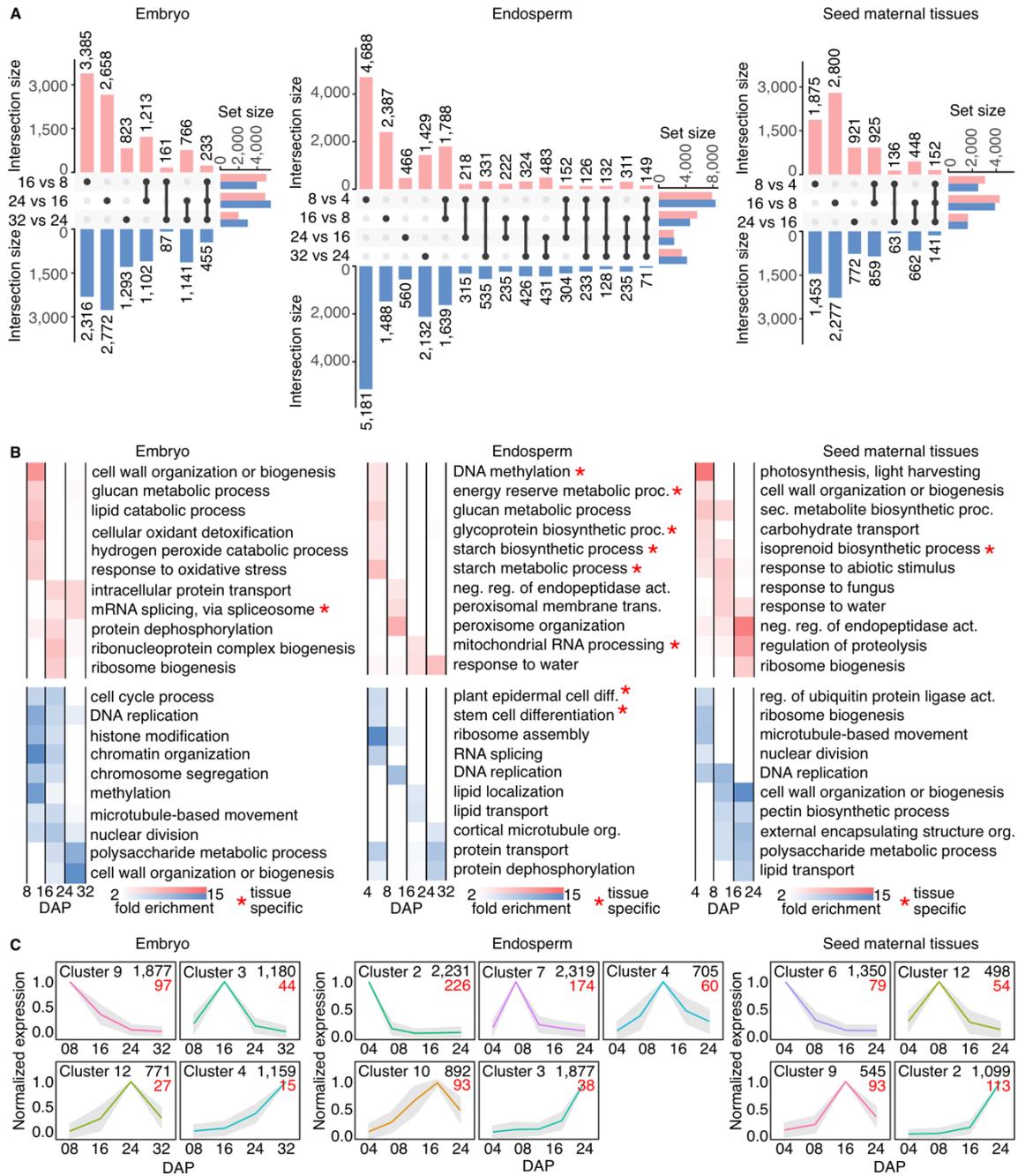


Figure 12: Tissue and timepoint specificity of gene expression during barley grain development. (A) The UpSet plots showing DEGs up-regulated (red) or down-regulated (blue) in embryo, endosperm and seed maternal tissues. The set size defines the total number of DEGs between two subsequent experimental timepoints. The intersection size shows the number of DEGs in stages and their combinations. (B) The top representative GO categories enriched among DEGs between developmental timepoints in embryo, endosperm, and seed maternal tissues. Color saturation corresponds to fold enrichment. * – tissue specific categories. (C) K-means co-expression clusters of DEGs, peaking at single consecutive developmental timepoints in embryo, endosperm and seed maternal tissues. The black numbers indicate gene count, the red numbers indicate subset of tissue-specific marker genes. DAP – days after pollination.

To identify correlations in the expression patterns among genes across our data, we conducted a weighted gene co-expression network analysis (WGCNA) on the 5,000 most divergent genes. The identified modules were organized temporally based on their profile, characteristic for the early, middle, or late stages of development (Figure 13A). Notably, the majority of genes accumulated in the early and late modules in individual tissues, suggesting that early and late seed developmental phases exhibit more dynamic transcriptional reprogramming. This approach was instrumental in uncovering TFs and regulatory motifs pivotal in barley grain development. We performed promoter motif enrichment analysis on sequences from -1,500 to -1 bp relative to the TSS of genes included in WGCNA modules. These sequences, representing their putative cis-regulatory regions, were analyzed for the presence of known TF binding motifs using the Arabidopsis HOMER database. Resulting collections of significantly enriched motifs (p -val < 0.05) were clustered based on their similarity, and the proportion of major motif clusters (MCs) in each collection was investigated (Figure 13B). Two MCs (MC2 and MC9) with consensus motifs corresponding to the G-box (CACGTG) and P-box (AWAAAG), respectively, were consistently enriched across numerous WGCNA modules in all tissues. Specific DNA-binding motifs within MC9 included members of the Dof TF family, known to bind the P-box motif (Figure 13C). The MC9 was enriched, particularly at middle embryo and endosperm development, and the relevance of these predicted TF families was corroborated by our transcriptomic data (Figure 13D). This includes the expression of the barley Dof TF *PROLAMIN BINDING FACTOR (PBF)*, known to regulate seed storage protein synthesis in barley and maize (Mena et al., 1998; Yanagisawa and Sheen, 1998). PBF activates the expression of *HORDEIN* genes in the endosperm by binding to the P-box motif in their promoter regions (Mena et al., 1998). We extended the knowledge about spatial and temporal expression pattern of this gene by *in situ* hybridization in barley endosperm (Figure 13E). The *PBF* transcript was first observed around the embryonic pole at 8 DAP and expanded to the endosperm periphery at the ventral side towards 16 DAP. Further investigation of the most expressed genes in the END_M4 module, where the P-box motif was enriched, revealed many enzymes involved in oligo- and polysaccharide synthesis (such as sucrose synthase, alpha-glucan-branching enzymes, and starch synthase), along with major endosperm proteins represented by low molecular weight glutenin subunit.

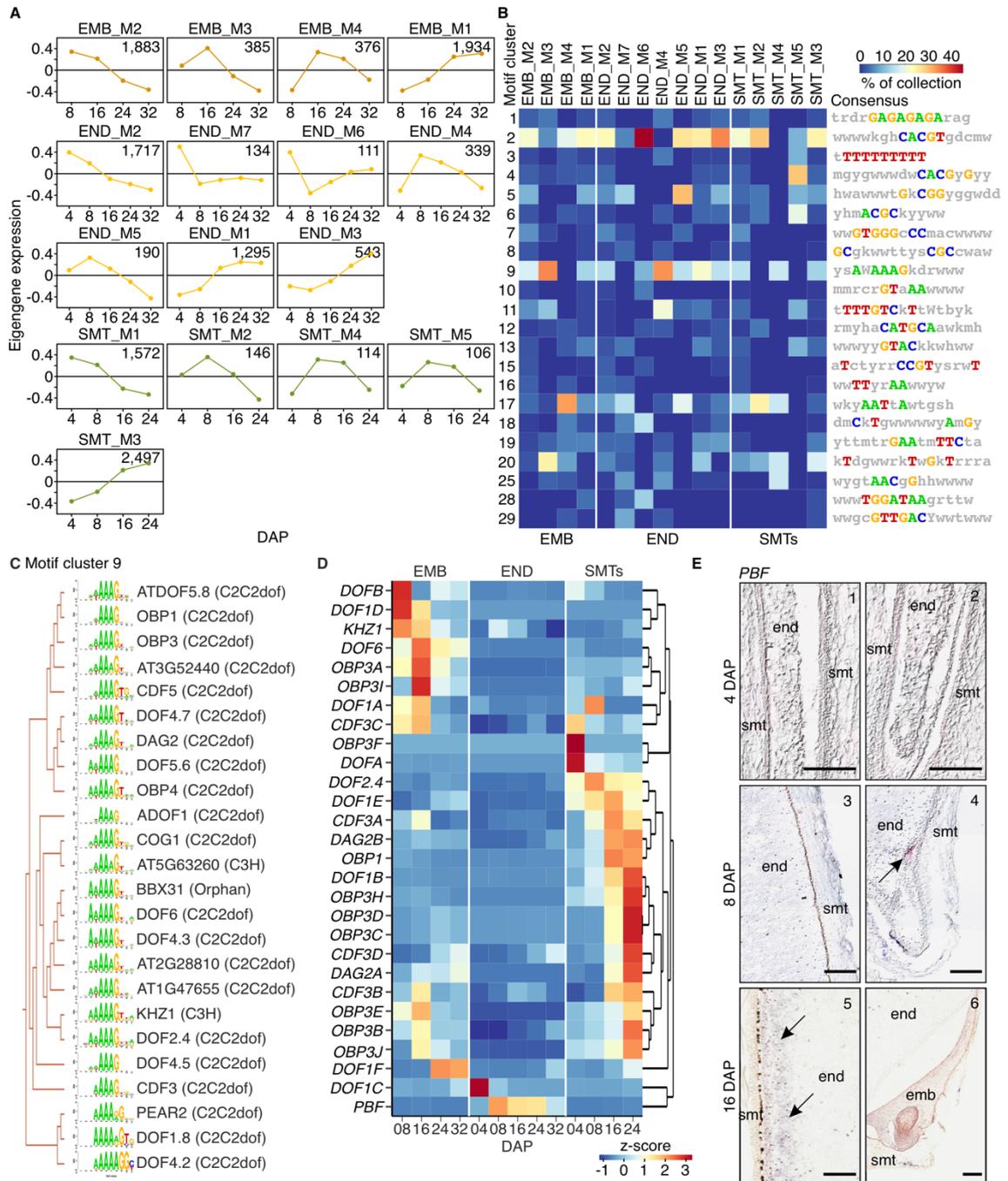


Figure 13: Gene co-expression network analysis and promoter motif enrichment. (A) Selected modules of weighted gene co-expression network analysis (WGCNA). The graphs show eigengene expression in each module. The numbers in the upper right corner indicate gene count in individual clusters. DAP – days after pollination. (B) Clusters of transcription factor binding motifs observed in individual WGCNA modules. The consensual motif in each motif cluster (MC, rows) is shown on the right side and its representation (%) in WGCNA module is indicated by the color scale. Heatmap depicts number of motifs from individual clusters identified in each WGCNA module containing at least 10 motifs. (C) Detailed clustering of TF binding sequence motifs in MC9 as to their similarity. The motifs of Arabidopsis TFs and their families are shown.

(D) Heatmap of hierarchically clustered expression for barley orthologs of Arabidopsis TFs from (C) in EMB, END and SMTs. (E) RNA *in situ* hybridization of *PBF* with antisense probe at 4, 8 and 16 DAP grains. emb – embryo; end – endosperm; smt – seed maternal tissues. Arrows indicate signal deposition in the endosperm.

To gain insights into endosperm differentiation, we conducted a comparative analysis using 12 markers for individual endosperm domains described in maize and rice (Skriver et al., 1992; Kalla et al., 1994; Opsahl-Ferstad et al., 1997; Hueros et al., 1999; Gómez et al., 2002; Magnard et al., 2003; Bate et al., 2004; Wisniewski and Rogowsky, 2004; Balandín et al., 2005). Utilizing reciprocal blast, we identified 29 barley homologs of these markers. Most of the markers exhibited a typical peak of expression around 4 to 8 DAP, indicating early endosperm differentiation (Figure 14A). In barley, the SE differentiation starts around 6 DAP, and the first detected expression of three copies of *STARCH BRANCHING ENZYME 1 (SBE1A-C)* was observed at 8 DAP, with *SBE1C* showing increased expression until peaking at 24 DAP. The expression of barley homologs of maize *ALEURONE 9 (AL9A and AL9B)*, rice subunit B1A of the *NUCLEAR TRANSCRIPTION FACTOR Y (NF-YB1A)* and maize *COLORED ALEURONE 1 (CIA)* peaked at 8, 16 and 24 DAP, respectively (Figure 14A-B). This suggests that the establishment of the AL in barley occurs later during endosperm development. However, RNA *in situ* hybridization of *NF-YB1A* revealed an interesting accumulation of the signal at the basal pole of the seed at 4 DAP endosperm (Figure 14C), indicating that aleurone identity might be defined in some endosperm nuclei even before cellularization, starting at the embryonic pole. The expression of ETCs markers *MYB-RELATED PROTEIN 1 (MRP1A and MRP1C)* and *EMBRYO SAC/BASAL ENDOSPERM TRANSFER LAYER/EMBRYO SURROUNDING REGION 1 (EBE1A and EBE1D)* was initiated at 4 DAP and peaked at 8 DAP. This timing suggests that transfer cell fate specification in barley likely occurs within a narrow time frame during the coenocyte stage of endosperm development. The ESR markers *INVERTASE INHIBITOR 1A (INVINH1A)*, *EMBRYO SURROUNDING REGION 6 (ESR6)*, *ANDROGENIC EMBRYO 1* paralogs *A* and *B (ANE1A and ANE1B)* and *ANDROGENIC EMBRYO 3* paralogs *A* and *B (ANE3A and ANE3B)* reached their expression peak at 4 DAP (Figure 14A). Our data confirmed the specificity of well-known endosperm markers from other cereals in barley and suggested that the specificity of the AL, ETC and ESR is defined at their founder nuclei even before endosperm cellularization.

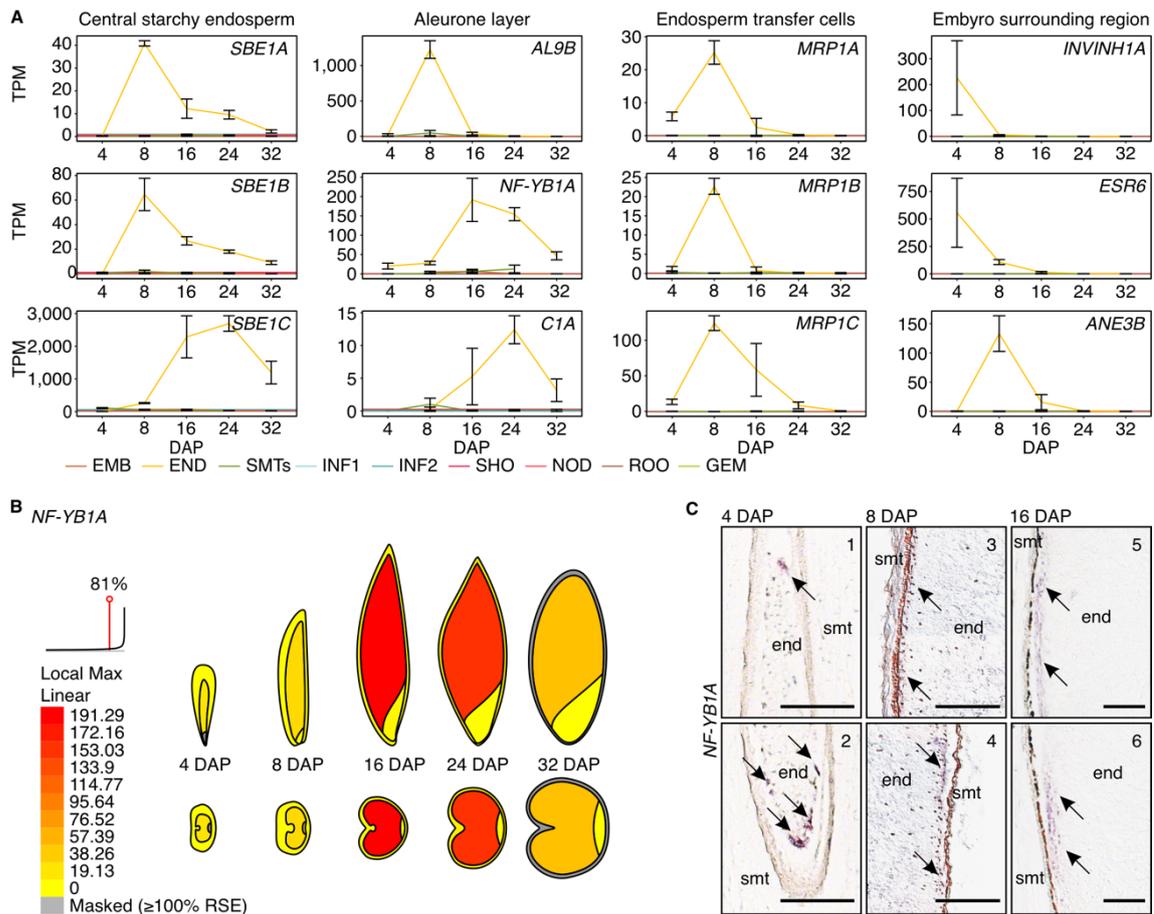


Figure 14: Expression from markers of endosperm domains in barley. (A) Expression profiles of barley orthologs of selected endosperm marker genes in other cereals grouped according to the domain of expression – central starchy endosperm, aleurone layer, endosperm transfer cells, and embryo surrounding region across different barley tissues (The International Barley Genome Sequencing Consortium, 2012). EMB – embryo; END – endosperm; SMTs – seed maternal tissues; ROO – root; GEM – germinating embryo; NOD – nodule; SHO – shoot; INF1 and INF2 – developing inflorescence 5 and 10 mm; CAR5 and CAR15 – caryopsis 5 and 15 days after pollination (DAP). Error bars indicate standard deviation. (B) Visualization of *NF-YB1A* expression (transcripts per million) in different tissues of developing barley grain at Barley ePlant. (C) RNA *in situ* hybridization of *NF-YB1A* in barley grains at 4, 8, and 16 days after pollination (DAP) using antisense probe. emb – embryo; end – endosperm; smt – seed maternal tissues. Scale bars 200 μ m.

3.3 Implications of PRC2 complex expression in H3K27me3 epigenetic landscape and identification of conserved imprinted genes in developing barley seed

In our study, we investigated the roles of histones and the PRC2, as indicated by GO term enrichment. We identified 152 out of 175 barley histone genes that were predominantly expressed during the coenocyte stage of endosperm development ($\text{TPM} \geq 1$), with their expression fading after cellularization (Figure 15). The peak of histone gene expression coincided with the period of DNA replication and nuclear division during coenocyte endosperm development. After cellularization, the expression of histone genes decreased, with only a few copies being expressed, mostly paralogs of non-canonical variants. The initial stages of embryo and SMTs development also exhibited peaks of histone expression, however at lower overall expression levels compared to the endosperm. A single *H2B* copy showed exclusive expression in 16 DAP and later embryo stages, and reciprocal BLAST revealed that it is a recently described histone H2B.S variant with seed-specific expression (Jiang et al., 2020).

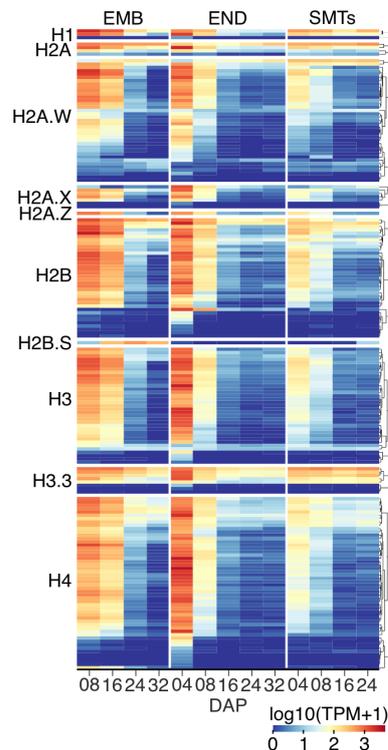


Figure 15: Heatmap of expression from 175 histone genes. The expression in different seed tissues is partitioned into groups according to core histones and their variants and hierarchically clustered within the groups. EMB – embryo; END – endosperm; SMTs – seed maternal tissues.

Regarding to PRC2, our analysis revealed that the barley genome contains single copies of *FIE* and *SDG10*, along with multiple copies of *MSI1*, *EMF2*, and *SDG1*. At least one copy of each subunit was well expressed (TPM > 10) throughout embryo and SMTs development (Figure 16A). In the endosperm, the expression pattern was more complex, with no or minimal functional PRC2 detected at 4 DAP, but expression of *EMF2A*, *FIE*, *MSI1B*, and *SDG10* observed from 8 DAP onwards. The expression of *SDG10* suggested a potential functional parallel with Arabidopsis endosperm-specific *SDG5/MEA* in barley. To investigate the potential phenotypic effects of this expression pattern, we performed H3K27me3 immunostaining (Figure 16B). Intense H3K27me3 signal was observed at the telomeric pole in embryo nuclei of all C-values and DAPs, whereas the signal was weaker in 8 DAP endosperm nuclei and almost absent at 24 DAP endosperm, indicating a global loss of H3K27me3 in older and high C-value endosperm nuclei. Considering the repressive function of H3K27me3, we anticipated that its significant reduction in barley endosperm would correlate with transcriptional activation of certain genes. A comparative analysis of our ChIP-seq data from 16 DAP endosperm and published data from seedlings (Baker et al., 2015) revealed a general reduction of H3K27me3 in the endosperm (Figure 16C). However, individual peak analysis showed more complex changes. The total number of H3K27me3 peaks was higher in the endosperm, but they were generally shorter and smaller compared to seedlings. Differential analysis of prominent peaks (fold enrichment ≥ 5) identified 17,194 regions with a significant loss ($\log_2\text{FoldChange} < 0$) and 13,845 regions with a significant gain ($\log_2\text{FoldChange} > 0$) of H3K27me3 in endosperm compared to seedlings, corresponding to 1,856 and 1,118 genes, respectively (Figure 16D-E). To assess the direct impact of H3K27me3 on the transcriptional regulation of these genes, we compared these subsets with sets of significantly up- and down-regulated genes in the endosperm. We identified 60 genes where H3K27me3 depletion correlated with increased expression, including several storage compound genes and inhibitors of sugar and protein degradation (Figure 16F-G). This suggests that the accumulation of storage compounds in the endosperm is accompanied by simultaneous inhibition of their degradation in an H3K27me3-dependent manner. Conversely, H3K27me3 enrichment was found for 238 genes, coinciding with reduced expression in the endosperm. This group predominantly consisted of *SENESCENCE-ASSOCIATED PROTEIN*s, and RNA Pol II subunit *MEDIATOR OF RNA POLYMERASE II TRANSCRIPTION SUBUNIT 12 (MED12)*.

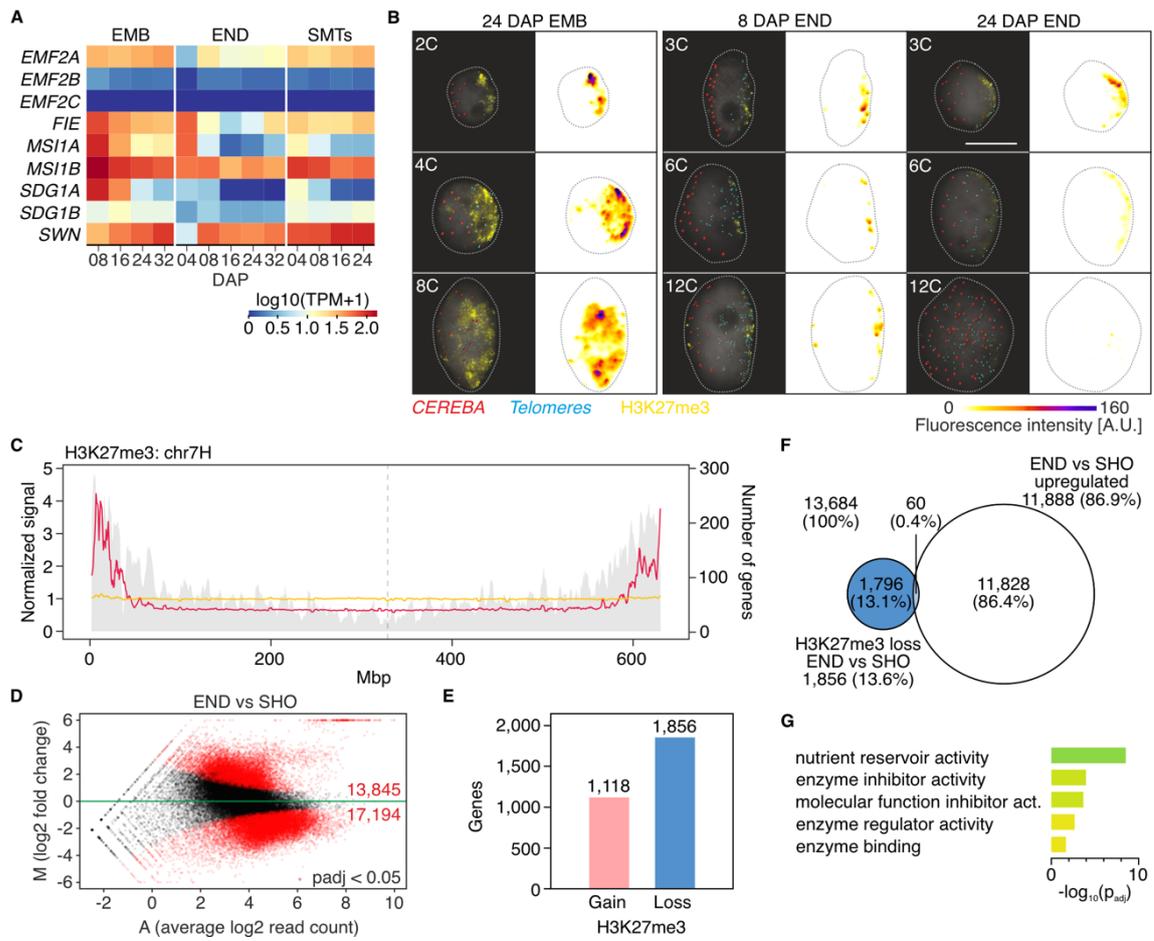


Figure 16: Dynamics of expression from Polycomb repressive complex 2 (PRC2) genes in barley grain tissues. (A) Heatmap of expression from genes coding PRC2 subunits at different days after pollination (DAP). EMB – embryo; END – endosperm; SMTs – seed maternal tissues. (B) Black background images show representative embryo and endosperm nuclei of different C-values collected at 8 and 24 DAP. DNA was stained with DAPI (grey), H3K27me3 was immunostained (yellow), and *CEREBEA* centromeric (red) and telomeric (blue) repeats were visualized by fluorescence *in situ* hybridization. White background images show H3K27me3 immunostaining fluorescence signal intensities in arbitrary units (A.U.). Scale bar – 10 μ m. (C) Normalized signal abundance of H3K27me3 in the endosperm (yellow) and 10-cm whole seedling (Baker et al., 2015) (magenta) across chromosome 7H. The gray background is gene density (secondary y-axis). (D) MAplot showing signal intensities in genomic intervals between endosperm and 10-cm seedling (SHO). Differential signal intensities (Benjamini-Hochberg FDR-adjusted p-value < 0.05) are in red. The red numbers indicate genomic intervals with significant gain or loss of H3K27me3. (E) Bar chart showing number of genes with loss of H3K27me3 in 10-cm seedlings (SHO; Baker et al., 2015) and genes up-regulated at 8, 16, 24, or 32 DAP endosperm. (F) Venn diagram showing number of genes with loss of H3K27me3 and genes up-regulated at 8, 16, 24, or 32 DAP endosperm. (G) Gene ontology (GO) term enrichment of genes with loss of H3K27me3 and significant up-regulation in the endosperm.

H3K27me3 plays an important role in epigenetic regulation of uniparental gene expression by genomic imprinting in Arabidopsis. To identify evolutionarily conserved imprinted genes in the barley endosperm, we conducted a comparative transcriptomic analysis (Figure 17A). We identified 21 barley orthologs of imprinted genes from rice, maize, and wheat, which were shared by at least two species, including 7 MEGs and 14 PEGs. These genes exhibited diverse expression patterns, falling into four main groups based on hierarchical clustering: those expressed across all seed tissues, those preferentially expressed in the embryo and SMTs, those restricted to endosperm expression, and a single gene expressed in early endosperm and 8 DAP SMTs (Figure 17B). To confirm their imprinted status, we performed reciprocal crosses between two genetically distant barley strains, elite cultivar Morex (MOR) and wild barley accession HOR 12560 (HS), and analyzed allelic expression of candidate genes in 8 DAP endosperm (Figure 17C-D). Three genes were excluded due to the lack of diagnostic SNPs and imprinting was tested for remaining 18 candidate genes that contained diagnostic SNPs in this parental combination. Through this analysis, we confirmed three MEGs – *CA-RESPONSIVE PROTEIN (CARP)*, *DNA-BINDING PROTEIN (ARP1)* and *PROLINE-RICH PROTEIN (PRP3)*. *CARP* was MEG in wheat and rice and was expressed only weakly in endosperm, compared to moderate expression in embryo. Rice and maize MEGs *ARP1* and *PRP3* were expressed specifically in barley endosperm. Additionally, four genes were confirmed as PEGs – *REGULATION OF NUCLEAR PRE-MRNA DOMAIN-CONTAINING PROTEIN 1B (RPRD1B)*, *DAI-RELATED 1 (DAR1)*, *AT-RICH INTERACTION DOMAIN 1B (ARID1B)* and *UBIQUITIN-SPECIFIC PROTEASE 16 (UBP16)*. The *RPRD1B* and *ARID1B* were found as PEGs in rice and maize. *RPRD1B* might play a role in endocytosis and cytoskeletal regulation, whereas Arabidopsis *ARIDI* is a transcriptional activator expressed in pollen development (De Camilli et al., 2002). *DAR1* was previously found as PEG in wheat and maize, and it is related to Arabidopsis *DA1*, a ubiquitin-activated peptidase playing a role in the regulation of endoreduplication, determination of plant architecture and possibly maternal control of seed mass (Peng et al., 2015; Gu et al., 2022). Lastly, we confirmed member of the UBP family *UBIQUITIN-SPECIFIC PROTEASE 16 (UBP16)* as evolutionary conserved PEG in maize, rice, wheat and in our study also in barley.

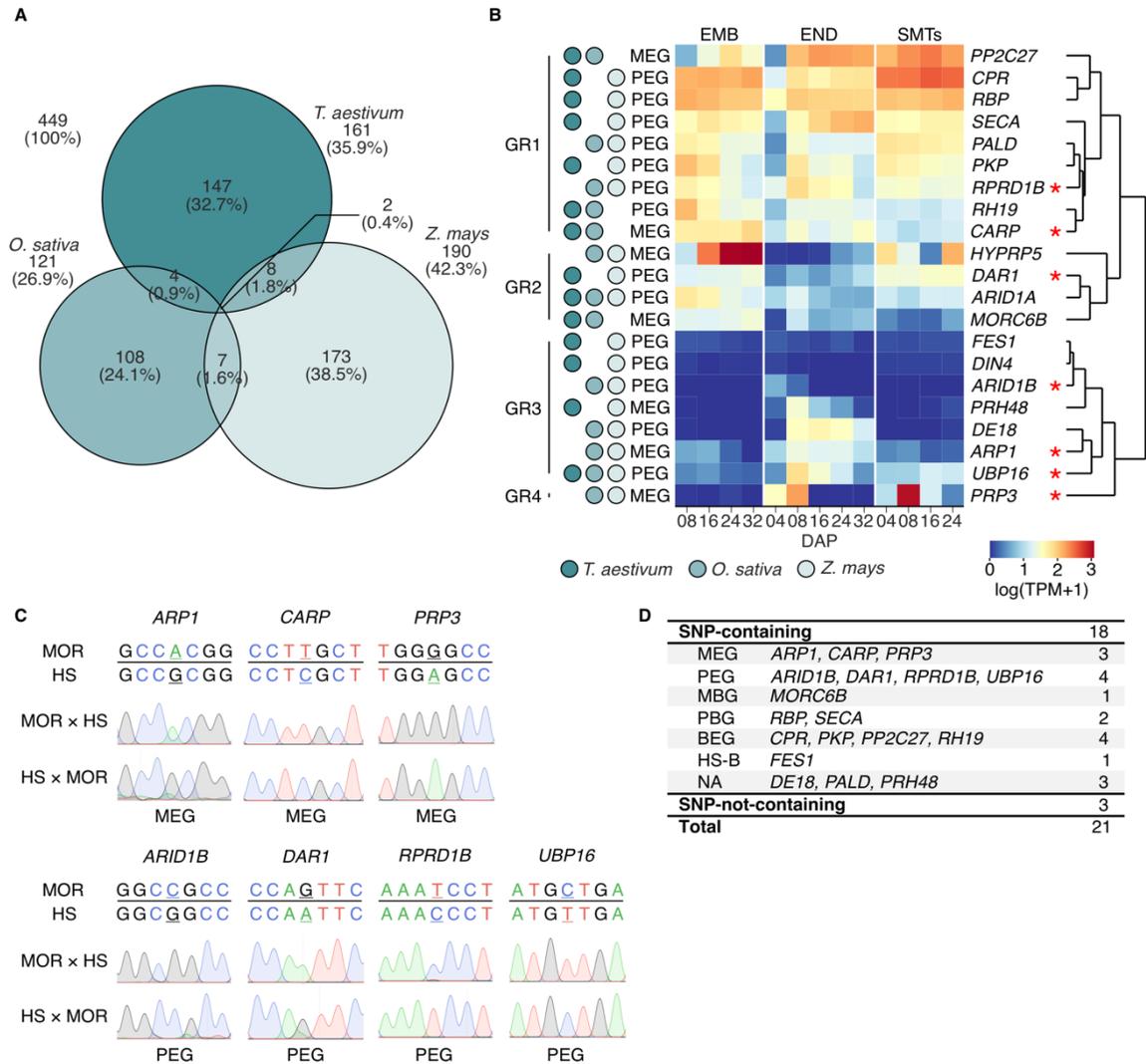


Figure 17: Identification and validation of imprinted genes in barley endosperm. (A) Venn diagram showing the number of genes imprinted in single and multiple cereal species. (B) Heatmap of expression from barley orthologs of 21 genes imprinted in more than one cereal species in different seed tissues at different days after pollination (DAP). The heatmap is partitioned into groups based on expression profiles. EMB – embryo; END – endosperm; SMTs – seed maternal tissues; MEG – maternally expressed gene; PEG – paternally expressed gene; * – imprinted in barley. (C) Validation of SNPs by Sanger sequencing for selected imprinted genes in 8 DAP endosperm tissues of F1 reciprocal hybrids obtained by crossing genetically distant strains Morex (MOR) and wild barley (HS). The maternal genotype is mentioned first in the crosses. The informative SNPs are underlined. (D) Summary of validation of imprinted genes in barley. MBG – maternally biased gene; PBG – paternally biased gene; BEG – biallelic expressed gene; HS-B – wild barley biased; NA – information not available.

3.4 Original publications

3.4.1 Isolation of High Purity Tissues from Developing Barley Seeds

(Appendix I)

3.4.2 Transcriptome landscape of developing barley seeds

(Appendix II)

3.4.1 Isolation of High Purity Tissues from Developing Barley Seeds

Kovačik M., Nowicka A., Pečinka A.

Journal of Visualized Experiments (164), e61681, 2020

DOI: 10.3791/61681

IF₂₀₂₀: 1.4

Abstract:

Understanding the mechanisms regulating the development of cereal seeds is essential for plant breeding and increasing yield. However, the analysis of cereal seeds is challenging owing to the minute size, the liquid character of some tissues, and the tight inter-tissue connections. Here, we demonstrate a detailed protocol for dissection of the embryo, endosperm, and seed maternal tissues at early, middle, and late stages of barley seed development. The protocol is based on a manual tissue dissection using fine-pointed tools and a binocular microscope, followed by ploidy analysis-based purity control. Seed maternal tissues and embryos are diploid, while the endosperm is triploid tissue. This allows the monitoring of sample purity using flow cytometry. Additional measurements revealed the high quality of RNA isolated from such samples and their usability for high-sensitivity analysis. In conclusion, this protocol describes how to practically dissect pure tissues from developing grains of cultivated barley and potentially also other cereals.

Contribution:

- Developed and optimized the protocol for tissue isolation,
- Involved in flow-cytometric purity verification,
- Performed RNA isolation,
- Analyzed transcript levels of known marker genes,
- Prepared all figures and wrote the initial draft of the publication.

3.4.2 Transcriptome landscape of developing barley seeds

Kovacik, M., Nowicka, A., Zwyrtková, J., Strejčková, B., Vardanega, I., Esteban, E., Pasha, A., Kaduchová, K., Krautsova, M., Červenková, M., Šafář, J., Provart, N., J., Simon, R., and Pecinka, A.

The Plant Cell, 2024

in press

IF₂₀₂₃: 12.085

Abstract:

Cereal grains are an important source of food and feed. To provide comprehensive spatiotemporal information about biological processes in developing seeds of cultivated barley, we performed a transcriptomic study of the embryo, endosperm, and seed maternal tissues collected from grains 4 to 32 days after pollination. Weighted gene co-expression network and motif enrichment analyses pointed out specific groups of genes and transcription factors potentially regulating barley seed tissue development. We defined a set of tissue-specific marker genes and families of transcription factors for functional studies of the pathways controlling barley grain development. Assessment of selected groups of chromatin regulators revealed that epigenetic processes are highly dynamic and likely to play a major role during barley endosperm development. Repressive modification H3K27me3 is globally reduced in endosperm tissues and at specific developmental and storage compound genes. Altogether, this atlas uncovers the complexity of the developmental regulation of gene expression in developing barley grains.

Contribution:

- Performed sample preparation and RNA extraction,
- Optimized bioinformatic pipeline for processing RNA-seq and ChIP-seq data,
- Performed all bioinformatic analyses,
- Performed RNA *in situ* hybridization,
- Performed sample preparation for chromatin immunoprecipitation,
- Prepared all figures, supplementary tables, and datasets, and wrote first draft of the publication.

3.5 Co-author publications

3.5.1 Dynamics of endoreduplication in developing barley seeds

(Appendix III)

3.5.2 Endopolyploidy variation in wild barley seeds across environmental gradients in Israel

(Appendix IV)

3.5.3 Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues

(Appendix V)

3.5.4 Core promoterome of barley embryo

(Appendix VI)

3.5.1 Dynamics of endoreduplication in developing barley seeds

Nowicka, A., Kovacik, M., Tokarz, B., Vrána, J., Zhang, Y., Weigt, D., Doležel, J., and Pecinka, A.

Journal of Experimental Botany 72(2), 268–282, 2021

DOI: 10.1093/jxb/eraa453

IF₂₀₂₁: 7.298

Abstract:

Seeds are complex biological systems comprising three genetically distinct tissues: embryo, endosperm, and maternal tissues (including seed coats and pericarp) nested inside one another. Cereal grains represent a special type of seeds, with the largest part formed by the endosperm, a specialized triploid tissue ensuring embryo protection and nourishment. We investigated dynamic changes in DNA content in three of the major seed tissues from the time of pollination up to the dry seed. We show that the cell cycle is under strict developmental control in different seed compartments. After an initial wave of active cell division, cells switch to endocycle and most endoreduplication events are observed in the endosperm and seed maternal tissues. Using different barley cultivars, we show that there is natural variation in the kinetics of this process. During the terminal stages of seed development, specific and selective loss of endoreduplicated nuclei occurs in the endosperm. This is accompanied by reduced stability of the nuclear genome, progressive loss of cell viability, and finally programmed cell death. In summary, our study shows that endopolyploidization and cell death are linked phenomena that frame barley grain development.

Contribution:

- Performed tissue isolation for estimation of nuclear DNA content,
- Was involved in estimation of nuclear DNA content,
- Performed correlation analysis.

3.5.2 Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel

Nowicka, A., Sahu, P., P., Kovacik, M., Weigt, D., Tokarz, B., Krugman, T., and Pecinka, A.

Genes (Basel) 12(5), 711, 2021

DOI: 10.3390/genes12050711

IF₂₀₂₁: 4.141

Abstract:

Wild barley is abundant, occupying large diversity of sites, ranging from the northern mesic Mediterranean meadows to the southern xeric deserts in Israel. This is also reflected in its wide phenotypic heterogeneity. We investigated the dynamics of DNA content changes in seed tissues in ten wild barley accessions that originated from an environmental gradient in Israel. The flow cytometric measurements were done from the time shortly after pollination up to the dry seeds. We show variation in mitotic cell cycle and endoreduplication dynamics in both diploid seed tissues (represented by seed maternal tissues and embryo) and in the triploid endosperm. We found that wild barley accessions collected at harsher xeric environmental conditions produce higher proportion of endoreduplicated nuclei in endosperm tissues. Also, a comparison of wild and cultivated barley strains revealed a higher endopolyploidy level in the endosperm of wild barley, that is accompanied by temporal changes in the timing of the major developmental phases. In summary, we present a new direction of research focusing on connecting spatiotemporal patterns of endoreduplication in barley seeds and possibly buffering for stress conditions.

Contribution:

- Involved in estimation of nuclear DNA content,
- Performed statistical analysis.

3.5.3 Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues

Nowicka, A., Ferková, L., Said, M., Kovacik, M., Zwyrtková, J., Baroux, C., and Pecinka, A.

Journal of Experimental Botany 74(8), 2527–2541, 2023

DOI: 10.1093/jxb/erad036

IF₂₀₂₃: 6.9

Abstract:

Rabl organization is a type of interphase chromosome arrangement with centromeres and telomeres clustering at opposite nuclear poles. Here, we analyzed nuclear morphology and chromosome organization in cycling and endoreduplicated nuclei isolated from embryo and endosperm tissues of developing barley seeds. We show that endoreduplicated nuclei have an irregular shape, less sister chromatid cohesion at 5S rDNA loci, and a reduced amount of centromeric histone CENH3. While the chromosomes of the embryo and endosperm nuclei are initially organized in Rabl configuration, the centromeres and telomeres are intermingled within the nuclear space in the endoreduplicated nuclei with an increasing endoreduplication level. Such a loss of chromosome organization suggests that Rabl configuration is introduced and further reinforced by mitotic divisions in barley cell nuclei in a tissue- and seed age-dependent manner.

Contribution:

- Performed statistical analysis.

3.5.4 Core promoterome of barley embryo

Pavlu, S., Nikumbh, S., Kovacik, M., An, T., Lenhard, B., Simkova, H. and Navratilova, P.

Computational and Structural Biotechnology Journal 23(2024) 264-277, 2024

DOI: 10.1016/j.csbj.2023.12.003

IF₂₀₂₃: 6.155

Abstract:

Precise localization and dissection of gene promoters are key to understanding transcriptional gene regulation and to successful bioengineering applications. The core RNA polymerase II initiation machinery is highly conserved among eukaryotes, leading to a general expectation of equivalent underlying mechanisms. Still, less is known about promoters in the plant kingdom. In this study, we employed cap analysis of gene expression (CAGE) at three embryonic developmental stages in barley to accurately map, annotate, and quantify transcription initiation events. Unsupervised discovery of de novo sequence clusters grouped promoters based on characteristic initiator and position-specific core-promoter motifs. This grouping was complemented by the annotation of transcription factor binding site (TFBS) motifs. Integration with genome-wide epigenomic data sets and gene ontology (GO) enrichment analysis further delineated the chromatin environments and functional roles of genes associated with distinct promoter categories. The TATA-box presence governs all features explored, supporting the general model of two separate genomic regulatory environments. We describe the extent and implications of alternative transcription initiation events, including those that are specific to developmental stages, which can affect the protein sequence or the presence of regions that regulate translation. The generated promoterome dataset provides a valuable genomic resource for enhancing the functional annotation of the barley genome. It also offers insights into the transcriptional regulation of individual genes and presents opportunities for the informed manipulation of promoter architecture, with the aim of enhancing traits of agronomic importance.

Contribution:

- Sharing RNA-sequencing data and annotation of barley transcription factors.

3.6 Published abstracts

3.6.1 Analysis of transcriptome landscape in developing barley seeds

(Appendix VII)

poster presentation

3.6.2 Transcriptome landscape of endosperm in developing barley seeds

(Appendix VIII)

poster presentation

3.6.3 Developing an atlas of gene expression during barley grain development

oral presentation

3.6.1 Analysis of transcriptome landscape in developing barley seeds

Kovačik M., Nowicka A., Pecinka A.

In: Book of abstracts of the Plant Biotechnology: Green for Good V. Poster 23, Olomouc, Czech Republic, 2019

Abstract:

Cereals are important source of calories for human and domestic animals. Cereal seeds are the raw material for food production, goods, and biofuels. This is owing to their large endosperm. Seed development is started by double-fertilization, during which two sperm nuclei migrate into embryo sac and one fuses with the egg cell nucleus, while the second fuses with the central cell nucleus, giving rise to the diploid embryo, respectively. Developing embryo and endosperm are surrounded by a diploid seed coat of maternal origin. In order to understand molecular and cellular mechanisms governing cereal seed development, we use barley, which is the model crop for temperate zone cereals. To gain basic knowledge on the processes taking place in different seed tissues and during developmental stages, we performed transcriptomic analysis. Total RNA was isolated from endosperm and embryo nuclei at 4, 8, 16 and 24 DAP (days after pollination) to cover main events of endosperm development and RNA-sequenced. At the meeting we will discuss the first results of our analysis.

3.6.2 Transcriptome landscape of endosperm in developing barley seeds

Kovačik M., Nowicka A., Vardanega I., Provart N. J., Simon R., Pecinka A.

In: Book of abstracts of the 26th International Conference on Sexual Plant Reproduction
P. 203. Prague, Czech Republic, 2022

Abstract:

Cereal grains are the major source of food and feed. The largest part of cereal seeds is occupied by triploid endosperm, representing a specialized tissue for embryo protection and nourishment. In order to understand molecular and cellular mechanisms governing cereal seed development, we used barley (*Hordeum vulgare*), which is a diploid temperate zone cereal crop. To provide a spatiotemporal information about the seed developmental process, we performed transcriptomic study of endosperm, embryo and seed maternal tissues at early, middle, and late stages of barley grain development. Analysis of differential gene expression and co-expression networks pointed out to the major biological processes going on in different grain tissues at different times after fertilization, such as cellularization, differentiation and storage components synthesis. Furthermore, we defined a set of tissue-specific marker genes, which can be used to follow tissue origin and stage of seed development. This opens new avenues towards functional analysis of barley seed development.

3.6.3 Developing an atlas of gene expression during barley grain development

Kovačik M., Nowicka A., Vardanega I., Zwyrkova J., Provar N. J., Simon R., Pecinka A.

In: Book of abstracts of the 13th International Barley Genetics Symposium (IBGS 13) Lecture, Riga, Latvia, 2022

Abstract:

Cereal seeds are an important source of food, feed and raw materials for biofuels and other technical products. We use barley (*Hordeum vulgare*) to study molecular and cellular mechanisms governing seed development. To provide comprehensive spatiotemporal information about barley grain developmental processes, we performed an RNA-seq-based transcriptomic study of different seed tissues (embryo, endosperm, seed maternal tissues) from 4 until 32 days after pollination. Analysis of differential gene expression and gene clustering based on their expression profiles revealed the major biological processes ongoing in different grain tissues at different stages. Gene co-expression network and motif enrichment analysis pointed out specific groups of genes and transcription factors with possible impacts on the regulation of endosperm development. We also defined a set of tissue-specific marker genes and families of transcription factors that can help understand the major pathways controlling barley seed development. Altogether, our atlas of gene expression during barley grain development will be a useful resource for both basic research scientists and also cereal breeders.

4 Discussion

4.1.1 Isolation of High Purity Tissues from Developing Barley Seeds

Recent advances in high-throughput sequencing have substantially enhanced the potential for developmental studies of individual seed tissues. A major challenge in these studies lies in the compact structure and tight adhesion of individual seed tissues, which complicates their separation and isolation (Sreenivasulu et al., 2010). Various approaches have been used to obtain high-resolution transcriptomes of individual seed tissues, including manual dissection, laser capture microdissection, single-cell sequencing, and spatial transcriptomics (Radchuk et al., 2009; Zhan et al., 2015; Yi et al., 2019; Hertig et al., 2023; Peirats-Llobet et al., 2023; Yuan et al., 2024). Laser capture microdissection provides precise delineation of the study area, minimizing tissue contamination. Single-cell sequencing, a more recent development, offers significantly higher resolution by enabling the measurement of expression in single cells, thus facilitating the detection and exclusion of tissue contamination from downstream analyses. However, these methods do not provide positional information about the transcripts. This limitation led to the development of spatial transcriptomics, able to preserve the positional context of transcriptional activity (Williams et al., 2022). Despite these advancements, manual dissection remains a prevalent technique in developmental studies, though it often lacks validation of tissue purity prior to RNA sequencing, leading to potential contamination from adjacent tissues.

A critical factor in developmental studies is the estimation of the growth phase. Although both self-pollination and manual flower emasculation and pollination require certain expertise, initial trials with both methods showed that self-pollinated spikes, which mature from the middle to the edges, produce seeds of variable developmental stages, limiting sampling to approximately 50% of the seeds. However, due to the low effectiveness of hand pollination and the change in seed phenotype caused by cutting the tops of the hull, we opted for self-pollination, which proved to be a more reliable method, yielding higher number of seeds.

The difficulty of tissue dissection varies with developmental stage and tissue type. Together with senescence associated RNA degradation it was the major limiting factor. Dissecting SMTs and endosperm was feasible but isolating the embryo before 8 DAP was

challenging due to insufficient amount of tissue for purity control and downstream analyses. Techniques such as micromanipulation, laser microdissection or fluorescence-activated nuclei sorting (FANS) could potentially improve the yield of embryonic tissue (Pirrello et al., 2018; Liew et al., 2020). Dissection after 8 DAP became considerably easier, though tissue separability decreased again during seed desiccation at later stages. Additionally, natural RNA degradation associated with aging prevented the isolation of RNA with high quality from SMTs after 24 DAP. Nevertheless, RIN while indicative of 18S and 25S rRNA abundance, may not accurately reflect RNA quality in transcriptionally inactive tissues.

The purity of extracted tissue was verified using flow cytometry and molecular level assessment in RNA-sequencing data. Flow cytometry, based on the differing ploidies of seed tissues, effectively distinguished the triploid endosperm from the diploid embryo and SMTs, significantly reducing the most common type of contamination. However, this method cannot differentiate between embryo and SMTs due to their identical ploidy level, though such contamination is less likely given their lack of adhesion. To further evaluate tissue purity at the molecular level, we analyzed the transcript abundance of known marker genes for each tissue in RNA-sequencing data, offering a high throughput evaluation. In tissues with well-annotated marker genes, qPCR could serve as a more straightforward alternative.

Despite discussed limitations, the application of our protocol represents a significant experimental advancement, enhancing the resolution and specificity of RNA sequencing data.

4.1.2 Transcriptome landscape of developing barley seeds

Spatiotemporal analysis of gene expression is a powerful tool for understanding developmental programs. Prior studies have examined the transcriptional profiles of barley grains and their components at various developmental stages using a range of technologies (Zhang et al., 2004; Druka et al., 2006; Thiel et al., 2008; Bian et al., 2019; Peirats-Llobet et al., 2023). Our goal was to conduct the most comprehensive transcriptomic profiling of developing barley grain to date. We compiled an expression atlas covering the transcriptome of the three primary seed tissues throughout the nearly complete developmental process of the barley seed, using RNA-sequencing technology and the latest genome assembly and annotation. However, the annotation of the barley

genome is not as comprehensive, with many genes lacking detailed descriptions. This limitation, combined with the scarce availability of mutants and their relatively long generation period, necessitated basing our conclusions on homologous genes found in another species.

Our differential expression analysis and subsequent gene clustering generated co-expression clusters that facilitated the identification of tissue marker genes. The presence of stage- and tissue-specific genes in the 4 and 8 DAP clusters within the embryo and endosperm could reflect the differing developmental rates at early and later stages. The coenocytic stage of endosperm development, characterized by rapid nuclear division without cell wall formation, is particularly significant in *Hordeum* species, generating up to 2,000 coenocytic nuclei within the first 3 DAP (Bennett et al., 1975; Walbot, 1994). The rapid early development in the endosperm is also indicated by the notably distant 4 and 8 DAP endosperm samples in PCA. Furthermore, specific processes such as coenocyte endosperm development and cellularization, which are relatively brief and specific to early developmental stages, contrast with prolonged processes like storage compound accumulation or desiccation. This diversity in developmental dynamics likely underpins the observed profiles of stage- and tissue-specific gene expression.

Investigating the role of TFs in seed development, we examined both the expression of TFs and the presence of their binding motifs in the promoter regions of expressed genes. These aspects often correlate and provide complementary insights. Despite the lack of a comprehensive database of known TF binding motifs specific to barley, we extrapolated some cases based on homologs and known binding motifs from *Arabidopsis*. For instance, the expression of the DOF transcription factor PBF and its association with seed storage protein genes is one such example. However, a deeper understanding of the regulatory network in barley, necessitates complex information on histone modifications and chromatin structure, particularly in the endosperm. Integrating these epigenetic and structural data would greatly enhance our understanding of the regulatory mechanisms governing seed development in barley.

The fully differentiated endosperm of cereals typically includes at least four main morphologically distinct domains (Olsen, 2004; Sabelli and Larkins, 2009). Although markers for these endosperm domains have been identified in various cereals, the number of such markers specifically identified for barley remains limited (Skriver et al., 1992; Kalla et al., 1994; Opsahl-Ferstad et al., 1997; Hueros et al., 1999; Gómez et al., 2002; Magnard et al., 2003; Bate et al., 2004; Wisniewski and Rogowsky, 2004; Balandín et al.,

2005). The most of the defined orthologs for these markers showed endosperm-specific expression, suggesting their utility in barley as well. The expression peaks at 4 and 8 DAP, coupled with RNA *in situ* hybridization results, indicated that endosperm differentiation begins before cellularization is complete. Recent research in barley supports this, showing that ETC and AL cells are determined during the coenocyte stage (Hertig et al., 2023). The observed morphology of the endosperm during cellularization, which displays a clear developmental gradient, suggests that positional signaling directs cell specification already at the coenocyte stage. This early cellular identity determination highlights the complexity of endosperm development in cereals, including barley.

4.1.3 Expression of PRC2 complex, H3K27me3 epigenetic landscape and conserved imprinted genes in seed development

The role of chromatin in seed development, extensively studied in the Arabidopsis model, has shed light on numerous epigenetic processes and molecular factors critical to this development (Le et al., 2010). However, the extent to which these findings are applicable to cereals, including barley, remains largely unknown. Our research delineated the expression patterns of two key groups of epigenetic regulators: histones and PRC2 complex. The observed expression of histones during the early stages of embryo and endosperm development coincides with periods of rapid DNA replication and cell division. Interestingly, several components of the PRC2 complex exhibited low level of expression, corresponding with a significant reduction of the H3K27me3 modification, at both cytological and molecular level. The reduction, identified comparative analysis of transcriptomic data and chromatin profiling, facilitated the identification of genes directly regulated by H3K27me3. However, in most cases, this global decrease in H3K27me3 did not correlate with significant transcriptional changes, suggesting the potential involvement of other mechanisms in the regulation of these genes.

Furthermore, we employed a basic comparative analysis to establish a foundation for future research on genomic imprinting in barley. This approach aimed to identify candidate imprinted genes conserved across other cereal species. One surprising outcome was the relatively low number of imprinted genes shared by at least two of the major cereal species – maize, rice, and wheat. This finding suggests a rapid evolution of imprinted genes in cereals. Among the conserved candidate genes we identified, nearly 40% were verified as imprinted in barley. These validated candidates highlight the

potential role of ubiquitin-based regulation in genomic imprinting within cereals. However, a comprehensive, genome-wide identification of imprinted genes in barley, coupled with a more sophisticated comparative analysis, is needed for a deeper understanding of genomic imprinting and its evolutionary conservation across different cereal species.

5 Conclusions

Cultivated barley, as a temperate crop, is gaining importance as a diploid model species for cereal research (Jayakodi et al., 2020; Rotasperti et al., 2020). In this context, we have developed a protocol for the dissection of different barley seed tissues and provided the most detailed gene expression atlas for seed development in barley to date.

The protocol which we developed, enables the isolation of high-purity barley seed tissues, validated through rigorous quality control analysis using a flow cytometer. Its effectiveness has been proven across several two- and six-row spring and winter barley cultivars and can be easily adapted for use with other members of the *Triticeae* tribe, such as wheat, oat, rye, or triticale.

The transcriptome atlas we have created, utilizing the current version of the barley genome assembly and annotation, serves as a solid base for subsequent biological investigations into the key factors involved in barley grain development. While our study primarily focused on endosperm tissues, the dataset provides equal resolution for the embryo and the often-neglected SMTs. We have focused mainly on TFs critical for developmental transitions, marker genes useful to study endosperm differentiation, and the role of epigenetic processes and molecular factors in seed development. We have investigated expression of PRC2 complex subunits and genome-wide distribution of H3K27me3 modification. Within our last aim we also identified conserved imprinted genes in barley. While much of this research has been thoroughly explored in other model systems like *Arabidopsis* and maize, similar information in barley has been largely unknown.

In conclusion, all three aims of the thesis were fulfilled. The combination of our dissection protocol and the comprehensive transcriptomic analysis constitutes a valuable resource that will promote and facilitate further research of barley grain development.

6 References

- Akar Taner, Avci Muzaffer, and Dusunceli Fazil** (2004). BARLEY post-harvest operations-post-harvest compendium.
- Angelovici, R., Galili, G., Fernie, A.R., and Fait, A.** (2010). Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci* **15**: 211–218.
- Babb, S. and Muehlbauer, G.J.** (2003). Genetic and morphological characterization of the barley unicum2 (cul2) mutant. *Theoretical and Applied Genetics* **106**: 846–857.
- Badr, A., Müller, K., Schäfer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H.H., Pozzi, C., Rohde, W., and Salamini, F.** (2000). On the Origin and Domestication History of Barley (*Hordeum vulgare*). *Mol Biol Evol* **17**: 499–510.
- Bai, F. and Settles, A.M.** (2015). Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Front Plant Sci* **5**: 1–10.
- Baker, K., Dhillon, T., Colas, I., Cook, N., Milne, I., Milne, L., Bayer, M., and Flavell, A.J.** (2015). Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J* **84**: 111–124.
- Balandín, M., Royo, J., Gómez, E., Muniz, L.M., Molina, A., and Hueros, G.** (2005). A protective role for the embryo surrounding region of the maize endosperm, as evidenced by the characterisation of ZmESR-6, a defensin gene specifically expressed in this region. *Plant Mol Biol* **58**: 269–282.
- Barlow, D.P.** (1995). Gametic imprinting in mammals. *Science* **270**: 1610–1613.
- Barlow, D.P.** (1994). Imprinting: a gamete's point of view. *Trends Genet* **10**: 194–199.
- Baroux, C. and Grossniklaus, U.** (2019). Seeds—An evolutionary innovation underlying reproductive success in flowering plants. In *Curr Top Dev Biol* (Academic Press), pp. 605–642.
- Bartels, D., Singh, M., and Salamini, F.** (1988). Onset of desiccation tolerance during development of the barley embryo. *Planta* **175**: 485–492.
- Bate, N.J., Niu, X., Wang, Y., Reimann, K.S., and Helentjaris, T.G.** (2004). An invertase inhibitor from maize localizes to the embryo surrounding region during early kernel development. *Plant Physiol* **134**: 246–254.
- Batista, R.A. and Köhler, C.** (2020). Genomic imprinting in plants-revisiting existing models. *Genes Dev* **34**: 24–36.

- Becraft, P.W.** (2001). Cell fate specification in the cereal endosperm. *Semin Cell Dev Biol* **12**: 387–394.
- Becraft, P.W. and Gutierrez-Marcos, J.** (2012). Endosperm development: Dynamic processes and cellular innovations underlying sibling altruism. *Wiley Interdiscip Rev Dev Biol* **1**: 579–593.
- Bennett, M.D., Smith, J.B., and Barclay, I.** (1975). Early seed development in the Triticeae. *Philos Trans R Soc Lond B Biol Sci* **272**: 199–227.
- Bernal-Lugo, I. and Leopold, A.C.** (1995). Seed stability during storage: Raffinose content and seed glassy state. *Seed Sci Res* **5**: 75–80.
- Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E.** (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS One* **3**.
- Bewley, J.D.** (1997). Seed Germination and Dormancy. *Plant Cell*: 1055–1066.
- Bewley, J.D., Black, M., and Halmer, P.** (2006). *The encyclopedia of seeds: science, technology and uses* (Cabi).
- Bhatty, R.S. and Rosnagel, B.G.** (1980). Lipids and fatty acid composition of Riso 1508 and normal barley. *Cereal Chem* **57**: 382–386.
- Bian, J., Deng, P., Zhan, H., Wu, X., Nishantha, M., Yan, Z., Du, X., Nie, X., and Song, W.** (2019). Transcriptional Dynamics of Grain Development in Barley (*Hordeum vulgare* L.). *Int J Mol Sci* **20**: 962.
- Bonett, T.O.** (1966). Inflorescences of maize, wheat, rye, barley, and oats: their initiation and development/721. *Bulletin* (University of Illinois (Urbana-Champaign campus). Agricultural Experiment Station); no. 721.
- Breuer, C., Ishida, T., and Sugimoto, K.** (2010). Developmental control of endocycles and cell growth in plants. *Curr Opin Plant Biol* **13**: 654–660.
- Brown, R.C., Lemmon, B.E., Nguyen, H., and Olsen, O.-A.** (1999). Development of endosperm in *Arabidopsis thaliana*. *Sex Plant Reprod* **12**: 32–42.
- Brown, R.C., Lemmon, B.E., and Olsen, A.** (1994). Endosperm development in barley: Microtubule involvement in the morphogenetic. *Plant Cell* **6**: 1241–1252.
- Bureau, T.E. and Wessler, S.R.** (1994a). Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses. *Proc Natl Acad Sci U S A* **91**: 1411–1415.

- Bureau, T.E. and Wessler, S.R.** (1994b). Stowaway: A New Family of Inverted Repeat Elements Associated with the Genes of Both Monocotyledonous and Dicotyledonous Plants. *Plant Cell* **6**: 907.
- Buttrose, M.S.** (1963). Ultrastructure of the Developing Aleurone Cells of Wheat Grain. *Aust J Biol Sci* **16**: 768–774.
- 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., and Martienssen, R.A.** (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151**: 194–205.
- De Camilli, P., Chen, H., Hyman, J., Panepucci, E., Bateman, A., and Brunger, A.T.** (2002). The ENTH domain. *FEBS Lett* **513**: 11–18.
- CGIAR: Science for humanity's greatest challenges** (2023).
- Chandler, J., Nardmann, J., and Werr, W.** (2008). Plant development revolves around axes. *Trends Plant Sci* **13**: 78–84.
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L.** (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**: 33–42.
- Chojekci, A.J.S., Bayliss, M.W., and Gale, M.D.** (1986). Cell Production and DNA Accumulation in the Wheat Endosperm, and their Association with Grain Weight. *Ann Bot* **58**: 809–817.
- Cochrane, M.P. and Duffus, C.M.** (1979). Morphology and Ultrastructure of Immature Cereal Grains in Relation to Transport. *Ann Bot* **44**: 67–72.
- Cochrane, M.P. and Duffus, C.M.** (1983). Observations on the Development of the Testa and Pericarp in Barley. In *Third International Symposium on Pre-Harvest Sprouting in Cereals*, J.E. Kruger and D.E. LaBerge, eds (CRC Press).
- Cocucci, A.E.** (2005). Morphogenetic seed types of Spermatophyta. *Plant Syst Evol* **250**: 1–6.
- Concia, L. et al.** (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol* **21**: 1–20.
- Costa, L.M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., and Gutierrez-Marcos, J.F.** (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr Biol* **22**: 160–165.
- Crop Prospects and Food Situation #1, March 2023** (2023). (FAO).

- Cuerda-Gil, D. and Slotkin, R.K.** (2016). Non-canonical RNA-directed DNA methylation. *Nat Plants* **2**: 1–8.
- Dante, R.A., Larkins, B.A., and Sabelli, P.A.** (2014). Cell cycle control and seed development. *Front Plant Sci* **5**: 1–14.
- Derkacheva, M. and Hennig, L.** (2014). Variations on a theme: Polycomb group proteins in plants. *J Exp Bot* **65**: 2769–2784.
- Doan, D.N.R., Linnestad, C., and Olsen, A.** (1996). Isolation of molecular markers from the barley endosperm coenocyte and the surrounding nucellus cell layers. *Plant Mol Biol* **31**: 877–886.
- Doležel, J., Čížková, J., Šimková, H., and Bartoš, J.** (2018). One Major Challenge of Sequencing Large Plant Genomes Is to Know How Big They Really Are. *Int J Mol Sci* **19**: 3554.
- Dong, X., Zhang, M., Chen, J., Peng, L., Zhang, N., Wang, X., and Lai, J.** (2017). Dynamic and Antagonistic Allele-Specific Epigenetic Modifications Controlling the Expression of Imprinted Genes in Maize Endosperm. *Mol Plant* **10**: 442–455.
- Druka, A. et al.** (2006). An atlas of gene expression from seed to seed through barley development. *Funct Integr Genomics* **6**: 202–211.
- Duck, N.B.** (1994). RNA in situ hybridization in plants. In *Plant Molecular Biology Manual*, S.B. Gelvin and R.A. Schilperoort, eds (Springer Netherlands: Dordrecht), pp. 335–347.
- Duffus, C.M. and Rosie, R.** (1975). Biochemical changes during embryogeny in *Hordeum distichum*. *Phytochemistry* **14**: 319–323.
- Dupuy, L., MacKenzie, J., Rudge, T., and Haseloff, J.** (2008). A system for modelling cell-cell interactions during plant morphogenesis. *Ann Bot* **101**: 1255–1265.
- Engell, K.** (1989). Embryology of barley: Time course and analysis of controlled fertilization and early embryo formation based on serial sections. *Nord J Bot* **9**: 265–280.
- Erbasol Serbes, I., Palovaara, J., and Groß-Hardt, R.** (2019). Development and function of the flowering plant female gametophyte. In *Curr Top Dev Biol* (Academic Press), pp. 401–434.
- Espelund, M., De Bedout, J.A., Outlaw, W.H., and Jakobsen, K.S.** (1995). Environmental and hormonal regulation of barley late-embryogenesis-abundant (Lea) mRNAs is via different signal transduction pathways. *Plant Cell Environ* **18**: 943–949.

- Feschotte, C., Jiang, N., and Wessler, S.R.** (2002). Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* **3**: 329–341.
- Fetch, T.G., Steffenson, B.J., and Nevo, E.** (2003). Diversity and Sources of Multiple Disease Resistance in *Hordeum spontaneum*. *Plant Dis* **87**: 1439–1448.
- Figueiredo, D.D., Batista, R.A., Roszak, P.J., and Köhler, C.** (2015). Auxin production couples endosperm development to fertilization. *Nat Plants* **1**: 1–6.
- Filipescu, D., Müller, S., and Almouzni, G.** (2014). Histone H3 Variants and Their Chaperones During Development and Disease: Contributing to Epigenetic Control. *Annu Rev Cell Dev Biol* **30**: 615–646.
- Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C.** (2008). Molecular aspects of seed dormancy. *Annu Rev Plant Biol* **59**: 387–415.
- Fischer, C. and Neuhaus, G.** (1996). Influence of auxin on the establishment of bilateral symmetry in monocots. *Plant J* **9**: 659–669.
- Food and Agriculture Organization of the United Nations** (2023).
- Freeman, P.L. and Palmer, G.H.** (1984). The structure of the pericarp and testa of barley. *J Inst Brew* **90**: 88–94.
- Fukagawa, T. and Earnshaw, W.C.** (2014). The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Dev Cell* **30**: 496–508.
- Gallego-Bartolomé, J.** (2020). DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytol* **227**: 38–44.
- Gehring Lab Plant Imprinting Database** (2020).
- Gehring, M.** (2004). Imprinting and seed development. *Plant Cell* **16**: S203–S213.
- Gehring, M., Bubb, K.L., and Henikoff, S.** (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**: 1447–1451.
- Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**: 495–506.
- Gehring, M., Missirian, V., and Henikoff, S.** (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS One* **6**: e23687.
- Gaiimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggreffe, T.** (2019). The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* **12**: 1–22.

- Gilles, G.J., Hines, K.M., Manfre, A.J., and Marcotte, W.R.** (2007). A predicted N-terminal helical domain of a Group 1 LEA protein is required for protection of enzyme activity from drying. *Plant Physiol Biochem* **45**: 389–399.
- Goldberg, R.B., De Paiva, G., and Yadegari, R.** (1994). Plant embryogenesis: zygote to seed. *Science* **266**: 605–614.
- Gómez, E., Royo, J., Guo, Y., Thompson, R., and Hueros, G.** (2002). Establishment of cereal endosperm expression domains identification and properties of a maize transfer cell-specific transcription factor, ZmMRP-1. *Plant Cell* **14**: 599–610.
- Gómez, E., Royo, J., Muñiz, L.M., Sellam, O., Paul, W., Gerentes, D., Barrero, C., López, M., Perez, P., and Hueros, G.** (2009). The Maize transcription factor myb-related protein-1 is a key regulator of the differentiation of transfer cells. *Plant Cell* **21**: 2022–2035.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldán-Arjona, T., David, L., and Zhu, J.K.** (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**: 803–814.
- Grossniklaus, U., Vielle-Calzada, J.P., Hoepfner, M.A., and Gagliano, W.B.** (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science* **280**: 446–450.
- Gu, B., Dong, H., Smith, C., Cui, G., Li, Y., and Bevan, M.W.** (2022). Modulation of receptor-like transmembrane kinase 1 nuclear localization by DA1 peptidases in Arabidopsis. *Proc Natl Acad Sci U S A* **119**: e2205757119.
- Gubatz, S. and Shewry, P.R.** (2010). The Development, Structure, and Composition of the Barley Grain. *Barley: Production, Improvement, and Uses*: 391–448.
- Gutiérrez-Marcos, J.F., Costa, L.M., Biderre-Petit, C., Khbaya, B., O’Sullivan, D.M., Wormald, M., Perez, P., and Dickinson, H.G.** (2004). maternally expressed gene1 Is a Novel Maize Endosperm Transfer Cell-Specific Gene with a Maternal Parent-of-Origin Pattern of Expression. *Plant Cell* **16**: 1288.
- Haberle, V. and Stark, A.** (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**: 621–637.
- Haig, D. and Westoby, M.** (1989). Parent-Specific Gene Expression and the Triploid Endosperm. *Am Nat* **134**: 147–155.
- Hampsey, M.** (1998). Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiol Mol Biol Rev* **62**: 465–503.

- Hand, S.C., Menze, M.A., Toner, M., Boswell, L., and Moore, D.** (2011). LEA Proteins During Water Stress: Not Just for Plants Anymore. *Annu Rev Physiol* **73**: 115–134.
- He, X.J., Chen, T., and Zhu, J.K.** (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res* **21**: 442–465.
- Heller, M.J.** (2002). DNA Microarray Technology: Devices, Systems, and Applications. *Annu Rev Biomed Eng* **4**: 129–153.
- Hertig, C., Rutten, T., Melzer, M., Schippers, J.H.M., and Thiel, J.** (2023). Dissection of Developmental Programs and Regulatory Modules Directing Endosperm Transfer Cell and Aleurone Identity in the Syncytial Endosperm of Barley. *Plants* **12**.
- Heslop-Harrison, J.S.** (2000). Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. *Plant Cell* **12**: 617–36.
- Holt, R.A. and Jones, S.J.M.** (2008). The new paradigm of flow cell sequencing. *Genome Res* **18**: 839–846.
- Hornslien, K.S., Miller, J.R., and Grini, P.E.** (2019). Regulation of parent-of-origin allelic expression in the endosperm. *Plant Physiol* **180**: 1498–1519.
- Hsieh, T.F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D.** (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science* **324**: 1451–1454.
- Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L.** (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci U S A* **108**: 1755–1762.
- Huang, F., Zhu, Q.H., Zhu, A., Wu, X., Xie, L., Wu, X., Helliwell, C., Chaudhury, A., Finnegan, E.J., and Luo, M.** (2017). Mutants in the imprinted PICKLE RELATED 2 gene suppress seed abortion of fertilization independent seed class mutants and paternal excess interploidy crosses in Arabidopsis. *Plant J* **90**: 383–395.
- Hudakova, S., Michalek, W., Presting, G.G., Hoopen, R. Ten, Dos Santos, K., Jasencakova, Z., and Schubert, I.** (2001). Sequence organization of barley centromeres. *Nucleic Acids Res* **29**: 5029.
- Hueros, G., Royo, J., Maitz, M., Salamini, F., and Thompson, R.D.** (1999). Evidence for factors regulating transfer cell-specific expression in maize endosperm. *Plant Mol Biol* **41**: 403–414.
- Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L.** (2008). Cellular programming of plant gene imprinting. *Cell* **132**: 735–744.

- Ibarra, C.A. et al.** (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360–1364.
- Idziak, D., Robaszkiewicz, E., and Hasterok, R.** (2015). Spatial distribution of centromeres and telomeres at interphase varies among *Brachypodium* species. *J Exp Bot* **66**: 6623–6634.
- Ingouff, M., Selles, B., Michaud, C., Vu, T.M., Berger, F., Schorn, A.J., Autran, D., Van Durme, M., Nowack, M.K., Martienssen, R.A., and Grimanelli, D.** (2017). Live-cell analysis of DNA methylation during sexual reproduction in *Arabidopsis* reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev* **31**: 72–83.
- Ingram, G.C.** (2004). Between the sheets: inter-cell-layer communication in plant development. *Philos Trans R Soc Lond B Biol Sci* **359**: 891–906.
- Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y. V, Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D.** (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol* **16**: 763–768.
- Jayakodi, M. et al.** (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**: 284–289.
- Jiang, D. and Berger, F.** (2017). DNA replication–coupled histone modification maintains Polycomb gene silencing in plants. *Science* (1979) **357**: 1146–1149.
- Jiang, D., Borg, M., Lorković, Z.J., Montgomery, S.A., Osakabe, A., Yelagandula, R., Axelsson, E., and Berger, F.** (2020). The evolution and functional divergence of the histone H2B family in plants. *PLoS Genet* **16**: e1008964.
- Jin, L., Lloyd, R. V, and Lloyd, R.** (1997). In Situ Hybridization: Methods and Applications. *J Clin Lab Anal* **11**: 2–9.
- Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F.** (2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in *Arabidopsis*. *Curr Biol* **16**: 486–492.
- Kalla, R., Shimamoto, K., Potter, R., Nielsen, P.S., Linnestad, C., and Olsen, O. -A** (1994). The promoter of the barley aleurone-specific gene encoding a putative 7 kDa lipid transfer protein confers aleurone cell-specific expression in transgenic rice. *Plant J* **6**: 849–860.
- Kent, N.L.** (1994). *Kent's Technology of Cereals: An introduction for students of food science and agriculture* (elsevier).

- Kermode, A.R. and Finch-Savage, B.E.** (2002). Desiccation sensitivity in orthodox and recalcitrant seeds in relation to development. In *Desiccation and survival in plants: drying without dying* (CABI Publishing: UK), pp. 149–184.
- Kiesselbach, T.A.** (1999). *The structure and reproduction of corn* (Cold spring harbor laboratory press).
- Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (1999). Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. *Plant Cell* **11**: 1945–1952.
- Kirby, J. and Keasling, J.D.** (2009). Biosynthesis of plant isoprenoids: Perspectives for microbial engineering. *Annu Rev Plant Biol* **60**: 335–355.
- Klemsdal, S.S., Hughes, W., Lönneborg, A., Aalen, R.B., and Olsen, O.A.** (1991). Primary structure of a novel barley gene differentially expressed in immature aleurone layers. *Mol Gen Genet* **228**: 9–16.
- Kowles, R. V.** (2009). The importance of DNA endoreduplication in the developing endosperm of maize. *Maydica* **54**: 387–399.
- Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., Zhang, Y., Handa, A.K., and Zhu, J.K.** (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc Natl Acad Sci U S A* **114**: E4511–E4519.
- Larkins, B.A., Dilkes, B.P., Dante, R.A., Coelho, C.M., Woo, Y.M., and Liu, Y.** (2001). Investigating the hows and whys of DNA endoreduplication. *J Exp Bot* **52**: 183–192.
- Law, J.A. and Jacobsen, S.E.** (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220.
- Le, B.H. et al.** (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A* **107**: 8063–8070.
- Leah, R., Skriver, K., Knudsen Soren, Ruud-Hansen, J., Raikhel, N. V., and Mundy, J.** (1994). Identification of an enhancer/silencer sequence directing the aleurone-specific expression of a barley chitinase gene. *Plant J* **6**: 579–589.
- Leprince, O. and Buitink, J.** (2010). Desiccation tolerance: From genomics to the field. *Plant Sci* **179**: 554–564.
- Leroux, B.M., Goodyke, A.J., Schumacher, K.I., Abbott, C.P., Clore, A.M., Yadegari, R., Larkins, B.A., and Dannenhoffer, J.M.** (2014). Maize early

- endosperm growth and development: From fertilization through cell type differentiation. *Am J Bot* **101**: 1259–1274.
- Li, E., Liu, H., Huang, L., Zhang, X., Dong, X., Song, W., Zhao, H., and Lai, J.** (2019). Long-range interactions between proximal and distal regulatory regions in maize. *Nat Commun* **10**: 1–14.
- Li, X. et al.** (2008). High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell* **20**: 259–276.
- Liew, L.C., Narsai, R., Wang, Y., Berkowitz, O., Whelan, J., and Lewsey, M.G.** (2020). Temporal tissue-specific regulation of transcriptomes during barley (*Hordeum vulgare*) seed germination. *Plant J* **101**: 700–715.
- Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E.** (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* (1979) **292**: 2077–2080.
- Lingle, S.E. and Chevalier, P.** (1985). Development of the Vascular Tissue of the Wheat and Barley Caryopsis as Related to the Rate and Duration of Grain Filling1. *Crop Sci* **25**: 123–128.
- Liu, B., Liu, Y., Wang, B., Luo, Q., Shi, J., Gan, J., Shen, W.H., Yu, Y., and Dong, A.** (2019). The transcription factor OsSUF4 interacts with SDG725 in promoting H3K36me3 establishment. *Nat Commun* **10**: 1–14.
- Locascio, A., Roig-Villanova, I., Bernardi, J., and Varotto, S.** (2014). Current perspectives on the hormonal control of seed development in Arabidopsis and maize: A focus on auxin. *Front Plant Sci* **5**: 97587.
- Lockhart, D.J. and Winzler, E.A.** (2000). Genomics, gene expression and DNA arrays. *Nature* **405**: 827–836.
- Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R.A.** (2012). Revisiting Global Gene Expression Analysis. *Cell* **151**: 476.
- Lu, F., Cui, X., Zhang, S., Jenuwein, T., and Cao, X.** (2011). Arabidopsis REF6 is a histone H3 lysine 27 demethylase. *Nat Genet* **43**: 715–719.
- Lu, J. and Magnani, E.** (2018). Seed tissue and nutrient partitioning, a case for the nucellus. *Plant Reprod* **31**: 309–317.

- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J.** (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–260.
- Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A.** (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc Natl Acad Sci U S A* **97**: 10637–10642.
- Luxová, M.** (1986). The seminal root primordia in barley and the participation of their non-meristematic cells in root construction. *Biol Plant* **28**: 161–167.
- Madrid, S.M.** (1991). The Barley lipid transfer protein is targeted into the lumen of the endoplasmic reticulum. *Plant Physiol. Biochem.* **29**: 695–703.
- Magnard, J.L., Lehouque, G., Massonneau, A., Frangne, N., Heckel, T., Gutierrez-Marcos, J.F., Perez, P., Dumas, C., and Rogowsky, P.M.** (2003). ZmEBE genes show a novel, continuous expression pattern in the central cell before fertilization and in specific domains of the resulting endosperm after fertilization. *Plant Mol Biol* **53**: 821–836.
- Maheshwari, P.** (1950). An introduction to the embryology of angiosperms.
- Marand, A.P., Eveland, A.L., Kaufmann, K., and Springer, N.M.** (2023). cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annu Rev Plant Biol* **74**: 111–137.
- Marmorstein, R. and Zhou, M.M.** (2014). Writers and readers of histone acetylation: Structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol* **6**.
- Martínez, G., Panda, K., Köhler, C., and Slotkin, R.K.** (2016). Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nat Plants* **2**.
- Mascher, M. et al.** (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**: 427–433.
- Mascher, M. et al.** (2021). Long-read sequence assembly: A technical evaluation in barley. *Plant Cell* **33**: 1888–1906.
- Matus, I., Corey, A., Filichkin, T., Hayes, P.M., Vales, M.I., Kling, J., Riera-Lizarazu, O., Sato, K., Powell, W., and Waugh, R.** (2003). Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. *Genome* **46**: 1010–1023.

- McDonald, J.F., Matzke, M.A., and Matzke, A.J.** (2005). Host defenses to transposable elements and the evolution of genomic imprinting. *Cytogenet Genome Res* **110**: 242–249.
- McKevith, B.** (2004). Nutritional aspects of cereals. *Nutr Bull* **29**: 111–142.
- Meagher, T.R. and Vassiliadis, C.** (2005). Phenotypic impacts of repetitive DNA in flowering plants. *New Phytol* **168**: 71–80.
- Mehrotra, S. and Goyal, V.** (2014). Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics Proteomics Bioinformatics* **12**: 164–171.
- Melaragno, J.E., Mehrotra, B., and Coleman, A.W.** (1993). Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *Plant Cell* **5**: 1661–1668.
- Melters, D.P. et al.** (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**: R10.
- Mena, M., Vicente-Carbajosa, J., Schmidt, R.J., and Carbonero, P.** (1998). An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. *Plant J* **16**: 53–62.
- Merry, J.** (1941). Studies on the Embryo of *Hordeum sativum*-I. The Development of the Embryo. *J Torrey Bot Soc* **68**: 585.
- Middleton, C.P., Stein, N., Keller, B., Kilian, B., and Wicker, T.** (2013). Comparative analysis of genome composition in Triticeae reveals strong variation in transposable element dynamics and nucleotide diversity. *Plant J* **73**: 347–356.
- Mirabet, V., Das, P., Boudaoud, A., and Hamant, O.** (2011). The role of mechanical forces in plant morphogenesis. *Annu Rev Plant Biol* **62**: 365–385.
- Mo, Y. and Jiao, Y.** (2022). Advances and applications of single-cell omics technologies in plant research. *Plant J* **110**: 1551–1563.
- Moreno-Romero, J., Jiang, H., Santos-González, J., and Köhler, C.** (2016). Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *EMBO J* **35**: 1298–1311.
- Morris, P.C., Jewer, P.C., and Bowles, D.J.** (1991). Changes in water relations and endogenous abscisic acid content of wheat and barley grains and embryos during development. *Plant Cell Environ* **14**: 443–446.

- Morton, T., Petricka, J., Corcoran, D.L., Li, S., Winter, C.M., Carda, A., Benfey, P.N., Ohler, U., and Megraw, M.** (2014). Paired-End Analysis of Transcription Start Sites in Arabidopsis Reveals Plant-Specific Promoter Signatures. *Plant Cell* **26**: 2746–2760.
- Mozgova, I. and Hennig, L.** (2015). The polycomb group protein regulatory network. *Annu Rev Plant Biol* **66**: 269–296.
- Němečková, A., Kolářčková, V., Vrána, J., Doležel, J., and Hřibová, E.** (2020). DNA replication and chromosome positioning throughout the interphase in three-dimensional space of plant nuclei. *J Exp Bot* **71**: 6262–6272.
- Nguyen, H.N., Sabelli, P.A., and Larkins, B.A.** (2007). Endoreduplication and Programmed Cell Death in the Cereal Endosperm. In *Endosperm* (Springer Berlin Heidelberg: Berlin, Heidelberg), pp. 21–43.
- Nielsen, K.A.** (1990). Polyamine Content in Relation to Embryo Growth and Dedifferentiation in Barley (*Hordeum vulgare* L.). *J Exp Bot* **41**: 849–854.
- van Norman, J.M., Breakfield, N.W., and Benfey, P.N.** (2011). Intercellular Communication during Plant Development. *Plant Cell* **23**: 855.
- Nowicka, A., Ferková, E., Said, M., Kovacik, M., Zwyrtková, J., Baroux, C., and Pecinka, A.** (2023). Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues. *J Exp Bot* **74**: 2527–2541.
- Nowicka, A., Kovacik, M., Tokarz, B., Vrána, J., Zhang, Y., Weigt, D., Doležel, J., and Pecinka, A.** (2021a). Dynamics of endoreduplication in developing barley seeds. *J Exp Bot* **72**: 268–282.
- Nowicka, A., Sahu, P.P., Kovacik, M., Weigt, D., Tokarz, B., Krugman, T., and Pecinka, A.** (2021b). Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel. *Genes (Basel)* **12**: 711.
- Olsen, O.-A.** (2001). ENDOSPERM DEVELOPMENT: Cellularization and cell fate specification. *Annu Rev Plant Biol* **52**: 233–267.
- Olsen, O.-A.** (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* **16**: S214–S227.
- Opsahl-Ferstad, H.G., Deunff, E. Le, Dumas, C., and Rogowsky, P.M.** (1997). ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *Plant J* **12**: 235–246.
- Orman-Ligeza, B. et al.** (2020). LYS3 encodes a prolamin-box-binding transcription factor that controls embryo growth in barley and wheat. *J Cereal Sci* **93**: 102965.

- Ortega-Galisteo, A.P., Morales-Ruiz, T., Ariza, R.R., and Roldán-Arjona, T. (2008).** Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol Biol* **67**: 671–681.
- Park, K., Kim, M.Y., Vickers, M., Park, J.S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., and Scholten, S. (2016).** DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proc Natl Acad Sci U S A* **113**: 15138–15143.
- Pavlu, S., Nikumbh, S., Kovacik, M., An, T., Lenhard, B., Simkova, H., and Navratilova, P. (2024).** Core promoterome of barley embryo. *Comput Struct Biotechnol J* **23**: 264–277.
- Peirats-Llobet, M., Yi, C., Liew, L.C., Berkowitz, O., Narsai, R., Lewsey, M.G., and Whelan, J. (2023).** Spatially resolved transcriptomic analysis of the germinating barley grain. *Nucleic Acids Res* **51**: 7798–7819.
- Peng, Y. et al. (2019).** Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat Commun* **10**: 1–11.
- Peng, Y., Chen, L., Lu, Y., Wu, Y., Dumenil, J., Zhu, Z., Bevan, M.W., and Lia, Y. (2015).** The ubiquitin receptors DA1, DAR1, and DAR2 redundantly regulate endoreduplication by modulating the stability of TCP14/15 in Arabidopsis. *Plant Cell* **27**: 649–662.
- Penterman, J., Zilberman, D., Jin, H.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007).** DNA demethylation in the Arabidopsis genome. *Proc Natl Acad Sci U S A* **104**: 6752–6757.
- Peris, C.I.L., Rademacher, E.H., and Weijers, D. (2010).** Green Beginnings – Pattern Formation in the Early Plant Embryo. In *Curr Top Dev Biol* (Elsevier Inc.), pp. 1–27.
- Pires, N.D. and Grossniklaus, U. (2014).** Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Rep* **6**.
- Pirrello, J. et al. (2018).** Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. *Plant J* **93**: 387–398.
- Price, P.B. and Parsons, J. (1979).** Distribution of Lipids in Embryonic Axis, Bran-Endosperm, and Hull Fractions of Hullless Barley and Hullless Oat Grain. *J Agric Food Chem* **27**: 813–815.
- Rabl, C. (1885).** *Über Zelltheilung. Morphologisches Jahrbuch*: 214–330.

- Radchuk, V. V., Borisjuk, L., Sreenivasulu, N., Merx, K., Mock, H.P., Rolletschek, H., Wobus, U., and Weschke, W.** (2009). Spatiotemporal Profiling of Starch Biosynthesis and Degradation in the Developing Barley Grain. *Plant Physiol* **150**: 190–204.
- Radchuk, V., Weier, D., Radchuk, R., Weschke, W., and Weber, H.** (2011). Development of maternal seed tissue in barley is mediated by regulated cell expansion and cell disintegration and coordinated with endosperm growth. *J Exp Bot* **62**: 1217–1227.
- Raghavan, V.** (2006). Double Fertilization — A Defining Feature of Flowering Plants. In *Double Fertilization* (Springer, Berlin, Heidelberg), pp. 1–28.
- Rathjen, J.R., Strounina, E. V., and Mares, D.J.** (2009). Water movement into dormant and non-dormant wheat (*Triticum aestivum* L.) grains. *J Exp Bot* **60**: 1619–1631.
- Rebernick, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T., and Köhler, C.** (2015). Non-reciprocal Interspecies Hybridization Barriers in the *Capsella* Genus Are Established in the Endosperm. *PLOS Genet* **11**: e1005295.
- Robertson, K.D.** (2005). DNA methylation and human disease. *Nat Rev Genet* **6**: 597–610.
- Rodríguez, M. V., Barrero, J.M., Corbineau, F., Gubler, F., and Benech-Arnold, R.L.** (2015). Dormancy in cereals (not too much, not so little): About the mechanisms behind this trait. *Seed Sci Res* **25**: 99–119.
- Roeder, A.H.K., Chickarmane, V., Cunha, A., Obara, B., Manjunath, B.S., and Meyerowitz, E.M.** (2010). Variability in the Control of Cell Division Underlies Sepal Epidermal Patterning in *Arabidopsis thaliana*. *PLoS Biol* **8**: e1000367.
- Rolletschek, H., Weschke, W., Weber, H., Wobus, U., and Borisjuk, L.** (2004). Energy state and its control on seed development: starch accumulation is associated with high ATP and steep oxygen gradients within barley grains. *J Exp Bot* **55**: 1351–1359.
- Rost, T.L., Artucio, P.I. de, and Risley, E.B.** (1984). Transfer cells in the placental pad and caryopsis coat of *Pappophorum subbulbosum* arch. (Poaceae). *Am J Bot* **71**: 948–957.
- Rotasperti, L., Sansoni, F., Mizzotti, C., Tadini, L., and Pesaresi, P.** (2020). Barley's second spring as a model organism for chloroplast research. *Plants (Basel)* **9**: 1–25.

- Roth, M., Florez-Rueda, A.M., and Städler, T.** (2019). Differences in Effective Ploidy Drive Genome-Wide Endosperm Expression Polarization and Seed Failure in Wild Tomato Hybrids. *Genetics* **212**: 141–152.
- Roudier, F. et al.** (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* **30**: 1928–1938.
- Sabelli, P.A.** (2012). Replicate and die for your own good: Endoreduplication and cell death in the cereal endosperm. *J Cereal Sci* **56**: 9–20.
- Sabelli, P.A. and Larkins, B.A.** (2009). The development of endosperm in grasses. *Plant Physiol* **149**: 14–26.
- Schenborn, E. and Groskreutz, D.** (1999). Reporter gene vectors and assays. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology* **13**: 29–44.
- Schnable, P.S. et al.** (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* (1979) **326**: 1112–1115.
- Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.S., Park, G.T., Fischer, R.L., and Tamaru, H.** (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. *Proc Natl Acad Sci U S A* **108**: 8042–8047.
- Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G.** (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. *Development* **125**: 3329–3341.
- Segundo-Val, I.S., Asthma, C.S.-L.-M. genetics of, and 2016, U.** (2016). Introduction to the gene expression analysis. In *Methods in Molecular Biology* (Humana Press Inc.), pp. 29–43.
- Sekine, D., Ohnishi, T., Furuumi, H., Ono, A., Yamada, T., Kurata, N., and Kinoshita, T.** (2013). Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J* **76**: 792–799.
- Shahan, R., Nolan, T.M., and Benfey, P.N.** (2021). Single-cell analysis of cell identity in the Arabidopsis root apical meristem: insights and opportunities. *J Exp Bot* **72**: 6679–6686.
- Shahbazian, M.D. and Grunstein, M.** (2007). Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu Rev Biochem* **76**: 75–100.
- Shan, W., Kubová, M., Mandáková, T., and Lysak, M.A.** (2021). Nuclear organization in crucifer genomes: nucleolus-associated telomere clustering is not a universal interphase configuration in Brassicaceae. *Plant J* **108**: 528–540.

- Shaw, R., Tian, X., and Xu, J.** (2021). Single-Cell Transcriptome Analysis in Plants: Advances and Challenges. *Mol Plant* **14**: 115–126.
- Shen, Q.J., Casaretto, J.A., Zhang, P., and Ho, T.H.D.** (2004). Functional definition of ABA-response complexes: The promoter units necessary and sufficient for ABA induction of gene expression in barley (*Hordeum vulgare* L.). *Plant Mol Biol* **54**: 111–124.
- Shewry, P.R. and Ullrich, S.E.** (2016). *Barley: Chemistry and Technology: Second Edition*. Barley: Chemistry and Technology: Second Edition: 1–322.
- Shih, M. der, Hoekstra, F.A., and Hsing, Y.I.C.** (2008). Late Embryogenesis Abundant Proteins. *Adv Bot Res* **48**: 211–255.
- Simpson, G.M.** (1990). Seed Dormancy in Grasses. *Seed Dormancy in Grasses*.
- Skriver, K., Leah, R., Müller-Uri, F., Olsen, F.L., and Mundy, J.** (1992). Structure and expression of the barley lipid transfer protein gene *Ltp1*. *Plant Mol Biol* **18**: 585–589.
- Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A.** (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461–472.
- De Smet, I. and Beeckman, T.** (2011). Asymmetric cell division in land plants and algae: the driving force for differentiation. *Nat Rev Mol Cell Biol* **12**: 177–188.
- Smith, A.M.** (1999). Making starch. *Curr Opin Plant Biol* **2**: 223–229.
- Smith, L.M., Handley, J., Li, Y., Martin, H., Donovan, L., and Bowles, D.J.** (1992). Temporal and spatial regulation of a novel gene in barley embryos. *Plant Mol Biol* **20**: 255–266.
- Sreenivasulu, N., Borisjuk, L., Junker, B.H., Mock, H.-P., Rolletschek, H., Seiffert, U., Weschke, W., and Wobus, U.** (2010). Barley Grain Development. In *Int Rev Cell Mol Biol*, pp. 49–89.
- Sreenivasulu, N., Radchuk, V., Strickert, M., Miersch, O., Weschke, W., and Wobus, U.** (2006). Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA-regulated maturation in developing barley seeds. *Plant J* **47**: 310–327.
- Sreenivasulu, N., Usadel, B., Winter, A., Radchuk, V., Scholz, U., Stein, N., Weschke, W., Strickert, M., Close, T.J., Stitt, M., Graner, A., and Wobus, U.** (2008). Barley grain maturation and germination: Metabolic pathway and regulatory

- network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant Physiol* **146**: 1738–1758.
- Stacy, R.A.P., Norcleng, T.W., Culiáñez-Macià, F.A., and Aalen, R.B.** (1999). The dormancy-related peroxiredoxin anti-oxidant, PER1, is localized to the nucleus of barley embryo and aleurone cells. *Plant J* **19**: 1–8.
- Stoute, A.I., Varenko, V., King, G.J., Scott, R.J., and Kurup, S.** (2012). Parental genome imbalance in *Brassica oleracea* causes asymmetric triploid block. *Plant J* **71**: 503–516.
- Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S.E., and Gutierrez, C.** (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **109**: 5370–5375.
- Talbert, P.B. et al.** (2012). A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* **5**: 1–19.
- Talbert, P.B. and Henikoff, S.** (2016). Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* **18**: 115–126.
- Tang, K., Lang, Z., Zhang, H., and Zhu, J.K.** (2016). The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat Plants* **2**: 1–10.
- The International Barley Genome Sequencing Consortium** (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**: 711–716.
- Thiel, J. et al.** (2021). Transcriptional landscapes of floral meristems in barley. *Sci Adv* **7**: eabf0832.
- Thiel, J., Weier, D., Sreenivasulu, N., Strickert, M., Weichert, N., Melzer, M., Czauderna, T., Wobus, U., Weber, H., and Weschke, W.** (2008). Different Hormonal Regulation of Cellular Differentiation and Function in Nucellar Projection and Endosperm Transfer Cells: A Microdissection-Based Transcriptome Study of Young Barley Grains. *Plant Physiol* **148**: 1436–1452.
- Tillich, H.J.** (2007). Seedling diversity and the homologies of seedling organs in the order poales (monocotyledons). *Ann Bot* **100**: 1413–1429.
- Turinetto, V. and Giachino, C.** (2015). Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic Acids Res* **43**: 2489–2498.

- Vicient, C.M., Suoniemi, A., Anamthawat-Jónsson, K., Tanskanen, J., Beharav, A., Nevo, E., and Schulman, A.H.** (1999). Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* **11**: 1769–1784.
- Waddington, S.R., Cartwright, P.M., and Wall, P.C.** (1983). A quantitative scale of spike initial and pistil development in barley and wheat. *Ann Bot* **51**: 119–130.
- Walbot, V.** (1994). Overview of Key Steps in Aleurone Development. *The Maize Handbook*: 78–80.
- Walker, J.M.** (2011). *RNA H*. Nielsen, ed (Humana Press: Totowa, NJ).
- Wang, G., Jiang, H., Del Toro de León, G., Martinez, G., and Köhler, C.** (2018). Sequestration of a Transposon-Derived siRNA by a Target Mimic Imprinted Gene Induces Postzygotic Reproductive Isolation in *Arabidopsis*. *Dev Cell* **46**: 696–705.e4.
- Wang, Z., Gerstein, M., and Snyder, M.** (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57–63.
- Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M.** (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**: 4221–4233.
- Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., and Köhler, C.** (2010). H3K27me3 Profiling of the Endosperm Implies Exclusion of Polycomb Group Protein Targeting by DNA Methylation. *PLoS Genet* **6**: e1001152.
- Wendrich, J.R. and Weijers, D.** (2013). The *Arabidopsis* embryo as a miniature morphogenesis model. *New Phytol* **199**: 14–25.
- Werker, E.** (1997). *Seed Anatomy* (Schweizerbart Science Publishers: Stuttgart, Germany).
- Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H., and Wobus, U.** (2003). The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. *Plant J* **33**: 395–411.
- Weschke, W., Panitz, R., Sauer, N., Wang, Q., Neubohn, B., Weber, H., and Wobus, U.** (2000). Sucrose transport into barley seeds: Molecular characterization of two transporters and implications for seed development and starch accumulation. *Plant J* **21**: 455–467.

- West, M.A.L. and Harada, J.J.** (1993). Embryogenesis in higher plants: An overview. *Plant Cell* **5**: 1361–1369.
- Wicker, T. et al.** (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973–982.
- Wicker, T. et al.** (2017). The repetitive landscape of the 5100 Mbp barley genome. *Mob DNA* **8**: 22.
- Wisniewski, J.P. and Rogowsky, P.M.** (2004). Vacuolar H⁺-translocating inorganic pyrophosphatase (Vpp1) marks partial aleurone cell fate in cereal endosperm development. *Plant Mol Biol* **56**: 325–337.
- Wolff, P., Jiang, H., Wang, G., Santos-González, J., and Köhler, C.** (2015). Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *Elife* **4**.
- Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C.** (2011). High-Resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet* **7**.
- Wollmann, H., Holec, S., Alden, K., Clarke, N.D., Jacques, P.É., and Berger, F.** (2012). Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the *Arabidopsis* transcriptome. *PLoS Genet* **8**: 28–31.
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C.** (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2**: 642–652.
- Yamamoto, Y., Ichida, H., Abe, T., Suzuki, Y., Sugano, S., and Obokata, J.** (2007). Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. *Nucleic Acids Res* **35**: 6219–6226.
- Yan, W., Chen, D., Smaczniak, C., Engelhorn, J., Liu, H., Yang, W., Graf, A., Carles, C.C., Zhou, D.X., and Kaufmann, K.** (2018). Dynamic and spatial restriction of Polycomb activity by plant histone demethylases. *Nat Plants* **4**: 681–689.
- Yanagisawa, S. and Sheen, J.** (1998). Involvement of Maize Dof Zinc Finger Proteins in Tissue-Specific and Light-Regulated Gene Expression. *Plant Cell* **10**: 75–89.
- Yelagandula, R. et al.** (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* **158**: 98–109.
- Yi, F. et al.** (2019). High temporal-resolution transcriptome landscape of early maize seed development. *Plant Cell* **31**: 974–992.

- Yin, R., Xia, K., and Xu, X.** (2023). Spatial transcriptomics drives a new era in plant research. *Plant J* **116**: 1571–1581.
- Yocca, A.E. and Edger, P.P.** (2022). Current status and future perspectives on the evolution of cis-regulatory elements in plants. *Curr Opin Plant Biol* **65**: 102139.
- Yuan, J., Chen, S., Jiao, W., Wang, L., Wang, L., Ye, W., Lu, J., Hong, D., You, S., Cheng, Z., Yang, D.L., and Chen, Z.J.** (2017). Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol* **216**: 373–387.
- Yuan, Y. et al.** (2024). Decoding the gene regulatory network of endosperm differentiation in maize. *Nat Commun* **15**: 1–19.
- Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N.M., Arakaki, A.M., Burnett, W.J., Logan, K.O., Wang, D., Wang, X., Drews, G.N., and Yadegari, R.** (2015). RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* **27**: 513–531.
- Zhang, H. et al.** (2004). Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J* **40**: 276–290.
- Zhang, H., Lang, Z., and Zhu, J.K.** (2018). Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol* **19**: 489–506.
- Zhang, K., Sridhar, V. V., Zhu, J., Kapoor, A., and Zhu, J.K.** (2007). Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS One* **2**.
- Zhang, M. et al.** (2014). Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res* **24**: 167–176.
- Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Zhao, H., Guan, H., Hu, X., Jiao, Y., Song, W., and Lai, J.** (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A* **108**: 20042–20047.
- Zhang, R., Tucker, M.R., Burton, R.A., Shirley, N.J., Little, A., Morris, J., Milne, L., Houston, K., Hedley, P.E., Waugh, R., and Fincher, G.B.** (2016). The Dynamics of Transcript Abundance during Cellularization of Developing Barley Endosperm. *Plant Physiol* **170**: 1549–1565.

- Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S.E.** (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* **10**: 1–14.
- Zhao, T., Zhan, Z., and Jiang, D.** (2019). Histone modifications and their regulatory roles in plant development and environmental memory. *J Genet Genomics* **46**: 467–476.
- Zheng, S. et al.** (2019). The *Arabidopsis* H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor. *Nat Commun* **10**.
- Zöllner, N.R., Bezruczyk, M., Laureyns, R., Nelissen, H., Simon, R., and Frommer, W.B.** (2021). An RNA in situ hybridization protocol optimized for monocot tissue. *STAR Protoc* **2**: 100398.

7 List of appendices

Original publications

Appendix I Isolation of High Purity Tissues from Developing Barley Seeds

Appendix II Transcriptome landscape of developing barley seeds

Co-author publications

Appendix III Dynamics of endoreduplication in developing barley seeds

Appendix IV Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel

Appendix V Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues

Appendix VI Core promoterome of barley embryo

Published abstracts – poster presentation

Appendix VII Analysis of transcriptome landscape in developing barley seeds

Appendix VIII Transcriptome landscape of endosperm in developing barley seeds

Appendix I

Isolation of High Purity Tissues from Developing Barley Seeds

Kovacik, M., Nowicka, A., and Pecinka, A.

Journal of Visualized Experiments

DOI: 10.3791/61681

IF (2020): 1.4

Isolation of High Purity Tissues from Developing Barley Seeds

Martin Kovacik¹, Anna Nowicka^{1,2}, Ales Pecinka¹

¹ Czech Acad Sci, Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany ²The Polish Academy of Sciences, The Franciszek Górski Institute of Plant Physiology

Corresponding Author

Ales Pecinka
pecinka@ueb.cas.cz

Citation

Kovacik, M., Nowicka, A., Pecinka, A. Isolation of High Purity Tissues from Developing Barley Seeds. *J. Vis. Exp.* (164), e61681, doi:10.3791/61681 (2020).

Date Published

October 26, 2020

DOI

10.3791/61681

URL

jove.com/video/61681

Abstract

Understanding the mechanisms regulating the development of cereal seeds is essential for plant breeding and increasing yield. However, the analysis of cereal seeds is challenging owing to the minute size, the liquid character of some tissues, and the tight inter-tissue connections. Here, we demonstrate a detailed protocol for dissection of the embryo, endosperm, and seed maternal tissues at early, middle, and late stages of barley seed development. The protocol is based on a manual tissue dissection using fine-pointed tools and a binocular microscope, followed by ploidy analysis-based purity control. Seed maternal tissues and embryos are diploid, while the endosperm is triploid tissue. This allows the monitoring of sample purity using flow cytometry. Additional measurements revealed the high quality of RNA isolated from such samples and their usability for high-sensitivity analysis. In conclusion, this protocol describes how to practically dissect pure tissues from developing grains of cultivated barley and potentially also other cereals.

Introduction

Seeds are complex structures composed of several tissues of maternal and filial origin¹. Cereal grains represent a special type of seed, with the largest part being formed by endosperm, a specialized triploid tissue that protects and nourishes the embryo. Cereals provide around 60% of global food resources and are the most valuable output from plant production². The knowledge of molecular processes controlling cereal seed development is important

due to their economic prominence and central role in plant reproduction^{1,3}.

Cultivated barley (*Hordeum vulgare* subsp. *vulgare*; $2n = 2x = 14$; $1C = 5.1$ Gbp) is the fourth most important cereal crop worldwide. It is used for animal feed, food, and biotechnology⁴. Besides that, it is also a classical temperate zone cereal crop model species of growing importance⁵. Barley genomic resources include genetic maps, collections of cultivars, landraces and mutants, high-quality genome assemblies and annotations as well as transcriptomic data of

the major developmental stages^{5,6,7}. Also, barley genes are used for genetic improvements of other cereals. Resistance to abiotic stresses such as drought and salinity, specific pathogens, and high content of beneficial compounds (e.g., β -glucan) make barley a valuable source of traits for wheat breeding⁸.

Seed development is initiated by fertilization on the day of pollination (DOP). DOP is defined by evaluation of the morphology of stigma and anthers according to the Waddington scale (W10.0)⁹. The spikes containing non-pollinated flowers were characterized by compact (unbranched) stigma and green anthers, whereas pollinated spikes contained extended spiklets, extended and widely branched stigma, swollen ovule, opened anthers and free pollen. The flowers at DOP represented an intermediate phenotype. The anthers had a yellow color, disrupted easily and then released pollen. Stigma had widely spread sigmatic branches of the pistil (**Figure 1C**).

Barley seed development includes three partially overlapping stages^{1,10}. The stage I (0 – 6 days after pollination; DAP) is launched by double fertilization, typified by cell proliferation and the absence of starch synthesis; stage II (7 – 20 DAP) comprises differentiation and great biomass gain accompanied by the production of starch and protein storage molecules; stage III (after 21 DAP) corresponds to seed maturation, weight reduction by desiccation and the

onset of dormancy. Alternatively, the phases are called early, middle and late, respectively¹¹.

Barley grain is covered by hulls, which consist of the lemma, palea, and glumes¹². In most barley genotypes, the hulls tightly wrap dry seeds. The seed itself is formed by the embryo, endosperm and seed maternal tissues (**Figure 1A**). The diploid embryo originates from the fertilization of the egg cell by one sperm cell nucleus. In the fully developed seed, the embryo consists of the embryonic axis with the coleorhiza surrounding the radicle, the coleoptile enclosing the shoot meristem and primary leaves, and the scutellum (cotyledon)^{1,10,13,14}. The triploid endosperm is the result of fertilization of the diploid central cell by the second sperm cell nucleus. The proliferation of endosperm begins with the syncytial (coenocyte) stage, where the dividing nuclei are pushed to the periphery by the central vacuole. At the end of the syncytial phase, microtubules form a radial network around the nuclei and indicate the anticlinal cell wall formation and the onset of endosperm cellularization. Endosperm differentiation occurs simultaneously with the cellularization and results in five major tissues: the starchy endosperm, the transfer cells, the aleurone and subaleurone layers, and the embryo surrounding region. Seed maternal tissues are a multi-layered diploid structure of maternal origin containing pericarp and seed coats^{10,12}. Seed maternal tissues include a nucellar projection on the dorsal side of the grain that has a transport-related function, and becomes embedded in endosperm at later stages of seed development¹⁵.

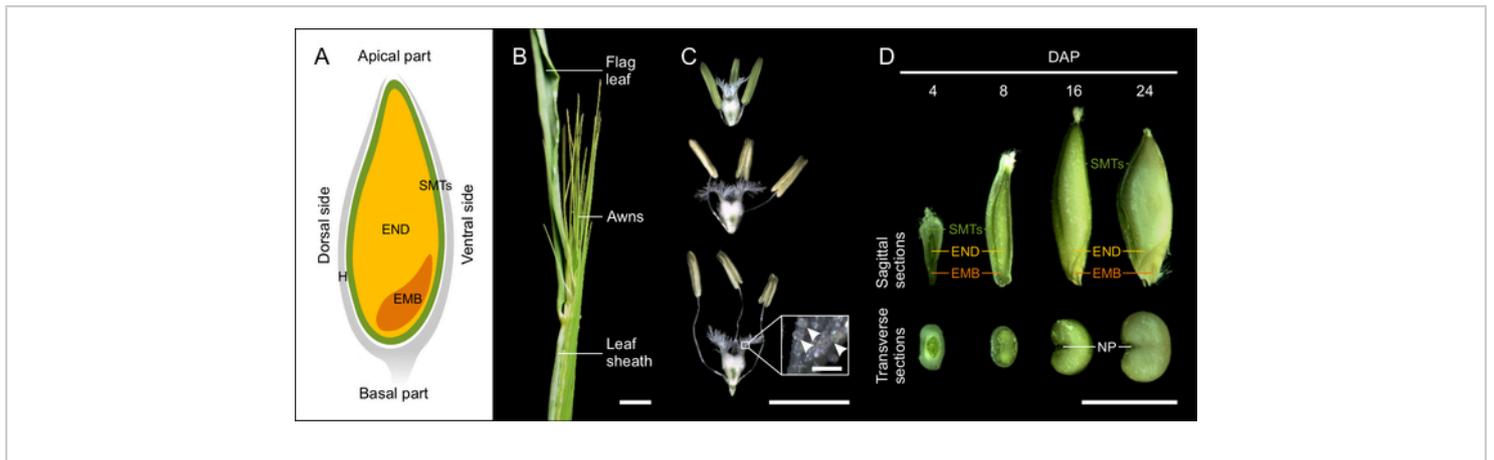


Figure 1: Developing barley seeds. (A) The schematic drawing of cereal grain at the sagittal plan with indicated seed maternal tissues (SMTs, green), endosperm (END, yellow), embryo (EMB, orange) and hulls (H, grey). (B) Morphology of barley spike close to the anthesis. Scale bar = 1 cm. (C) Morphology of stigma and anthers at the stages before, during and after pollination. Inset shows detail of the stigma with pollen grains (arrowheads). Scale bar = 5 mm, inset bar = 200 μ m. (D) Sagittal and transverse sections of 4, 8, 16 and 24 DAP seeds. (NP, nucellar projection) Scale bar = 5 mm. [Please click here to view a larger version of this figure.](#)

Recent progress in high-throughput genomics provides the tools for the study of individual seed tissue development. However, the major obstacle of this purpose is the compact structure and tight adhesion of the seed tissues¹. We developed a protocol for high purity dissection of seed tissues from developing barley seeds with possibility to subsequent use for highly sensitive analyses, such as RNA-sequencing. In addition, the presented protocol can be easily adapted to other cereals.

Protocol

1. Growing plants

NOTE: Considering that a single barley plant usually has 5 to 6 tillers and only the middle 5 to 6 spikelets of each spike should be used for dissection, then a maximum yield per plant is 72 seeds for two-row and 216 seeds for six-row cultivars.

1. To germinate barley seeds, prepare a Petri dish padded with three layers of cellulose tissue paper covered with one layer of filter paper. Moisturize it with distilled water, so there is no excess water, put the seeds on the surface and close the Petri dish. Filtration paper avoids growing the roots through the cellulose tissue. Germinate the seeds for 3 days at 25 °C in the dark.
NOTE: Alternatively, germinate seeds by putting them directly in a wet soil mixture (see step 1.2).
2. Transfer germinated seeds with a visible radicle and shoot of about 5 cm into 5 cm x 5 cm peat pots with a mixture of soil and sand (3:1, v/v). Water regularly. After 10 days, transfer the plants into 12 cm x 12 cm pots filled with the same soil mixture.

- Grow the plants in a climatic chamber under the controlled long-day regime (16 h day 20 °C, 8 h night 16 °C; light intensity 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$; humidity 60%).

NOTE: Spring barley requires approximately 8-10 weeks from sowing to the beginning of anthesis, with no requirement for vernalization. Winter barley needs 7-8 weeks of vernalization (short-day conditions, 8 h day 4 °C, 16 h night 4 °C; light intensity 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$; humidity 85%) to induce flowering.

2. Determination of pollination

NOTE: Precise determination of pollination is needed for proper estimation of developmental progression. Barley is a self-pollinating species. To define day of pollination (DOP), we monitored the day of self-pollination. This trait is cultivar specific, but starting protrusion of the awns from the leaf sheath is a good indicator of approaching DOP (**Figure 1B**).

- Open the leaf-sheath covering the spike. Use fine-pointed tweezers to check anthers and the ovary inside the spikelets in the central part of the spike. Spikelets with yellow anthers and “fluffy” stigma will pollinate within few hours¹⁶ and are considered as DOP (**Figure 1B**).
- Clip off the spike near the tip of the last spikelet, and remove the flag leaf and the upper part of the awns. Then, clip off the top 1/3 of hulls in each spikelet. This dries the anthers and leads to their more synchronized opening and release of pollen.
- Cover the spike with a glassine bag with the spike ID, plant number and defined DOP date. This also prevents cross-pollination, which may compromise specific experiments.

- Note the information to a tabular editor. Use the following formula to calculate the day after pollination (DAP) when tissue isolation should take place.

$$xDAP = DOP + x$$

x = expected DAP

NOTE: The values should have ‘Data’ format.

- For seed tissues dissection, collect the spikes at DAP according to the prepared tabular calendar.

3. Dissection of the seed tissues

NOTE: The following steps should be performed using a stereomicroscope. Remove the hulls before dissection using tweezers. Note that hulls become drier and more adherent from around 16 DAP. To keep physiological conditions and avoid drying of the plant materials during dissection, moisten the samples by putting them into a drop of 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , 1.8 mM KH_2PO_4 , pH = 7.4). Use a new seed for dissection of each tissue to avoid DNA, RNA, or protein degradation due to extended sample collection time. For RNA isolation from dissected material, use only RNase-free materials and chemicals. Do not exceed the total dissection time 15 minutes for one sample consisting typically from tissues dissected from 5-10 seeds to minimize RNA degradation.

- Remove the rest of the hulls using fine-pointed tweezers before tissue dissection. Moisturizing with 1x PBS for 1 minute helps to remove dry residues of the spikelet.
- Place the peeled seed on a Petri dish with a drop of 1x PBS and dissect individual parts using fine-pointed tweezers, fine-needle, and microcapillary pipette. A slightly different strategy is applied for dissection of individual tissues: the seed maternal tissues (step 3.3), the embryo (step 3.4) and the endosperm (step 3.5).

3. Dissection of seed maternal tissues

1. Dissection from seeds up to 8 DAP

1. Place a seed on the dorsal side and gently cut the seed along a longitudinal axis, and peel off seed maternal tissues except the last layer bordering endosperm from the apical to the basal part using tweezers.
2. Collect seed maternal tissues from 5 to 10 seeds into a 1.5 mL tube with 50 μ L of 1 \times PBS, discard the buffer using a pipette, rinse the tissue twice with 100 μ L of PBS, remove excessive buffer by pipetting and close the tube and freeze in liquid nitrogen or use directly for flow cytometric ploidy measurement. The amount of material is sufficient for typically one downstream application (e.g., RNA isolation or flow cytometric ploidy measurement).

2. Dissection from seeds after 8 DAP

1. Place a seed on the dorsal side, gently cut in the middle of the ventral side of seed maternal tissues and gradually peel off the tissue around whole seed including nucellar projection. For each downstream application collect and wash the tissue from 5 to 10 seeds as described in step 3.3.1.

4. Dissection of embryo

1. Dissection from seeds at 8 DAP and younger

1. Place a seed on the dorsal side and cut basal 1/3 of the seed. Carefully split separated part in half and release the embryo. For each downstream application collect and wash the embryos from 10 to 20 seeds as described in 3.3.1.

2. Dissection from seeds after 8 DAP

1. Place a seed on the dorsal side and remove seed maternal tissues from the basal part of the ventral side. Gently disturb the thin layer of endosperm around the perimeter of the embryo by fine-needle or fine-pointed tweezers and peel out the embryo. For each downstream application collect and wash embryos from up to 5 seeds as described in 3.3.1.

5. Dissection of endosperm

1. Dissection of syncytial endosperm from 4 DAP seeds.

1. Place a seed on the dorsal side, and remove seed maternal tissues except the last layer of cells bordering endosperm. Gently puncture layer in the middle of the ventral side by a thin needle, and suck the syncytial endosperm by capillary action using a microcapillary pipette.

2. For each downstream application collect liquid endosperms from 10 to 15 seeds into a new 1.5 mL tube with buffer suitable for the planned downstream analysis (i.e., 1 \times PBS, RNA isolation buffer, flow cytometry buffer). Buffer volume should reflect the protocol for the planned downstream application. Freeze in liquid nitrogen.

2. Dissection of cellularizing endosperm from 5 to 8 DAP seeds

1. Place a seed on the dorsal side, and remove all seed maternal tissues and embryo. For each downstream application collect and wash the endosperm from 10 to 15 seeds into a new 1.5 mL tube with 1 \times PBS and freeze in liquid nitrogen.

3. Dissection of cellularized endosperm from seed after 8 DAP

1. Place a seed on the dorsal side, remove all seed maternal tissues and embryo. For each downstream application collect and wash endosperm from a single seed per tube as described in step 3.3.1.

6. Store the tubes with isolated material at -80°C until use.

NOTE: The protocol can be paused here.

4. Control of tissue purity using flow cytometry

NOTE: The sample purity can be checked using flow cytometry before RNA isolation. Proper instrument calibration is critical for the biological sample analysis. The flow cytometer/ploidy analyzer optics should be adjusted using calibration beads (fluorescently stained polystyrene microspheres highly uniform with respect to their size and fluorescence intensity) until the maximal peak sharpness, typically reaching the coefficient of variation (CV) $< 2\%$. Cereal seed tissues contain mainly populations of G1, G2 and endoreduplicated nuclei; therefore, using a logarithmic scale is recommended. Start with a leaf tissue that contains mostly G1 nuclei and serves as a basal ploidy control.

1. Use freshly prepared samples kept on ice (see step 3.3.1) or frozen tissue as described¹⁷.

NOTE: Because the whole < 8 DAP sample is used for flow cytometry, this represents only an indirect control. We recommend researchers performing multiple isolations and measurements until reaching high proportion of pure samples ($>90\%$) before proceeding to RNA isolation with < 8 DAP samples.

2. Release the nuclei from the 4 and 8 DAP embryo samples (for other samples see step 4.4) by homogenizing the

tissues by 5 to 10 turns of the plastic pestle in 1.5 mL tube containing 300 μL of Otto I buffer (0.1 M citric acid monohydrate, 0.5% (v/v) Tween 20, filtered through a 0.22 μm filter)¹⁸.

3. Filter the crude suspension through 50 μm nylon mesh into a flow-cytometry analysis tube and add 600 μL of Otto II buffer (0.4 M $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$) containing 2 $\mu\text{g mL}^{-1}$ DAPI (4',6-diamidino-2-phenylindole)¹⁸ to stain DNA.
4. Place all other tissues (including 16 DAP or older embryos) on a Petri dish containing 500 μL of Otto I buffer and homogenized by chopping with a razor blade. Filter the suspension as in step 4.3 and stain with 1 mL of Otto II buffer containing DAPI.

NOTE: Manipulation with a sharp double edge razor blade requires special attention. To reduce the risk of injury, there are a single edge razor blades or special blade holders available.

5. Estimate the nuclear DNA content of the sample using a flow cytometer. At least 2000 particles per sample are required for analyzing the sample purity.

5. RNA isolation and quality measurement

1. Use frozen tissue to prevent RNA degradation by endogenous ribonucleases. From seed maternal tissues and embryo samples isolate RNA using commercially available kits or TRIzol reagent¹⁹. Due to a high starch content in endosperm tissues, isolate total RNA from all samples using commercial on-column RNA extraction protocols for problematic tissues (e.g., Spectrum Plant Total RNA Kit) with an on-column DNase I treatment²⁰.

2. Measure RNA concentration and integrity using a dedicated protocol for RNA gel electrophoresis or Agilent 2100 Bioanalyzer.

NOTE: Intact total RNA has a clear 18S and 25S rRNA bands/peaks of size around 1.9 and 3.7 kb respectively. The 25S rRNA band should be approximately two times more intense than the 18S rRNA band.

Representative Results

To perform a tissue-specific transcriptomic analysis of barley seed development, we established a protocol for high purity tissue isolation. The protocol is based on the manual dissection of embryo, endosperm and seed maternal tissues from peeled (after manual hull removal) grains (**Figure 1A**). The protocol was successfully used for isolating materials from several two- and six-row spring barley cultivars, and the spikes were harvested at a given DAP and directly used for extraction without fixation (**Figure 1D**).

The definition of DOP was a critical parameter to be estimated before dissection. Barley spikes mature from the middle towards the edges. Therefore, the middle flower was used for DOP evaluation. At most, six seeds from the middle of the spikes were used for tissue dissection. The success rate

of self-pollination was close to 100%. In comparison, manual pollination had much lower success.

During the selected developmental window (4 to 24 DAP), the difficulty of tissue dissection generally decreased over time. However, it has to be noted that the separation of tissues becomes again more difficult during and after seed desiccation (> 24 DAP) due to tissue adherence. During the dissection of the tissues from 4 DAP seeds, a critical part was not to collapse the squashy syncytial endosperm (**Figure 2A**). Therefore, we removed seed maternal tissues by gentle cut and peel off. The embryo had to be protected against drying by adding buffer. At 8 DAP grain, the isolation strategy was analogous to younger seeds, but the nucellar projection (a part of seed maternal tissues on the dorsal side of the seed) appeared at this stage and its careful removal was required to avoid contamination of endosperm tissues. Conversely, this tissue is an important part of seed maternal tissues and should not be forgotten during isolation of this tissue. At later stages (16 and 24 DAP) seed maternal tissues were more cohesive. Our practical experience was that seed maternal tissues can be dissected and harvested in strips without damaging the endosperm (**Figure 2C**). The perimeter of the embryo should have a clear round shape and its original position in the seed should be clean of any rests of embryo tissues (**Figure 2C**).

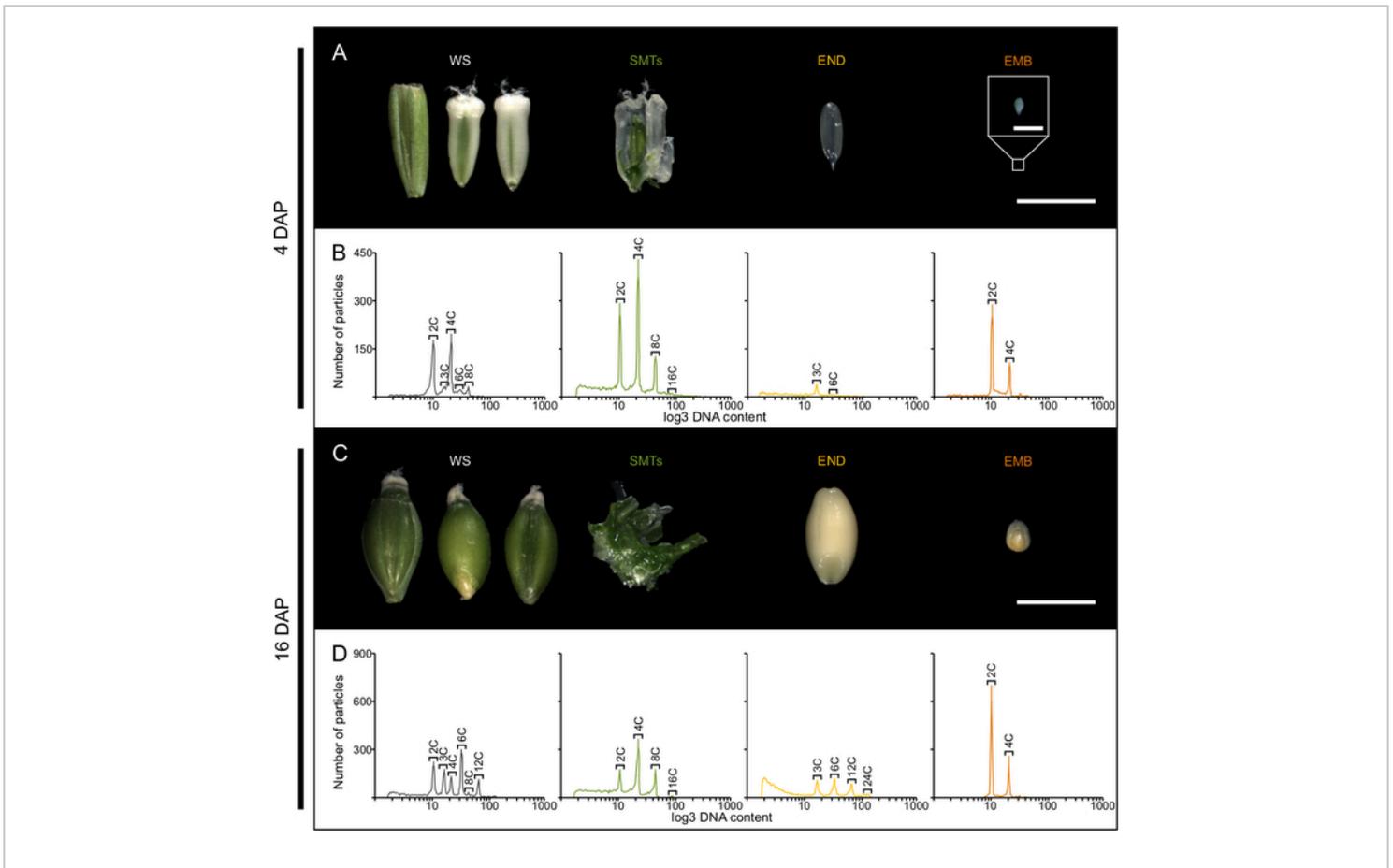


Figure 2: Flow cytometric estimation of the purity of dissected seed tissues. (A) Morphology of dissected seed tissues from 4 DAP seeds. The groups correspond to the whole seeds (WS) photographed as hulled (left) and peeled from ventral and dorsal sides (middle and right, respectively); the next are dissected seed maternal tissues (SMTs), endosperm (END) and embryo (EMB). Scale bar = 5 mm, except for the inset to which 100 μ m scale bar applies. **(B)** Representative histograms of nuclear DNA content obtained from described tissues. The histograms show marked C-value peaks for diploid EMB and/or seed maternal tissues (2C, 4C, 8C) and/or triploid END tissues (3C, 6C, 12C). The x-axis shows DNA content (relative fluorescence on log₃ scale) and the y-axis the number of measured particles. **NOTE:** The scale on the y-axis should not be compared between the histograms as it varies depending on the starting amount of material and duration of the measurement. Only the presence/absence of a peak and relative height within one histogram should be evaluated. **(C, D)** Morphology and flow cytometric profiles of 16 DAP seeds. The figures are organized as in A and B. The flow cytometric measurement reveals also 16C and 24C nuclei, from diploid and triploid tissues, respectively. Scale bar = 5 mm. [Please click here to view a larger version of this figure.](#)

To test the purity of isolated tissues, we estimated nuclear DNA content using flow cytometry (**Figure 2B and D**).

We used fresh barley leave to establish the position of the peak corresponding to diploid (2C) nuclei (**Figure 3A**).

This tissue contained > 95% nuclei with 2C and 4C DNA content, corresponding to G0/G1 and G2 phases of the cell cycle, respectively and < 5% nuclei with 8C and 16C DNA content, corresponding to endoreduplicated nuclei. Next, all subsequent seed-tissue samples were measured with the same flow cytometer settings. Flow cytometric histograms of the whole seeds contained C-value peaks for diploid seed

tissue (a mixture of an embryo and seed maternal tissues; 2C, 4C, 8C and 16C) and triploid endosperm tissues (3C, 6C, 12C and 24C)²¹. In properly dissected seed tissues, only C-value peaks for either the diploid or the triploid tissues were present (**Figure 2B and D**). Samples mixing tissues were identified based on the presence of contaminant diploid or triploid peaks (**Figure 3B and C**).

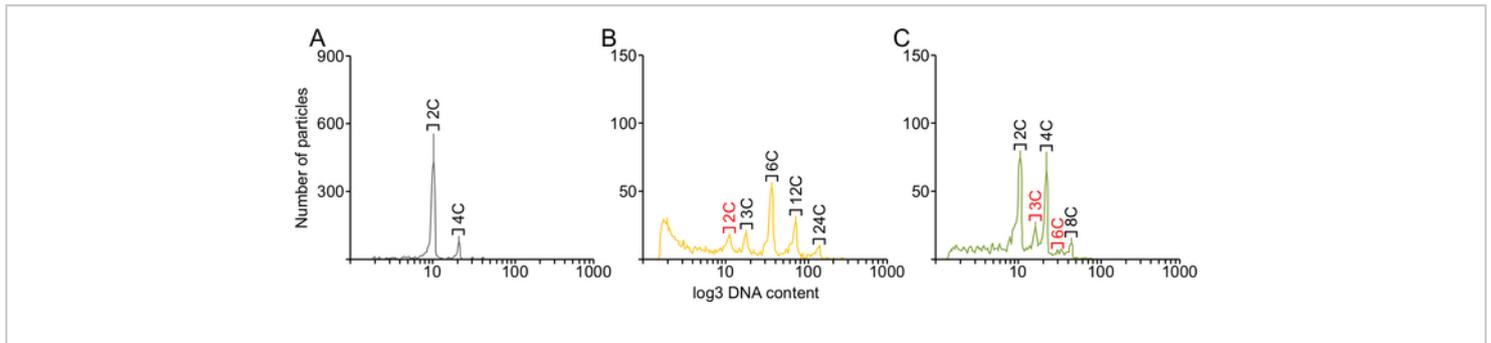


Figure 3: Examples of control and contaminated seed tissue samples as revealed by flow cytometry. (A)

Representative histogram of nuclear DNA content from 10 days old barley leaf representing somatic tissue control. **(B)** Example histogram of dissected 16 DAP endosperm (3C, 6C, 12C and 24C peaks) contaminated by a diploid tissue (2C peak – red-labeled). **(C)** Example histogram of 8 DAP dissected seed maternal tissues contaminated by endosperm tissues as indicated by the presence of 3C and 6C peaks (red-labeled). The x-axis shows DNA content (relative fluorescence on log₃ scale) and the y-axis the number of measured particles. **NOTE:** The scale on the y-axis should not be compared between the histograms as it varies depending on the starting amount of material and duration of the measurement. Only the presence/absence of a peak and relative height within one histogram should be evaluated. [Please click here to view a larger version of this figure.](#)

As an example of downstream use of the samples, we isolated RNA from separated seed tissues using either the commercial RNA isolation kits or TRIzol reagent. However, due to high starch content in endosperm of older seeds (after 16 DAP), we used a commercial column-based kit for RNA isolation from problematic tissues. RNA isolation from endosperm samples older than 16 DAP using TRIzol resulted in insufficient RNA quality and high level of protein contamination. To remove a residual DNA contamination, we

performed on-column DNase I treatment that is an optional step in the commercial kits. The amount of isolated total RNA per sample was 200 – 3,000 ng for endosperm, 600 – 15,000 ng for embryo and 1,500 – 15,000 ng for seed maternal tissues. Next, we assessed the quality of isolated RNA using a bioanalyzer. Although the pattern can differ between tissues, two sharp peaks/bands representing the large and the small ribosomal RNA subunits should be normally present in the spectra/gel (**Figure 4**). The presence of additional peaks and

high background in the fast and the inter-region indicates RNA degradation, whereas the 5S rRNA region includes various types of small rRNAs and those peaks do not affect the quality of the isolated RNA. A signal in the precursor region can

indicate residual genomic DNA contamination. The samples with RNA integrity number (RIN) ≥ 7 are considered of sufficient quality for analysis including reverse transcription PCR or RNA-sequencing.

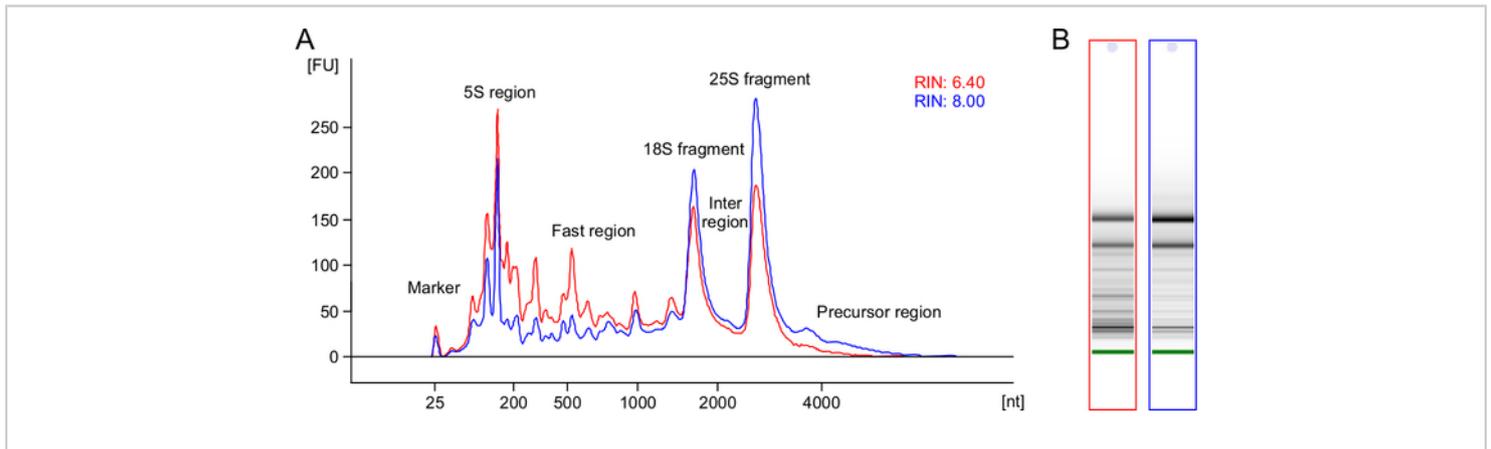


Figure 4: Quality control of isolated total RNA. (A) A representative spectrum of high quality (blue) and partially degraded (red) total RNA from 8 and 16 DAP seed maternal tissues respectively produced using Agilent 2100 Bioanalyzer with the RNA integrity number (RIN) 8.00 and 6.40 respectively. The graph shows the intensity of the peaks of the ribosomal RNA subunits: nuclear large-25S, small-18S and 5S RNA. nt = number of estimated nucleotides based on ladder; FU = relative fluorescence units. (B) The electrophoretic gel-like view of high quality (blue) and partially degraded (red) RNA indicating the subunit bands. [Please click here to view a larger version of this figure.](#)

To test the purity of tissues prepared using this protocol at the molecular level, we performed RNA-sequencing (Kovacik, Nowicka and Pecinka, unpublished data) and analyzed transcript levels for several well known marker genes of embryo, endosperm and seed maternal tissues development (**Figure 5**). As embryo markers, we selected barley homologs of maize *LEAFY COTYLEDON 1* (*HvLEC1*; HORVU.MOREX.r2.6HG0506770) and *GLOBULIN 2* (*HvGLB2*; HORVU.MOREX.r2.5HG0430450), which are important genes for embryogenesis and production of storage protein, respectively^{22,23}. *HvLEC1* transcript was highly abundant in embryo tissues at 8 DAP and its amount strongly decreased at 16 DAP and was absent at 24

DAP embryo and all stages on other analyzed tissues. In contrast, *HvGLB2* transcript started low at 8 DAP embryo, but greatly increased at 16 and 24 DAP embryos. *HvGLB2* showed also low level of transcript in endosperm and seed maternal tissues. Endosperm marker genes were represented by barley homologs of maize *ALEURONE 9* (*HvAL9*; HORVU.MOREX.r2.1HG0010310) and *GLUTENIN SUBUNIT* (*HvGS*; HORVU.MOREX.r2.1HG0001010) which are related to aleurone differentiation^{24,25,26} and energy storage, respectively. The transcripts were highly specific for endosperm tissues, and *HvAL9* peaked at 8 DAP while it was 16 and 24 DAP for *HvGS*, which is consistent with endosperm tissue differentiation and energy accumulation. Our markers

for seed maternal tissues were represented by barley *BETA AMYLASE* (*HvBA*; HORVU.MOREX.r2.2HG0113950) and *CHLOROPHYLL A/B BINDING PROTEIN* (*HvCAB*; HORVU.MOREX.r2.1HG0073450). *BETA AMYLASE* is connected with utilization of first storage protein deposition in seed maternal tissues at early seed development¹. As the

seed maternal tissues are the only green tissue in the seed, CAB proteins are important for photosynthesis. Highly tissue-specific profiles of all selected marker genes demonstrate that our protocol has potential for genotyping samples without or with only minimal contamination from surrounding tissues.

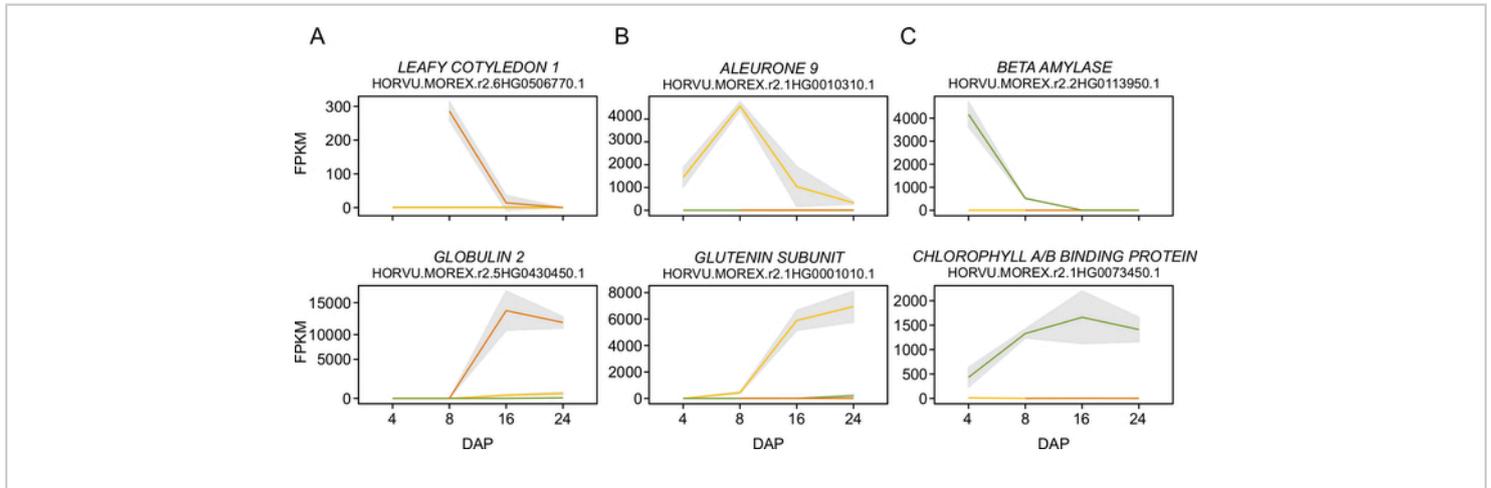


Figure 5: Examples of expression from marker genes. PolyA enriched mRNAs from barley seed tissues were sequenced in three biological replicates using Illumina platform. The graphs show an average fragments per kilobase per million reads (FPKM) at different days after pollination (DAP) in embryo (orange lines), endosperm (yellow lines) and seed maternal tissues (green lines). Standard deviation between biological triplicates is indicated by the gray field. Two examples of early (top row) and late (bottom row) marker genes are shown for (A) embryo, (B) endosperm and (C) seed maternal tissues.

[Please click here to view a larger version of this figure.](#)

Discussion

Here, we present a protocol that allows high purity isolation of barley seed tissues. Although it was developed and tested for barley, it can be easily adopted to other members of the *Triticeae* tribe such as wheat, oat, rye or triticale²⁷. The initial part of the protocol, focusing on seed tissue dissection, does not require any non-standard or expensive equipment and therefore should be accessible to many scientists. A highly specialized instrument such as a flow cytometer is required for the thorough quality control analysis. However, many plant

research institutions have a flow cytometer or ploidy analyzer operated by a trained research staff.

For plant pollination, we make use of the barley's ability to self-pollinate and rely on a set of simple morphological parameters that define the exact day of natural pollination. Hence, the protocol avoids manual flower emasculum and pollination that is a common approach applied to many plant species. We have initially tried both methods and the manual pollination method resulted in much smaller rate of successfully developing seeds (< 40%). Although the

monitoring of spontaneous pollination requires experience in estimating maturity of stigma and anthers, it can be very reliable method with a reduced hands-on time, and can produce higher numbers of seeds needed for dissections.

The difficulty of tissue dissection changes over the time of seed development. The most challenging is the isolation of tissues from the youngest (0 to 8 DAP) seeds, which are minute and easy to damage due to their soft texture and liquid character (endosperm). Therefore, fine tools are needed. Using the presented protocol, we were able to manually isolate seed maternal tissues and endosperm from 4 DAP or older seeds in sufficient amount and quality for complex assays. Dissection of embryo before 8 DAP was problematic and we were not able to collect sufficient amount of tissue for downstream analyses. We envisage that further improvements could be achieved with a micromanipulator. An alternative method could be tissue sectioning followed by a laser microdissection²⁸. Although this method offers great resolution it usually brings a very small amount of material that needs to be extensively PCR amplified before analysis. This may introduce certain biases or redundancy. The presence of cell walls in plant tissues prevents separation of intact cells and their simple isolation by fluorescence-activated cell sorting (FACs) as performed for animal or fungal cell cultures. In plants, fluorescence-activated nuclei sorting (FANs) based on the nuclei C-values is feasible²⁹. Although FANs is highly sensitive method, it represents only part of cellular information (e.g., cytoplasmic RNA and proteins are largely missing), generates small amounts of sample, and requires highly advanced instruments. It has to be emphasized that the protocol presented here provides relatively large amounts of material and does not involve PCR-based amplification before library preparation. Dissection of tissues older than 8 DAP with a cellularized endosperm is considerably easier,

but drying of seed parts at later stages may decrease tissue separability. A simple solution is moisturizing the tissues with a physiological buffer.

An important factor affecting all downstream analyses is the purity of extracted tissues. In highly sensitive experiments such as RNA-sequencing, tissue contamination results in decreased resolution and false information. Therefore, we implemented a flow cytometry-based purity control step in the protocol that is based on different ploidies of seed tissues. Distribution of nuclei within the tissue is not always homogenous and some parts can be absent of nuclei while other part of the same tissue may still contain nuclei (e.g., central starchy endosperm and aleurone layer, respectively). We were able to detect nuclei in each completely dissected tissue in the selected developmental window. Based on their triploid nature, the endosperm tissues can be easily distinguished from the diploid seed maternal tissues and embryo. The most common type of contamination observed in the samples was between seed maternal tissues and endosperm, possibly originating from the nucellar projection. However, when the sample contained less than 5% of different ploidy level, this could not be visible as a separate ploidy peak. It is also obvious that the purity control cannot distinguish contamination between seed maternal tissues and embryo. However, this type of contamination is less likely as both tissues do not adhere to each other and can be easily manually separated. In addition, to reduce tissue contamination after dissections, we applied multiple washes using a physiological buffer.

Since RNA can be easily destroyed by the activity of endogenous RNases, the quality of dissected tissue was determined based on RNA degradation level. As an index of quality we used RIN calculated automatically by a

bioanalyzer. Degradation is also a natural process caused by aging and together with unproper handling during dissection is the main factor affecting the quality of dissected tissues. This limitation did not allow isolation of good quality RNA from seed maternal tissues older than 24 DAP.

The isolated material is suitable for various types of downstream analyses including isolation of nucleic acids, proteins and other cellular compounds. We have already successfully used the tissues to isolate RNA and perform RNA-sequencing experiments. This is significant experimental improvement because until now, barley seed transcriptomic studies were done using either the mixture of embryo and endosperm³⁰ or even whole seeds³¹. Application of the optimized strategy will greatly increase the resolution and specificity of the RNA-sequencing data as also demonstrated on the expression values of several tissue-specific genes (**Figure 5**).

In conclusion, this protocol provides means for detailed analysis of individual seed tissues. This will help unravelling the mechanisms controlling seed development in barley and other cereals.

Disclosures

The authors have nothing to disclose.

Acknowledgments

We thank Dr. Jan Vrána and Dr. Mahmoud Said for the maintenance of flow cytometers, Eva Jahnová for preparation of buffers Marie Seifertová for list of materials and Zdenka Bursová for plant care. This work was supported primarily from the Czech Science Foundation grant 18-12197S. A.P. was further supported by the J. E. Purkyně Fellowship from the Czech Academy of Sciences and the ERDF project

"Plants as a tool for sustainable global development" (No. CZ.02.1.01/0.0/0.0/16_019/0000827).

References

1. Sreenivasulu, N. et al. Barley grain development: Toward an integrative view. *International Review of Cell and Molecular Biology*. **281** (C), 49-89 (2010).
2. FAO. Food and Agriculture Organization of the United Nations Cereal supply and demand brief. <http://www.fao.org> (2019).
3. Baik, B.K., Ullrich, S.E. Barley for food: Characteristics, improvement, and renewed interest. *Journal of Cereal Science*. **48** (2), 233-242 (2008).
4. Langridge, P. Economic and Academic Importance of Barley. In *The Barley Genome*. Edited by Stein, N., Muehlbauer, G. J., 1-10, Springer Cham. Switzerland, AG (2018).
5. Mascher, M. et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature*. **544** (7651), 427-433 (2017).
6. Mascher, M. Pseudomolecules and annotation of the second version of the reference genome sequence assembly of barley cv. *Morex [Morex V2]*. <https://edal.ipk-gatersleben.de> (2019).
7. Rapazote-Flores, P. et al. BaRTv1.0: An improved barley reference transcript dataset to determine accurate changes in the barley transcriptome using RNA-seq. *BMC Genomics*. **20** (1), 968 (2019).
8. Molnár-Láng, M., Linc, G., Szakács, É. Wheat-barley hybridization: The last 40 years. *Euphytica*. **195** (3), 315-329 (2014).

9. Waddington, S.R., Cartwright, P.M., Wall, P.C. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany*. **51** (1), 119-130 (1983).
10. Sabelli, P.A., Larkins, B.A. The development of endosperm in grasses. *Plant Physiology*. **149** (1), 14-26 (2009).
11. Dante, R.A., Larkins, B.A., Sabelli, P.A. Cell cycle control and seed development. *Frontiers in Plant Science*. **5** (September), 1-14 (2014).
12. Rodríguez, M. V., Barrero, J.M., Corbineau, F., Gubler, F., Benech-Arnold, R.L. Dormancy in cereals (not too much, not so little): About the mechanisms behind this trait. *Seed Science Research*. **25** (2), 99-119 (2015).
13. Sreenivasulu, N. et al. Gene expression patterns reveal tissue-specific signaling networks controlling programmed cell death and ABA-regulated maturation in developing barley seeds. *Plant Journal*. **47** (2), 310-327 (2006).
14. Olsen, O.A. Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell*. **16** (suppl_1), S214-S227 (2004).
15. Thiel, J. et al. Different hormonal regulation of cellular differentiation and function in nucellar projection and endosperm transfer cells: A microdissection-based transcriptome study of young barley grains. *Plant Physiology*. **148** (3), 1436-1452 (2008).
16. Weschke, W. et al. Sucrose transport into barley seeds: Molecular characterization of two transporters and implications for seed development and starch accumulation. *Plant Journal*. **21** (5), 455-467 (2000).
17. Staszak, A.M., Rewers, M., Sliwinska, E., Klupczyńska, E.A., Pawłowski, T.A. DNA synthesis pattern, proteome, and ABA and GA signalling in developing seeds of Norway maple (*Acer platanoides*). *Functional Plant Biology*. **46** (2), 152-164 (2019).
18. Otto, F. Chapter 11 DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. *Methods in Cell Biology*. **33** (C), 105-110 (1990).
19. Fisher Scientific, T. *Procedural guidelines*. (2020).
20. *Spectrum TM Plant Total RNA Kit*. (2020).
21. Nowicka, A., et al. Dynamics of endoreduplication in developing barley seeds. *Journal of Experimental Botany*. eraa453. (2020).
22. Chen, J. et al. Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiology*. **166** (1), 252-264 (2014).
23. Zhang, S., Laurie, A.E., Ae, W., Meng, L., Lemaux, P.G. Similarity of expression patterns of knotted1 and ZmLEC1 during somatic and zygotic embryogenesis in maize (*Zea mays* L.). *Planta*. **215** (2), 191-194 (2002).
24. Doll, N.M. et al. Transcriptomics at maize embryo/endosperm interfaces identifies a transcriptionally distinct endosperm subdomain adjacent to the embryo scutellum. *Plant Cell*. **32** (4), 833-852 (2020).
25. Yi, F. et al. High temporal-resolution transcriptome landscape of early maize seed development. *Plant Cell*. **31** (5), 974-992 (2019).
26. Gómez, E. et al. The maize transcription factor myb-related protein-1 is a key regulator of the differentiation of transfer cells. *Plant Cell*. **21** (7), 2022-2035 (2009).

27. Bewley, J.D., Black, M., Halmer, P. *The encyclopaedia of seeds: Science, technology and uses*. CABI. Wallingford, UK. (2006).
28. Liew, L.C. et al. Temporal tissue-specific regulation of transcriptomes during barley (*Hordeum vulgare*) seed germination. *Plant Journal*. **101** (3), 700-715 (2020).
29. Pirrello, J. et al. Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. *Plant Journal*. **93** (2), 387-398 (2018).
30. Sreenivasulu, N. et al. Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *Plant Journal*. **37** (4), 539-553 (2004).
31. Bian, J. et al. Transcriptional dynamics of grain development in barley (*Hordeum vulgare* L.). *International Journal of Molecular Sciences*. **20** (4), 962 (2019).

Appendix II

Transcriptome landscape of developing barley seeds

Kovacik, M., Nowicka, A., Zwyrtková, J., Strejčková, B., Vardanega, I., Esteban, E., Pasha, A., Kaduchová, K., Krautsova, M., Červenková, M., Šafář, J., Provar, N., J., Simon, R., and Pecinka, A.

The Plant Cell

in press

IF (2023): 12.085

1 **Large-Scale Biology Article**

2

3 **The transcriptome landscape of developing barley seeds**

4

5 Martin Kovacik^{1,2}, Anna Nowicka^{1,3}, Jana Zwyrtková¹, Beáta Strejčková¹, Isaia Vardanega⁴, Eddi
6 Esteban⁵, Asher Pasha⁵, Kateřina Kaduchová¹, Maryna Krautsova¹, Marie Červenková¹, Jan
7 Šafář¹, Nicholas J. Provart⁵, Rüdiger Simon⁴ & Ales Pecinka^{1,2*}

8

9 ¹ Institute of Experimental Botany, Czech Acad Sci, Centre of Plant Structural and Functional
10 Genomics, Šlechtitelů 31, 779 00 Olomouc, Czech Republic.

11 ² Department of Cell Biology and Genetics, Faculty of Science, Palacký University, Šlechtitelů 27,
12 779 00 Olomouc, Czech Republic.

13 ³ Franciszek Górski Institute of Plant Physiology Polish Academy of Sciences, Niezapominajek
14 21, 30 239 Kraków, Poland.

15 ⁴ Institute for Developmental Genetics, Heinrich-Heine-University, Universitätsstraße 1,
16 Düsseldorf 40225, Germany.

17 ⁵ Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and
18 Function, University of Toronto, 25 Willcocks St., Toronto, ON M5S 3B2, Canada.

19

20 **Running title:** Transcriptome atlas of developing barley seeds

21

22 The author responsible for the distribution of materials integral to the findings presented in this
23 article in accordance with the policy described in the Instructions for Authors
24 (<https://academic.oup.com/plcell/pages/General-Instructions>) is: Ales Pecinka
25 (pecinka@ueb.cas.cz).

26

27 *Corresponding author: Ales Pecinka (pecinka@ueb.cas.cz)

28

29 Keywords: seed, endosperm, embryo, transcriptomics, H3K27me3, barley, genomic imprinting

30

31 **Abstract**

32 Cereal grains are an important source of food and feed. To provide comprehensive spatiotemporal
33 information about biological processes in developing seeds of cultivated barley, we performed a
34 transcriptomic study of the embryo, endosperm, and seed maternal tissues collected from grains
35 4 to 32 days after pollination. Weighted gene co-expression network and motif enrichment
36 analyses pointed out specific groups of genes and transcription factors potentially regulating
37 barley seed tissue development. We defined a set of tissue-specific marker genes and families of
38 transcription factors for functional studies of the pathways controlling barley grain development.
39 Assessment of selected groups of chromatin regulators revealed that epigenetic processes are
40 highly dynamic and likely to play a major role during barley endosperm development. Repressive
41 modification H3K27me3 is globally reduced in endosperm tissues and at specific developmental
42 and storage compound genes. Altogether, this atlas uncovers the complexity of the developmental
43 regulation of gene expression in developing barley grains.

44
45 **Introduction**

46 Seeds are a crucial stage in the life cycle of many plants that allow the survival of long periods of
47 unfavorable conditions and colonization of new sites. High nutritive value makes seeds prime
48 targets of plant breeding and cereal grains are one of the most valuable agronomic products
49 (Carena, 2009). Cultivated barley (*Hordeum vulgare* L. subsp. *vulgare*) is the fourth most important
50 cereal worldwide and it is used as feed (70%), for the production of malt (21%), and as food (9%)
51 (Food and Agriculture Organization of the United Nations, 2023). Barley grains have three main
52 compartments: embryo, endosperm, and seed maternal tissues (SMTs), each consisting of
53 different parts and cell types (Gubatz et al., 2007). The diploid embryo originates from a fusion of
54 the egg cell nucleus and sperm cell nucleus. A mature embryo consists of the embryonic axis
55 (coleoptile, plumule, shoot apical meristem, radicle, coleorhiza) and the scutellum. The triploid
56 endosperm derives from a fusion of the diploid central cell nucleus and the second sperm cell
57 nucleus. After an initial unicellular multinucleate coenocyte stage, cereal endosperm cellularizes
58 and differentiates into the central starchy endosperm (CSE) that serves as the main storage of
59 complex sugars, the aleurone layer (AL), the embryo-surrounding region (ESR), and the
60 endosperm transfer cells (ETCs) (Olsen, 2001). Finally, an outer layer of SMTs, consisting of a
61 pericarp and two layers of seed coat, maintains stable conditions, transports assimilates, protects
62 the embryo and endosperm, and serves as the first source of starch during early grain
63 development (Weschke et al., 2003; Radchuk et al., 2009). The molecular mechanisms governing
64 the development of individual seed tissues remain little understood in barley.

65 Spatiotemporal analysis of gene expression plays a crucial role in understanding
66 developmental programs. Previous studies have explored transcriptional profiles of developing
67 barley grains and their parts. An early study using expressed sequence tags revealed
68 transcriptional reprogramming at 0 to 7 days after flowering in seed maternal and filial tissues and
69 early to late whole caryopses (Zhang et al., 2004). Genes encoding many protein- and lipid-
70 activating enzymes were up-regulated while genes coding for reactive oxygen species (ROS)-
71 scavenging enzymes were down-regulated in SMTs, indicating mobilization of storage compounds
72 and programmed cell death (PCD), respectively. The filial tissues contained highly expressed
73 genes coding for factors involved in cell growth and cell wall biosynthesis. Expression microarray-
74 based barley genome-wide transcriptomic studies used samples from dissected embryos and
75 endosperm 16 to 25 days after pollination (DAP) and whole caryopses 5 to 16 DAP (Druka et al.,
76 2006; Sreenivasulu et al., 2008). This revealed dynamic changes in gene expression related to
77 metabolic and hormonal pathways during grain development and germination. RNA sequencing
78 (RNA-seq) of whole caryopses explored the extent of RNA editing in barley grain development
79 (Bian et al., 2019). Most recently, a spatially resolved cellular map for germinating barley seeds
80 was provided and showed significant enrichment of different processes in specific tissues (Peirats-
81 Llobet et al., 2023).

82 Here, we performed a comprehensive transcriptome profiling of barley seed tissues at
83 different DAP timepoints with an aim to create an atlas of gene expression and thus provide
84 detailed information on the temporal and spatial distribution of the key molecular processes acting
85 during barley grain development. Our focus on nucleus-driven processes will be fundamental for
86 future functional studies of the role of the epigenome in barley grain development.

87

88 **Results**

89 **Generating tissue-specific transcriptome profiles of developing barley seeds**

90 To identify transcriptional signatures of the major tissues of developing barley seeds, we
91 performed RNA-seq of the manually dissected embryo, endosperm, and SMTs of barley cultivar
92 (cv.) Morex at 4, 8, 16, 24, and 32 DAP (Figs. 1, A and B). Sample hierarchical clustering revealed
93 a strict grouping of the tissues and time points, except for the 4 DAP endosperm that clustered
94 separately, indicating its unique transcriptome (Supplemental Fig. S1). This pattern was further
95 corroborated by principal component analysis (PCA), which revealed a predominant grouping by
96 DAP (PC1) and tissue (PC2), explaining 70% of the total variability (Fig. 1C). The 4 and 8 DAP
97 endosperm samples were notably distant from the later time points, indicating a massive
98 transcriptional reprogramming during endosperm proliferation and cellularization. The closer

99 distance between the 24 and 32 DAP samples for both embryo and endosperm suggested
100 relatively minor transcriptional changes toward the end of grain development. No major changes
101 in clustering were found by inspecting PC3 (18% of the variability; Supplemental Fig. S2 A and
102 B). When compared with the transcriptomes of eight different tissues (The International Barley
103 Genome Sequencing Consortium, 2012), the overall distribution was defined by the seed samples.
104 The vegetative tissues (root, shoot, nodule, inflorescences) clustered together with the
105 germinating embryo (Fig. 1D and S2 C and D). On the contrary, caryopsis 5 and 15 days after
106 anthesis grouped with the 4 and 16 DAP endosperm, respectively, suggesting that the endosperm
107 tissues dominate the caryopsis transcriptome.

108 To allow easy visualization of the transcriptomic data in a user-friendly format, we
109 integrated our dataset into the Barley ePlant on the Bio-Analytic Resource for Plant Biology (BAR)
110 (Fig. 1E; https://bar.utoronto.ca/eplant_barley/). The expression levels for individual genes (shown
111 as transcripts per million, TPM; Supplemental Data Set S1) revealed striking differences between
112 the tissues. The endosperm median expression was 0.96 – 1.35 TPM compared to 1.80 – 2.03
113 TPM in the embryo and 1.81 – 2.11 TPM in SMTs (Fig. 1F). A significant portion of genes with low
114 expression (TPM 0 – 1; n = 31,571) were found in different parts of the seeds, specifically 67.8%
115 in the endosperm (n = 21,400), 50.7% in the embryo (n = 16,009), and 55.0% in SMT (n = 17,364)
116 (Fig. 1G). Despite having a 2-fold lower median expression, endosperm contained the majority of
117 the highest expressed genes (Fig. 1H). These genes were significantly enriched in the Gene
118 Ontology (GO) molecular functions of negative regulation of proteolysis, defense response,
119 development, response to wounding, lipid transport, and cell wall macromolecule catabolic
120 processes (Supplemental Table S1).

121
122 **Grain development is associated with extensive transcriptional changes**
123 Spatial and temporal changes in the seed transcriptome were assessed by plotting differentially
124 expressed genes (DEGs) at one or more time points for each tissue (Fig. 2A). During embryo
125 development, the major transcriptional changes occurred from 8 to 24 DAP. Most DEGs were
126 detected from 8 to 16 DAP (n = 8,952) and 16 to 24 DAP (n = 10,340) (FDR-adjusted $P < 0.05$;
127 Fig. 2A and S4A and Supplemental Data Set S2). The major changes in embryo occurred between
128 8 to 16 DAP (8,952 DEGs) and 16 to 24 DAP (10,340 DEGs), while in SMTs more DEGs were
129 observed from 4 to 8 DAP (5,604 DEGs) and 8 to 16 DAP (8,264 DEGs). The endosperm showed
130 massive transcriptional changes, including a subset of unique genes from 4 to 8 DAP (15,990
131 DEGs, 9,869 unique). The number of endosperm DEGs gradually decreased toward 32 DAP.
132 There was a trend similar to embryos also in SMTs, with more DEGs between the early and middle

133 time points (5,604 between 4 and 8 DAP and 8,264 between 8 and 16 DAP; FDR-adjusted $P <$
134 0.05; Fig. 2A and S4C and Supplemental Data Set S2) and less at the late stage of grain
135 development (3,953 DEGs at 16 to 24 DAP).

136 We associated transitions between individual DAPs with over-/under-represented Gene
137 Ontology terms (Fig. 2B and Supplemental Table S2). Embryos at 8 DAP showed enrichment in
138 terms linked to cell division (GO terms cell cycle, DNA replication, histone, and chromatin
139 modification) and their reduction before 16 DAP. The 16 DAP embryo transcriptome indicated
140 intense ribosome biogenesis and cell wall synthesis. However, these enriched GO terms are
141 typical for many actively growing tissues and were also shared to a large extent with endosperm
142 and SMTs (Supplemental Table S2 and S3). The GO term “mRNA splicing via spliceosome” was
143 found to be significantly enriched at 16 DAP and even moreso after 24 DAP (Fig. 2B). This is
144 consistent with the accumulation of mRNAs in developing seeds and subsequent translation
145 during germination (Sano et al., 2020). In endosperm, many tissue-specific enriched GO terms
146 were associated with storage compounds (GO terms: lipid-, starch-, glucan-, glycoprotein-
147 biosynthesis and metabolism; Fig. 2B, Supplemental Table S2 and S3), which underlies the role
148 of endosperm as the main nutritive tissue in grains. Furthermore, there was a tissue-specific
149 expression of genes coding for DNA methylation factors from 4 to 8 DAP. The most prominent
150 enriched GO terms in SMTs included photosynthesis and cell wall development (both peaking at
151 8 DAP), and a wave of expression from fungi and abiotic stress-responsive factors (Fig 2B,
152 Supplemental Table S2 and S3). The unique enriched GO term of SMTs was the upregulation of
153 isoprenoid biosynthesis from 4 to 16 DAP. Accumulation of isoprenoids is important for seed
154 nutritional and physiological quality (Vishal and Kumar, 2018).

155 Using k-means clustering, we defined molecular marker genes for individual tissues and
156 DAP timepoints (Fig. 2C and S5 and Supplemental Data Set S3). The embryo ($n = 17,214$),
157 endosperm ($n = 21,889$), and SMTs ($n = 15,034$) DEGs were divided into 12, 13, and 14 co-
158 expression clusters, respectively, where four to five clusters showed expression peaking at single
159 consecutive experimental points. We searched the clusters for tissue-specific genes defined as
160 having less than 5% TPM in other seed tissues (Fig. 2C and S5 – red numbers). Most of the stage-
161 and tissue-specific genes in embryo and endosperm were found in 4 and 8 DAP-specific clusters.

162 163 **Specific promoter motifs are associated with transcriptional regulation of seed** 164 **development**

165 To identify expression correlations among genes, we performed weighted gene co-expression
166 network analysis (WGCNA) for each tissue. The identified WGCNA modules were organized

167 temporally according to the early, middle, and late stages of development (Fig. 3A and S6 and
168 Supplemental Data Set S4). The majority of the genes were found to accumulate either at early
169 (EMB_M2, n = 1,883; END_M2/M7/M6, n = 1,962; and SMT_M1, n = 1,572) or late (EMB_M1, n
170 = 1,934; END_M1/M2, n = 1,838; and SMT_M3, n = 2,497) modules of individual tissues,
171 suggesting roles for these genes in transcriptional reprogramming.

172 Next, we analyzed transcription factors (TFs) and regulatory motifs important during barley
173 grain development by performing promoter motif enrichment analysis. The regions -1,500 to -1 bp
174 from the transcription start site of barley genes included in WGCNA modules were analyzed for
175 the presence of known TF binding motifs using the *Arabidopsis thaliana* HOMER database (Heinz
176 et al., 2010). This resulted in collections of significantly enriched motifs (p-val < 0.05; Supplemental
177 Data Set S5) that grouped into clusters based on similarity (Supplemental Table S4 and S5). The
178 proportion of major motif clusters (MCs) in each collection was quantified (Fig. 3B). The MC2 and
179 MC9 were enriched across many WGCNA modules of all tissues and their consensus motifs
180 corresponded to the G-box (CACGTG) or Prolamin-box (P-box; AWAAAG), respectively (Fig. 3B).
181 The common TF families predicted to bind MC2 DNA motifs were the basic/helix-loop-helix (bHLH)
182 family and basic region/leucine zipper (bZIP) family members known to bind the G-box and
183 TKACGT motif variants (Supplemental Fig. S7A). To test whether the predicted TFs in those
184 families might be further supported by our data, we identified barley orthologs of Arabidopsis TFs
185 (Supplemental Table S6), and retrieved their expression profiles from our data (Supplemental Fig.
186 S7B and Supplemental Table S6). The expression profiles of many TFs appeared early or late
187 during grain development, e.g. the barley homolog of *ABA INSENSITIVE 5 (ABI5)* exhibited an
188 expression profile peaking later in embryo and endosperm development. In Arabidopsis, ABI5 is
189 involved in signaling during seed maturation and regulates a subset of *LATE EMBRYOGENESIS*
190 *ABUNDANT (LEA)* genes (Goyal et al., 2005). Many *LEA* genes were highly expressed in the late
191 embryo module, and some were also found to be expressed in the late endosperm module
192 (Supplemental Data Set S4). The MC9 was enriched particularly with genes expressed during
193 middle embryo and endosperm development and these contained the characteristic Dof TF family
194 binding P-box motif (A/T)AAAG (Fig. 3, B and C and Supplemental Table S7). Dof type TFs
195 PROLAMIN BINDING FACTOR (PBF) and DOF1 have been shown to regulate seed storage
196 protein synthesis in barley and maize, respectively (Yanagisawa and Sheen, 1998; Mena et al.,
197 1998). Barley PBF presumably activates the expression of storage compounds in endosperm by
198 binding to the P-box motif present in the promoters of *Hordein* genes (Mena et al., 1998). Our *in*
199 *situ* hybridization data extended the knowledge about spatial and temporal expression patterns of
200 this gene in barley endosperm (Fig. 3, D, E and S8). The *PBF* transcript was detected around the

201 embryonic pole starting at 8 DAP and its pattern of expression expanded to the endosperm
202 periphery at the ventral side towards 16 DAP. Observed expression in embryos at 16 DAP further
203 confirms its suggested role in embryo development (Cook et al., 2018). As the P-box motif was
204 enriched in the middle endosperm module, we investigated the genes most strongly expressed in
205 the END_M4 module. We found increased transcript of many genes encoding enzymes involved
206 in oligo- and poly-saccharide synthesis (sucrose synthase, alpha-glucan-branching enzymes, and
207 starch synthase) and major endosperm proteins, represented by low molecular weight glutenin
208 subunit (Supplemental Data Set S4). To further extend the understanding of transcriptional
209 regulation of seed development, we performed *de novo* motif enrichment analysis within WGCNA
210 modules. Collections of *de novo* motifs from each WGCNA module were curated for false positive,
211 low complexity, and simple repeat motifs, resulting in a list of 168 motifs identified in embryo,
212 endosperm and SMTs (Supplemental Data Set S6). The major portion of *de novo* motifs were
213 observed in early and late modules, providing the first insight into novel regulatory motifs driving
214 the process of seed development.

215

216 **The barley endosperm differentiation is initiated before cellularization**

217 Several gene markers for individual endosperm domains have been identified across cereals
218 (Bate et al., 2004; Magnard et al., 2003; Opsahl-Ferstad et al., 1997; Gómez et al., 2002; Hueros
219 et al., 1999; Kalla et al., 1994), including a few markers for ETCs and AL (Leah et al., 1991; Kalla
220 et al., 1994; Doan et al., 1996; Hertig et al., 2020, 2023). Here we aimed to extend the list of
221 markers for barley by performing a comparative analysis of 12 endosperm marker genes
222 described in maize and rice. By reciprocal BLAST, we identified in total 29 homologs in barley
223 (Supplemental Table S8). Interestingly, these known marker genes typically reached a maximum
224 expression at younger (4 or 8 DAP) stages (Fig. 4 and S9 and Supplemental Table S8), suggesting
225 a biased selection. Nevertheless, several candidates might be used for identifying later (16 or 24
226 DAP) stages of endosperm development.

227 The timing of starch accumulation in endosperm differs across cereals. It begins around 6
228 DAP in barley and only later, around 10 DAP, in maize (Bennet et al., 1975; Charlton et al., 1995).
229 The CSE marker *STARCH BRANCHING ENZYME 1* has three gene copies in barley (*SBE1A-C*),
230 where *SBE1A* and *1B* were first detected at 8 DAP and then decreased their expression, while
231 *SBE1C* started low at 8 DAP and dramatically increased its expression (2,745 TPM) at 16 and 24
232 DAP (Fig. 4A). Rice *SUCROSE NON-FERMENTING-1 RELATED PROTEIN KINASE 1b*
233 (*SnRK1b*) has four homologs in barley. Expression of *SnRK1b* in rice was correlated with the
234 emergence of starch granules (Kanegae et al., 2005). However specific expression in barley

235 endosperm was observed only for the *SnRK1bA* (Fig. S9). Almost no expression was observed
236 for the homolog of maize *SHRUNKEN-1 (SH1)* encoding a sucrose synthase (Fig. S9) (Chourey
237 and Nelson, 1976). As to the aleurone, the barley *AL9A* and *AL9B* (homologs of maize
238 *ALEURONE9*) were expressed exclusively 8 DAP. Other AL markers, such as barley homologs
239 of rice subunit *B1A* of the *NUCLEAR FACTOR Y (NF-YB1A)* and of maize *COLORED*
240 *ALEURONE1 (C1A)*, reached their expression maxima at 16 and 24 DAP, respectively (Fig. 4, A
241 and B and S9). We further analyzed the expression of *NF-YB1A* at 4 to 16 DAP using RNA *in situ*
242 hybridization and found strong signals in the AL (Fig. 4C and S9). Findings in maize suggest that
243 AL differentiates from the outer layers of endosperm cells between 6 and 10 DAP soon after
244 alveolation and the first periclinal division of the cellularized endosperm (Olsen, 2001).
245 Surprisingly, *NF-YB1A* was also expressed in 4 DAP endosperm, where the signals accumulated
246 at the embryonic pole of the seed (Fig. 4C and S9). This provided new evidence supporting the
247 recent observation that the aleurone identity is already defined in specific endosperm nuclei before
248 cellularization begins (Hertig et al., 2023) and suggests that barley endosperm differentiation
249 starts at the embryonic pole. The homolog of maize *OUTER CELL LAYER 4 (OCL4)*, encoding a
250 homeo domain-leucine zipper IV transcription factor expressed in AL, did not show endosperm-
251 specific expression in barley (Sosso et al., 2010). ETCs markers, such as antimicrobial protein-
252 coding genes *EMBRYO SAC/BASAL ENDOSPERM TRANSFER LAYER/EMBRYO*
253 *SURROUNDING REGION 1* paralogs *A* to *D* and 2 (*EBE1A-D* and *EBE2*) and TF *MYB-RELATED*
254 *PROTEIN 1* paralogs *A* to *C* (*MRP1A-C*), were expressed at 4 DAP and reached expression
255 maxima at 8 DAP (Fig. 4A and S9). This indicates that transfer cell fate specification occurs during
256 a narrow temporal window of coenocyte development in barley, which is similar to the situation
257 described in maize (Costa et al., 2003; Hertig et al., 2023). ESR markers, such as sugar
258 management coding gene *INVERTASE INHIBITOR 1A (INVINH1A)*, and defensin-type gene
259 *EMBRYO SURROUNDING REGION 6 (ESR6)*, reached their peak of expression at 4 DAP (Fig.
260 4A). Another class of ESR markers was represented by genes coding for putative antifungal
261 proteins *ANDROGENIC EMBRYO 1A* to *B (ANE1A-B)* and *ANDROGENIC EMBRYO 3 A*
262 (*ANE3A*) that were highly expressed 4 and 8 DAP, respectively (Fig. 4A and S9). This corresponds
263 to findings in maize, where ESR cells differentiate upon completion of the endosperm
264 cellularization phase (Kiesselbach and Walker, 1952; Serna et al., 2001).

265 Although the endosperm markers from various cereals also exhibited tissue-specificity in
266 barley, their temporal expression dynamics differed in several cases. The AL, ETCs, and ESR
267 markers were already expressed at 4 and 8 DAPs. It was previously reported that the specificity

268 of AL and ETCs is already defined in their founder nuclei before the cellularization (Hertig et al.,
269 2023). Here we extend this observation to the ESR.

270

271 **Multiple pathways are controlled by H3K27me3 in barley endosperm**

272 Our GO term enrichment analysis indicated chromatin-based regulation of barley seed
273 development (Fig. 2B, Supplemental Table S2 and S3). To explore this further, we studied the
274 expression patterns of barley orthologs of histones and POLYCOMB REPRESSIVE COMPLEX 2
275 (PRC2) (Supplemental Table S9).

276 We found in total 188 barley histone genes corresponding to all canonical forms and
277 common plant variants (Supplemental Fig. S10 and Supplemental Table S10). In total, 152 out of
278 175 expressed histones (87%, TPM \geq 1) were part of the endosperm k-means CL2 (including 30
279 copies of *H3*; 31 copies of *H2A.W*; and one *H2A.Z*), with predominant expression levels during
280 the coenocyte stage (Fig. 2C and Supplemental Data Set S3). The peak of expression in CL2
281 coincided with the period of DNA replication and nuclei division during coenocyte development.
282 After cellularization, the expression of histone genes decreased, and only a few copies, mostly
283 paralogs of non-canonical variants, remained expressed (Supplemental Fig. S10 and
284 Supplemental Table S10). The initial stages of embryo and SMTs were also marked by the peak
285 of histone expression, but the overall maxima were lower compared to endosperm (Supplemental
286 Fig. S10). A single *H2B* copy showed a prominent expression in 16 DAP and later embryo stages.
287 Closer inspection of this copy revealed that it is a recently described seed-specific histone *H2B.S*
288 variant (Jiang et al., 2020). The *H2B.S* was not expressed in endosperm and SMTs. Altogether,
289 this indicates dynamic epigenetic control and changes in nucleosome composition during
290 endosperm and embryo development.

291 The PRC2 complex installs tri-methylation of lysine 27 at histone H3 (H3K27me3). This
292 modification transcriptionally represses its target genes and thus contributes to developmental
293 transitions, including endosperm cellularization (Mozgova and Hennig, 2015). Arabidopsis PRC2
294 consists of four evolutionary conserved subunits FERTILIZATION INDEPENDENT ENDOSPERM
295 (FIE), MULTICOPY SUPPRESSOR OF IRA 1 (MSI1), EMBRYONIC FLOWER 2 (EMF2), and the
296 catalytic subunit represented by the three homologs SET DOMAIN GROUP 1, 5 and 10 (SDG1,
297 5 and 10) alias CURLY LEAF (CLF), MEDEA (MEA) and SWINGER (SWN), respectively. All four
298 core PRC2 subunits are present in cereals (Strejčková et al., 2020). We found that the barley
299 genome lacks an *SDG5/MEA* homolog and contains single copies of *FIE* and *SDG10/SWN* and
300 multiple copies of *MSI1*, *EMF2*, and *SDG1/CLF*. At least one copy of each PRC2 subunit was well
301 expressed (TPM > 10) throughout the whole embryo and SMTs development (Fig. 5A and

302 Supplemental Table S10). The pattern in the endosperm was more complex. Genes coding for
303 *EMF2A/B*, *SDG1A/B*, and *SDG10* were silent or only weakly expressed at 4 DAP, (Ramirez-Parra
304 and Gutierrez, 2007) suggesting no or little function of the PRC2 complex in barley early
305 endosperm development. From 8 DAP onwards, *EMF2A*, *FIE*, *MSI1B*, and *SDG10* maintained
306 moderate expression levels, while both *SDG1* copies remained silent or weakly expressed.

307 To estimate whether these expression patterns have any impact on the H3K27me3 levels,
308 we performed immunostaining on 8 DAP and 24 DAP endosperm and 24 DAP embryo nuclei
309 isolated by flow-cytometry-based on their C-values (Fig. 5B). There was an intense H3K27me3
310 signal at the telomeric poles of all analyzed types of embryo nuclei. Such signal distribution is
311 caused by the accumulation of H3K27me3 at the gene-rich chromosome ends and the presence
312 of Rab1 chromosome organization in barley (Nowicka et al., 2023; Baker et al., 2015). On the
313 contrary, the H3K27me3 signals were weaker at 8 DAP endosperm nuclei and almost absent at
314 24 DAP (Fig. 5B and S11). The H3K27me3 loss was also progressive over increasing C-values,
315 suggesting a developmental stage and ploidy-dependent global loss of H3K27me3 in endosperm
316 nuclei.

317 To investigate H3K27me3 levels in endosperm at greater resolution, we performed ChIP-
318 seq on 16 DAP endosperm samples and compared the signals to H3K27me3 ChIP-seq data from
319 seedlings (Baker et al., 2015) (Supplemental Fig. S12 and S13A). The H3K27me3 signals
320 concentrated on gene-rich chromosome termini in seedlings but were strongly reduced in
321 endosperm. Although the overall trend was biased towards H3K27me3 reduction in endosperm,
322 the changes were more complex when looking at individual H3K27me3 peaks (Supplemental Fig.
323 S13B, C and D). The total number of H3K27me3 peaks was higher in the endosperm, but they
324 were generally shorter and smaller compared to seedlings. We focused on the most prominent
325 peaks (fold enrichment ≥ 5) and performed differential analysis. We found 17,194 regions with a
326 significant H3K27me3 loss and 13,845 regions with gain in endosperm relative to seedling
327 ($\log_2\text{FoldChange} < 0$ or > 0 , respectively, Benjamini-Hochberg FDR-adjusted p-value < 0.05) (Fig.
328 5D and Supplemental Data Set S7). The genomic regions that lost or gained H3K27me3 peaks
329 included 1,856 and 1,118 genes, respectively (Supplemental Data Set S8). To assess the direct
330 role of H3K27me3 in transcriptional regulation, we searched which of these genes were
331 significantly up-regulated or down-regulated ($\text{padj} < 0.05$) in at least one analyzed endosperm
332 time point (Fig. 5F and S13E, Supplemental Table S11 and S13). For 60 genes, H3K27me3 –
333 depletion in endosperm correlated with their increased expression. This included several storage
334 compound genes, such as *LOW-MOLECULAR-WEIGHT GLUTENIN SUBUNITS*, *C-HORDEIN*
335 and *OMEGA SECALIN* (Fig. 5F, Supplemental Table S11). Interestingly, other such genes

336 included inhibitors of sugar and protein degradation and were expressed in a seed stage-
337 dependent manner (Supplemental Table S11). The INVERTASE INHIBITORS block hydrolysis of
338 sucrose to fructose and glucose, and were highly expressed at 8 DAP but not during subsequent
339 stages, possibly allowing feeding of the growing embryo or endosperm. The inhibitor of protein
340 degradation *SERPIN* was expressed from 16 DAP (Supplemental Table S11). This suggests that
341 the accumulation of storage carbohydrates and proteins is accompanied by simultaneous
342 inhibition of their degradation in the endosperm tissues in an H3K27me3-dependent manner. This
343 was supported also by the enrichment of the storage protein-related GO terms for these 60 genes
344 (Fig. 5G and Supplemental Table S13).

345 There were 238 genes that showed a significant decrease in expression of at least one
346 analyzed endosperm time point, coinciding with enrichment of H3K27me3 (Supplemental Fig.
347 S13E, Supplemental Table S13). This cluster was dominated by the two prominent groups. The
348 first included putative *SENESCENCE-ASSOCIATED PROTEINS* (n = 72) which might be
349 regulating tissue maturation by inhibiting specific proteases (Roberts et al., 2012). The second
350 cluster included 27 copies of the RNA Pol II subunit *MEDIATOR OF RNA POLYMERASE II*
351 *TRANSCRIPTION SUBUNIT 12 (MED12)* (Supplemental Table S12). MED12 could be linked with
352 transcriptional regulation of specific genes in barley. In Arabidopsis, MED12 contributes to the
353 regulation of flowering genes and the mutants show late-flowering phenotype (Gillmor et al., 2014).
354 Among enriched GO categories for these genes, we observed terms related to respiration and
355 generation of energy (Supplemental Fig. S13F and Supplemental Table S14).

356 This suggests that grain filling, senescence, and several other biological pathways are
357 controlled by the H3K27me3 modification in barley seeds.

358

359 **Identification of conserved imprinted genes**

360 The H3K27me3 plays an important role in the epigenetic regulation of uniparental gene expression
361 by genomic imprinting in Arabidopsis (Batista and Köhler, 2020). Only eight imprinted genes have
362 been identified in barley to date, based on the homology with imprinted genes in rice (Chen et al.,
363 2018). We took an analogous approach with a broader dataset of 155, 156, and 697 imprinted
364 genes from rice, maize, and wheat, respectively (Waters et al., 2011; Zhang et al., 2011; Luo et
365 al., 2011; Yang et al., 2018; Chen et al., 2018), and performed a comparative search for
366 evolutionarily conserved imprinted genes in barley endosperm tissues. We identified 449 barley
367 orthologs (Fig. 6A and Supplemental Table S15) with 19 genes shared across two and 2 genes
368 shared across three species (together 4.3% of all candidates). Such a low number could indicate
369 a relatively fast evolution of imprinted genes in grasses.

370 To provide experimental validation of the 21 conserved candidates, we checked their
371 expression patterns in our seed transcriptomic data and identified four main expression pattern
372 groups (Fig. 6B and Supplemental Table S16). Group 1 contained nine genes that were expressed
373 across all seed tissues. Group 2 had four genes expressed in embryo and SMTs but weakly
374 expressed or silent in the endosperm. Group 3 consisted of seven genes with expression restricted
375 to endosperm, and Group 4 contained a single gene expressed at early endosperm, 8 DAP SMTs,
376 and silent in the embryo. Next, we analyzed parent-of-origin specific expression pattern of these
377 candidates using 8 DAP endosperms of reciprocal hybrids between Morex × HOR 12560 (HS)
378 genotypes. Three genes were excluded due to the lack of diagnostic SNPs in this parental
379 combination (Fig. 6C,D, S13 and S14). Amplification of cDNA from three genes – *DEFECTIVE*
380 *ENDOSPERM (DE18)*, *PALADIN (PALD)* and *PROTEIN PHOSPHATASE HOMOLOG 48*
381 *(PRH48)* – was not successful and therefore these candidates remain unclassified. The *FRIGIDA-*
382 *ESSENTIAL 1 (FES1)* gene showed an HS genotype-specific expression. Four genes
383 *CYTOCHROME P450 REDUCTASE (CPR)*, *PROTEIN KINASE FAMILY PROTEIN (PKP)*,
384 *PROBABLE PROTEIN PHOSPHATASE 2C 27 (PP2C27)* from Group 1 were expressed from both
385 parents. Three genes were classified as potentially imprinted because they showed either
386 maternally (*MORC6B*) or paternally (*RBP* and *SECA*) biased expression. Seven of the total 18
387 tested genes (38.9%) were confirmed as imprinted in barley (Fig. 6D). Three maternally expressed
388 genes (MEGs) included *CA-RESPONSIVE PROTEIN (CARP)*, *RNA-BINDING PROTEIN (ARP1)*
389 and *PROLINE-RICH PROTEIN (PRP3)*. Wheat and rice MEG *CARP* was part of barley Group 1
390 and was expressed only weakly in endosperm compared to moderate expression in embryos.
391 *ARP1* and *PRP3* were both MEGs in rice and maize and their expression was relatively specific
392 to barley endosperm. We also confirmed four paternally expressed genes (PEGs) *REGULATION*
393 *OF NUCLEAR PRE-MRNA DOMAIN-CONTAINING PROTEIN 1B (RPRD1B)*, *DA1-RELATED 1*
394 *(DAR1)*, *AT-RICH INTERACTION DOMAIN 1B (ARID1B)* and *UBIQUITIN-SPECIFIC PROTEASE*
395 *16 (UBP16)*. The *RPRD1B* and *ARID1B* genes were found as PEGs in rice and maize. *RPRD1Bs*
396 encode epsin N-terminal homology (ENTH) domain-containing proteins with roles in endocytosis
397 and cytoskeletal regulation (De Camilli et al., 2002). Arabidopsis *ARID1* is a transcriptional
398 activator expressed in pollen development, which could be consistent with its role as PEGs in
399 barley. *DAR1* was previously found as a PEG in wheat and maize (Fig. 6B). It is related to
400 Arabidopsis *DA1*, a ubiquitin-activated peptidase that plays a role in the regulation of
401 endoreduplication, determination of plant architecture and possibly maternal control of seed
402 weight (Gu et al., 2022; Peng et al., 2015). *DA1* functions antagonistically with its direct substrate
403 *UBIQUITIN-SPECIFIC PROTEASE 15 (UBP15)* in Arabidopsis (Du et al., 2014). Interestingly, we

404 confirmed another member of the UBC family, *UBIQUITIN-SPECIFIC PROTEASE 16 (UBP16)* as
405 an evolutionary conserved PEG in maize, rice, wheat (Luo et al., 2011; Waters et al., 2011; Yang
406 et al., 2018; Chen et al., 2018) and barley (this study).

407 These findings encourage future genome-wide searches of imprinted genes in barley and
408 provide an initial set of evolutionarily conserved candidates for functional studies.

409

410 **Discussion**

411 We generated a comprehensive gene expression atlas of developing barley seeds. This resource
412 offers a higher resolution to the tissues and time points compared to previous studies (Zhang et
413 al., 2004; Druka et al., 2006; Sreenivasulu et al., 2008). The stage- and tissue-specific marker
414 genes, especially those for the later developmental stages, can serve as a basis for the follow-up
415 functional investigations of crucial players involved in barley grain development. While we focused
416 mostly on the analysis of endosperm tissues, the dataset have an equal resolution for embryos
417 that give rise to the next generation and often neglected SMTs that play a critical role in protecting
418 seeds and defining their parameters, e.g. size (Radchuk et al., 2011). It has to be emphasized
419 that the embryo, endosperm and SMTs were manually dissected and each tissue consisted of
420 multiple cell types. Hence, our data show a tissue rather than cell type-specific transcriptional
421 profiles. Manual dissection of spatially complex grain tissues can lead to potential tissue cross-
422 contamination. We routinely monitored tissue purity by flow-cytometry-based ploidy
423 measurements, and samples with signs of nuclei from other tissue were discarded. In addition,
424 multiple wash steps were included due to the same ploidy level of embryo and seed maternal
425 tissues. However, the embryo tissue was relatively solid and easy to separate and expected
426 contamination between these two tissues should be minor ($\leq 5\%$). The transcriptomic data are
427 easily accessible through the Barley ePlant under the V2 and V3 genome releases.

428 TFs play crucial roles in developmental transitions and control of major maturation events,
429 including storage reserve accumulation, chlorophyll degradation, and dormancy (Alizadeh et al.,
430 2021). A detailed description of the transcriptional regulation exists for the embryo. We analysed
431 expression of TFs facilitating whole seed development, considering both TF expression and the
432 presence of its binding motifs in promoter regions of the expressed genes. These two views often
433 correlate and complement each other. We highlighted the case of DOF transcription factors and
434 the corresponding expression of seed storage proteins. However, understanding the regulatory
435 network comprising endosperm development will require further investigation of individual genes.

436 Using information from other cereals, we defined orthologs for individual markers of
437 endosperm compartments in barley (Olsen, 2001). Most of them showed endosperm-specific

438 expression, suggesting that these markers will be useful also in barley. Many of the marker genes
439 had expression maxima at 4 or 8 DAP, suggesting an early endosperm differentiation.
440 Observations in maize suggested endosperm differentiation upon completion of endosperm
441 cellularization. However, a recent study from barley (Hertig et al., 2023) and our results already
442 exhibit the expression of the aleurone marker before cellularization. This is reminiscent of the
443 situation in Arabidopsis, where endosperm differentiation has already started at the 16-nuclei
444 syncytial stage, and when cellularization is initiated around the embryo (Brown et al., 2003).

445 Studies in Arabidopsis revealed important roles of several epigenetic processes and
446 molecular factors in embryo and endosperm development (Le et al., 2010). How much are these
447 processes conserved in cereals, including barley, remains unknown. We performed several pilot
448 experiments that defined expression patterns of two important groups of epigenetic regulators –
449 histones as the basic constituents of nucleosomes and members of the PRC2 complex, the key
450 repressor of developmentally regulated genes (Probst et al., 2020; Zhang et al., 2002). The
451 histones were expressed mostly at early stages of embryo and endosperm development, which
452 correlates with a rapid round of DNA replication and cell division. Notably different was the *H2B.S*
453 variant (Jiang et al., 2020) that was expressed in the late stages of embryo development. This
454 indicates conserved function of H2B.S across higher plants. Low expression of several PRC2
455 subunits at the early stages of barley endosperm development was surprising and suggested
456 possible dynamics in the H3K27me3. By a combination of transcriptomics and chromatin profiling,
457 we also defined a set of endosperm genes that are directly controlled by H3K27me3. These
458 include genes encoding storage components and their modifying enzymes such as *LOW-*
459 *MOLECULAR-WEIGHT GLUTENIN SUBUNITS*, *C-HORDEIN* and *OMEGA SECALIN*, *SERPIN*
460 and *INVERTASE INHIBITORS*. We hypothesize that the global reduction in H3K27me3 might not
461 be sufficient for these transcriptional changes and it is likely that other mechanisms, possibly
462 involving specific histone demethylases, are involved. However, our data point toward the pivotal
463 role of chromatin in barley grain development and suggest the need for further functional studies.

464 Finally, we laid down a fundament for future analysis of genomic imprinting in barley. A
465 comparative approach identified 4.3% (n = 21) conserved (shared by at least two species)
466 imprinted genes in maize, rice and wheat. This is a surprisingly low number that points to the
467 potential high speed of evolution of imprinted genes among cereal species. Experimental
468 validation in barley confirmed the imprinted status of almost 40% of the conserved candidates.
469 When combined with the previous study (Chen et al., 2018), we conclude that there are currently
470 15 confirmed imprinted genes in barley. Interestingly, the confirmed candidates point to the role
471 of ubiquitin-based regulation in genomic imprinting. Proteasomal degradation of specific targets

472 of imprinted genes could be a fine balancing mechanism between the maternal and paternal
473 genome contributions. However, whole genome studies will be needed to uncover the full
474 spectrum of barley imprinted genes.

475 In conclusion, our study generated valuable data for functional research on barley grain
476 development and provided numerous unique resources that will enhance the capacity of barley
477 genomic research. Altogether, this will help in understanding the role of nucleus-governed
478 processes during cereal grain development

479

480 **Materials and methods**

481 **Plant materials and growth conditions**

482 Six-rowed spring barley *Hordeum vulgare* subsp. *vulgare* cv. Morex was used throughout the
483 study. For analysis of imprinted genes, also wild barley *Hordeum vulgare* subsp. *spontaneum*
484 strain HOR 12560 (HS) was used. For plant growth, the seeds were germinated on moisturized
485 cellulose tissue paper covered with one layer of filter paper for 3 days at 25°C in the dark.
486 Germinating grains were moved into 5 × 5 cm peat pots containing a mixture of soil and sand (3/1;
487 v/v) and were grown in a climatic chamber under a long-day regime (16 h day 20°C, 8 h night
488 16°C; light intensity 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$; humidity 60%). After 10 days, the plants were transferred
489 into 15 × 15 cm pots containing the same soil and cultivated in the same conditions until flowering.
490 The number of DAPs was defined by determining the day of self-pollination as described (Kovacik
491 et al., 2020). The developing seeds were collected at 4, 8, 16, 24, and 32 DAP in at least three
492 biological replicates. Only the central row seeds from the middle of the spikelets were used.
493 Corresponding tissues were manually dissected and checked for its purity by flow-cytometric
494 ploidy measurement as described (Kovacik et al., 2020). The 4 DAP embryo was omitted due to
495 its small size. For sampling 8 DAP embryos, at least 10 embryos were pooled per biological
496 replicate. For embryo samples at later stages, five embryos were pooled per biological replicate.
497 Young (4 and 8 DAP) and older endosperm (>8 DAP) tissues were isolated from twenty or three
498 seeds, respectively, per biological replicate as described (Kovacik et al., 2020). SMTs were
499 isolated from at least five grains per biological replicate, irrespective of the stage of development.
500 SMTs at 32 DAP were excluded due to their dry nature. Dissected tissues were immediately frozen
501 in liquid nitrogen and stored at -80°C until use.

502

503 **RNA extraction, sequencing, and analysis**

504 Total RNA was isolated using an RNeasy® Plant Mini Kit (QIAGEN) and Spectrum™ Plant Total
505 RNA Kit (Sigma-Aldrich) according to the manufacturer's instructions including on-column DNase

506 I (Sigma-Aldrich) treatment. The RNA quality was checked using Bioanalyzer 2100 with RNA 6000
507 Pico Chips (all Agilent). Samples with RNA integrity number >6.8 were processed into RNA-Seq
508 mRNA libraries using NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] with poly-A selection.
509 The mRNA-enriched libraries were sequenced as single-end 100 bp RNA seq reads on a
510 HiSeq2500 instrument (Illumina). The raw reads were trimmed for adaptors by Trim Galore v.0.4.1
511 (Martin, 2011) and aligned to the *H. vulgare* cv. Morex reference genome v3 (Mascher, 2021)
512 using HiSat2 v.2.1.0 genomic mapper (Kim et al., 2019). Aligned reads were assigned to features
513 and meta-features using Subread v.1.5.2 software (Liao et al., 2013) according to the genome
514 annotation. Differential expression analysis was performed using DESeq2 v.1.24.0 package (Love
515 et al., 2014) in R v.3.6.3 software (R Core Team, 2020). A gene was declared to be significantly
516 differentially expressed if the Benjamini-Hochberg FRD-adjusted p-value was <0.05. Published
517 barley RNA-Seq Supplemental Data Set Sets were retrieved at the NCBI SRA from Bioproject
518 PRJEB3149 and analyzed as described above. The PCAs were done after applying the variance
519 stabilization transformation (Anders and Huber, 2010). Venn diagrams were drawn using the
520 eulerr v.6.1.0 package in R (Larsson, 2020).

521

522 **Clustering analyses**

523 For k-means clustering, unique DEGs from all tested combinations within the tissue were clustered
524 using the k-means algorithm in R (R Core Team, 2020). K-means clustering was performed using
525 the Hartigan-Wong algorithm with 1,000 iterations. An optimal number of clusters was determined
526 by statistical testing using a gap statistics method (Tibshirani et al., 2001). For WGCNA a gene
527 co-expression network was constructed for each tissue with the raw read counts after the *rlog*
528 transformation using the WGCNA library in R (Langfelder and Horvath, 2008, 2012). An adjacency
529 matrix was made using the soft thresholded Pearson correlation (power = 18) among 5,000 genes
530 with the most varying expression among experimental points. Hierarchical clustering was
531 performed for grouping the genes with highly similar co-expression patterns. The modules were
532 identified using the Dynamic Hybrid Tree cut algorithm (Langfelder et al., 2008) keeping the
533 minimum size of the module to 15 and DeepSplit set to 4 to produce fine clusters. Each module
534 was represented by color coding, with 12-15 modules detected depending on the tissue. The
535 expression profile of each module was summarized by module eigengene defined as its first
536 principal component. The probes that did not fit any of the main modules were placed into the
537 “unspecific” module that was removed from further analysis.

538

539 **A seed view in Barley ePlant**

540 The Barley ePlant framework (Thiel et al., 2021) was modified to accept V3 Barley gene identifiers
541 (Mascher, 2021). To create a new Seed view in the Barley ePlant, the data described in this
542 publication in TPM were databased on the Bio-Analytic Resource for Plant Biology's server
543 (Toufighi et al., 2005). An SVG image depicting the parts of the seed that were sampled in this
544 work was generated and an XML file linking specific parts of the image with database sample
545 names was manually created to configure this new view.

546
547 **GO term enrichment analyses and annotation of transcription factors**
548 GO terms provided in Morex reference genome v3 annotation (Mascher, 2021) was used to
549 perform an enrichment analysis by the topGO v.2.44.0 package in R (Alexa and Rahnenfuhrer,
550 2021). Redundant GO terms were filtered using the web-based tool REVIGO (Supek et al., 2011)
551 with default settings and general terms were filtered using size selection (Yon Rhee et al., 2008).
552 Terms with a size sufficient for robust statistical analysis ($n > 100$) and fold enrichment > 3 were
553 investigated. Transcription factors were classified into families (Supplemental Data Set S9) based
554 on the presence of specific domains according to PlantTFDB (Jin et al., 2017).

555
556 **Cis-motif identification and clustering**
557 For cis-motif identification and enrichment analysis, 1500 bp upstream sequences from the
558 predicted start codon (ATG) of all WGCNA module genes were used. The analysis was carried
559 out using the "findMotif.pl" program from the HOMER suite (Heinz et al., 2010) that performs
560 known and *de novo* motif identification and enrichment analysis with default parameters. The
561 enrichment of identified motifs was calculated respective to all 1500 bp background sequences.
562 Collections of identified motifs in each WGCNA module were post-filtered for plant motifs and
563 clustered using the RSAT (Castro-Mondragon et al., 2017) with default parameters.

564
565 **RNA *in situ* hybridization**
566 Barley seeds were harvested at 4, 8, and 16 DAP, fixed with 4% freshly prepared
567 paraformaldehyde (w/v), with 2% Tween-20 and 2% Triton X-100 (pH 7, adjusted by HCl) for 1 h
568 under vacuum. For increased fixation efficiency vacuum was broken every 10 min and applied
569 again. Subsequently, seeds were transferred into fresh fixatives and stored overnight at 4°C,
570 dehydrated using ethanol series, cleared by ROTIHistol series, and embedded into Paraplast.
571 Longitudinal dorsoventral sections of 10 μm were cut with a Reichert-Jung 2030 microtome and
572 attached to Adhesion Slides Superfrost Ultra Plus (Thermo Fisher Scientific). DNA templates for
573 synthesis of RNA probes were amplified from cDNA (reverse transcribed by RevertAid H Minus

574 First Strand cDNA Synthesis Kit; Thermo Fisher Scientific) by PCR. Sense and antisense
575 digoxigenin (DIG)-labeled RNA probes were amplified using gene-specific primers containing T7
576 promoter sequences (Supplemental Table S17) and DIG-UTP (Thermo Fisher Scientific)
577 according to the manufacturer's instructions. After purification, the efficiency of DIG labeling was
578 verified by a modified dot blotting hybridization (Zöllner et al., 2021). For hybridization, slides with
579 tissue sections were washed in ROTIHistol, rehydrated, and treated with 0.2 M HCl for 10 min,
580 pronase (0.125 mg mL⁻¹) for 10 min, 0.2% glycine for 2 min, 4% formaldehyde for 10 min and
581 acetic anhydride (1% in 0.1 M Triethanolamine pH 8.0). Hybridizations with denatured probes
582 were carried out at 50°C using the hybridization buffer containing 100 µL 10× salts, 400 µL
583 deionized formamide, 200 µL 50% dextran sulfate, 10 µL of yeast tRNA (100 mg mL⁻¹), 20 µL 50×
584 Denhardt's Solution (Thermo Fisher Scientific) and 70 µL dH₂O. After washing, unbound RNA was
585 digested for 30 min at 37°C using RNase A (20 µg mL⁻¹). Immunological detection using DIG
586 antibodies (1:3000 in blocking solution) coupled with alkaline phosphatase and staining procedure
587 with 4-Nitro blue tetrazolium chloride (NBT) and 5-Bromo-4-chloro-3-indolyl-phosphate (BCIP)
588 was done for 24–36 h at room temperature in dark. Hybridization signal analysis was performed
589 using a light microscope BX60 (Olympus).

590

591 **Identification of chromatin genes in barley**

592 Identification was done using a subset of 47 *A. thaliana* genes encoding for histones and 12
593 encoding PRC2 complex subunits. Homology searches were performed using BLAST+ (Camacho
594 et al., 2009). The resulting hits were confirmed with reciprocal homology searching using the whole
595 genome of 48,359 *A. thaliana* genes. The candidates were further filtered by standard BLAST+ E-
596 value (≤ 0.01) and additional parameters counting with the comparison of (1) barley and *A. thaliana*
597 gene length ($\geq 40\%$), and (2) alignment coverage of the hit ($\geq 40\%$).

598

599 **ImmunoFISH, microscopy, and image analysis**

600 ImmunoFISH was performed on flow-sorted nuclei from 24 DAP embryos and 8 and 24 DAP
601 endosperm as described (Nowicka et al., 2023). Preparations were incubated with the rabbit anti-
602 H3K27me3 primary antibody (1:200; Abcam, 195477) at 4°C overnight and the secondary goat
603 anti-rabbit-Alexa Fluor 488 antibody (1:300, Molecular Probes, A11008) at 37°C for 90 min. Barley
604 centromeres were detected with a synthetic 28-mer oligonucleotide (5'-AGGGAGA-3')₄ *CEREBA*
605 probe labeled at the 5' end with Cy3 (Eurofins). Telomeres were visualized with a synthetic 28-
606 mer oligonucleotide probe (5'-CCCTAAA-3')₄ labeled at the 5 end with Cy5 (Eurofins).

607 The images were acquired with an AxioImager Z2 microscope (Zeiss, Oberkochen,
608 Germany) equipped with a pE-4000 LED illuminator light source (CoolLED), laser-free confocal
609 spinning disk device (DSD2, Andor, Belfast, UK) and with $\times 100/1.4$ NA Oil M27 Plan-Apochromat
610 (Zeiss) objective. Image stacks of 40-80 slides depending from the C-value of the nucleus, on
611 average, with 0.2 μm z-step were taken separately for each fluorochrome using the appropriate
612 excitation (DAPI $\lambda = 390/40$ nm, GFP $\lambda = 482/18$ nm, RFP $\lambda = 561/14$ nm, Cy5 = 640/14 nm) and
613 emission (DAPI $\lambda = 452/45$ nm, GFP $\lambda = 525/45$ nm, RFP $\lambda = 609/54$ nm, Cy5 = 676/29 nm) filters.
614 The 4.2 MPx sCMOS camera (Zyla 4.2) and the iQ 3.6.1 acquisition software (both Andor) were
615 used to drive the microscope and for fluorescence detection. The images were saved as maximum
616 intensity projection files with Imaris File Converter 9.2.1 (Bitplane, Zurich, Switzerland). Further,
617 Imaris 9.7 functions 'Surface' and 'Spots' were used for the nucleus surface, immunosignals, and
618 centromere and telomere 3D modeling. Fluorescence intensity measurements were performed in
619 FIJI using the 'Interactive 3D Surface plot' plugin.

620

621 **Analysis of imprinted genes**

622 The lists of cereal imprinted genes were extracted from published works (Waters et al., 2011;
623 Zhang et al., 2011; Luo et al., 2011; Yang et al., 2018; Chen et al., 2018) and their overlaps were
624 visualized using Venn diagrams in R package eulerr (Larsson, 2020). *H. vulgare* subsp.
625 *spontaneum* accession HOR 12560 was grown as described (Nowicka et al., 2021) and
626 synchronized for flowering with cv. Morex. The strains were reciprocally crossed, 8 DAP
627 endosperm was manually dissected as described (Kovacik et al., 2020), total RNA was isolated
628 using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich), and reverse transcribed into cDNA using
629 RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Primers were
630 designed to amplify 200–1,100 bp fragments of the candidate genes harboring informative SNP(s)
631 (Supplemental Table S17) using a standard PCR. The amplicons were gel purified using GeneJet
632 Gel Extraction Kit (Thermo Fisher Scientific) and subjected to Sanger sequencing followed by *in*
633 *silico* analysis using SnapGene v6.2 software (GSL Biotech LLC).

634

635 **H3K27me3 ChIP-seq and data analysis**

636 Approximately 2 g of 16 DAP endosperm tissue from cv. Morex were isolated in three biological
637 replicates and cross-linked under vacuum in 1% (w/v) formaldehyde for 15 min at room
638 temperature. The cross-linking was quenched by adding glycine to a final concentration of 0.1 M,
639 and the vacuum was applied for 5 min. Endosperm tissue was rinsed with water and frozen in
640 liquid nitrogen. ChIP was performed as described previously (Gendrel et al., 2005), with the

641 following modifications. Briefly, isolated nuclei were resuspended in Nuclei lysis buffer (50 mM
642 Tris-HCl pH 8.0, 1 mM EDTA, 0.5% SDS, cComplete™, EDTA-free Protease Inhibitor Cocktail
643 (Roche)) and incubated at 4°C under gentle agitation for 20 min. The chromatin solution was
644 sonicated using a Biorupter Plus (Diagenode) with 10 cycles of 30 sec pulse/90 sec cooling at
645 high power at 4°C. The resulting sheared chromatin was pooled and diluted 1:4 with ChIP dilution
646 buffer (16.7 mM Tris-HCl pH 8.0, 167 mM NaCl, 1.25 mM EDTA, 1.25% Triton X-100, 1×
647 cComplete™, EDTA-free Protease Inhibitor Cocktail). 600 µL aliquots of diluted chromatin were
648 incubated with 7 µL of the rabbit anti-trimethyl-Histone H3 (Lys27) antibody (Millipore, 07-449) in
649 a rotator at 4°C overnight. Samples without antibody were used as negative controls. The next
650 day, 40 µL of the Dynabeads Protein A (Invitrogen) were added to each tube, and the samples
651 were further incubated for 2 hours. Afterward, beads were washed with a sequence of buffers,
652 and immune complexes were eluted as described. Control chromatin aliquots ('input DNA') were
653 taken prior to immunoprecipitation. Reverse crosslinking was performed for all samples, and DNA
654 was extracted and purified using the ChIP DNA Clean & Concentrator™ kit (Zymo Research).
655 Sequencing libraries were prepared and 150 bp pair-end reads were sequenced using Illumina
656 NovaSeq 6000 platform (Illumina) by Novogene.

657 The raw reads of samples sequenced in our study and also publicly available H3K27me3
658 ChIP-seq data from shoot (Baker et al., 2015) were trimmed for adaptors and aligned to the Morex
659 reference genome, duplicates were removed using MarkDuplicates in GATK (der Auwera and
660 O'Connor, 2020) and the peak calling was performed using MACS2 (Zhang et al., 2008). The
661 peaks were filtered (fold enrichment ≥ 5) and tested for differential signal intensity between
662 endosperm and shoot samples using R package MAnorm2 (Tu et al., 2021). Testing was
663 performed in genomic intervals of size 2000 bp and intervals with differential signal intensities
664 localized in coding regions or 2000 bp upstream were related to genes. GO term enrichment
665 analysis was performed by web based tool g:Profiler (Raudvere et al., 2019) using barley GO
666 annotation for Morex V3 available at Ensembl Plants.

667

668 **Accession numbers**

669 Unique identifiers for all genes and their products mentioned in the text are provided in
670 Supplemental Table S18.

671

672 **Acknowledgments**

673 We thank H. Tvardíková for technical assistance, Z. Bursová for plant care, and P. Navrátil for IT
674 support.

675
676 **Author contributions**
677 M.Ko. performed sample preparation and RNA extraction, data analysis and interpretation. M.Ko.,
678 I.V. and M.Č. performed RNA *in situ* hybridization. A.N. performed immunostaining and
679 fluorescence *in situ* hybridization, microscopy, image analysis and barley crosses. J.Z. performed
680 the identification of chromatin genes. B.S. performed ChIP and contributed to data analysis.
681 N.J.P., E.E. and A.Pa. prepared a seed view in barley ePlant. K.K. and M.Kr. performed validation
682 of imprinted genes. R.S. and J.Š. supervised parts of the project. A.Pe. conceived, designed, and
683 supervised the project. A.Pe and M.Ko. wrote the manuscript. All authors have seen and agreed
684 upon the final version of the manuscript.

685
686 **Supplemental data**
687 The following materials are available in the online version of this article.

688 **Supplemental Figure S1.** Sample distance heatmap.

689 **Supplemental Figure S2.** Principal component analysis (PCA) of seed transcriptomes.

690 **Supplemental Figure S3.** Numbers of tissue-specific and shared genes among seed tissues
691 expressed at different levels.

692 **Supplemental Figure S4.** Volcano plots of differentially expressed genes (DEGs) in consecutive
693 developmental transitions.

694 **Supplemental Figure S5.** K-means co-expression clusters of DEGs.

695 **Supplemental Figure S6.** WGCNA corresponding modules and their hierarchical clustering.

696 **Supplemental Figure S7.** Transcription factor motif cluster 9 analysis.

697 **Supplemental Figure S8.** RNA *in situ* hybridization of *PBF* in barley grains at 4, 8, and 16 days
698 after pollination (DAP).

699 **Supplemental Figure S9.** Expression profiles of barley orthologs of selected endosperm marker
700 genes.

701 **Supplemental Figure S10.** Heatmap of expression from 175 histone genes.

702 **Supplemental Figure S11.** Cytology analysis of H3K27me3.

703 **Supplemental Figure S12.** Abundance of H3K27me3.

704 **Supplemental Figure S13.** Comparison of H3K27me3 profiles between endosperm and shoot.

705 **Supplemental Figure S14.** Validation of candidates for imprinted genes.

706 **Supplemental Table S1.** Gene ontology (GO) term category enrichment based on the subset of
707 genes with TPM > 1,000 in endosperm tissues.

708 **Supplemental Table S2.** Gene ontology (GO) term enrichment for down-regulated DEGs.

709 **Supplemental Table S3.** Gene ontology (GO) term enrichment for up-regulated DEGs.
710 **Supplemental Table S4.** List of motifs assigned to different motif clusters.
711 **Supplemental Table S5.** Contribution of each motif collection to the motif clusters identified by
712 RSAT.
713 **Supplemental Table S6.** Barley genes homologous to Arabidopsis genes with binding motifs
714 assigned to motif cluster 2.
715 **Supplemental Table S7.** Barley genes homologous to Arabidopsis genes with binding motifs
716 assigned to motif cluster 9.
717 **Supplemental Table S8.** Homologies of known rice and maize endosperm domains marker genes
718 with barley genes.
719 **Supplemental Table S9.** Barley histone and PRC2 genes.
720 **Supplemental Table S10.** Seed transcriptomic data for barley histone and PRC2 genes.
721 **Supplemental Table S11.** Up-regulated genes with significantly reduced H3K27me3 in 16 DAP
722 endosperm (END) compared to shoot (SHO).
723 **Supplemental Table S12.** Down-regulated genes with significantly reduced H3K27me3 in 16
724 DAP endosperm (END) compared to shoot (SHO).
725 **Supplemental Table S13.** GO term enrichment for up-regulated and H3K27me3 significantly
726 reduced genes in 16 DAP endosperm compared to shoot.
727 **Supplemental Table S14.** GO term enrichment for up-regulated and H3K27me3 significantly
728 reduced genes in 16 DAP endosperm compared to shoot.
729 **Supplemental Table S15.** Homology of barley genes to imprinted genes identified in maize, rice
730 and wheat.
731 **Supplemental Table S16.** Seed transcriptomic data for barley homologs of imprinted genes
732 identified in at least two other cereals.
733 **Supplemental Table S17.** Primers used in this study.
734 **Supplemental Table S18.** Gene names and IDs used in the text.
735 **Supplemental Data Set S1.** RNA-seq raw read counts for the embryo (EMB), endosperm (END)
736 and seed maternal tissues (SMTs).
737 **Supplemental Data Set S2.** Expression changes between the subsequent days after pollination
738 in the embryo (EMB), endosperm (END), and seed maternal tissues (SMTs).
739 **Supplemental Data Set S3.** Assignment of differentially expressed genes from the embryo
740 (EMB), endosperm (END) and seed maternal tissues (SMTs) to individual k-means clusters (K).
741 **Supplemental Data Set S4.** Gene assignment to WGCNA modules in the embryo (EMB),
742 endosperm (END) and seed maternal tissues (SMTs).

743 **Supplemental Data Set S5.** Identified plant motifs within WGCNA modules in embryo, endosperm
744 and seed maternal tissues.

745 **Supplemental Data Set S6.** Identified *de novo* plant motifs within WGCNA modules in embryo,
746 endosperm and seed maternal tissues.

747 **Supplemental Data Set S7.** List of genomic intervals with significantly differential H3K27me3
748 signal intensity in endosperm (END) compared to shoot (SHO) in ChIP-seq.

749 **Supplemental Data Set S8.** List of genes with the differential signal intensity of H3K27me3 in
750 endosperm (END) compared to shoot (SHO).

751 **Supplemental Data Set S9.** List of transcription factors and their assignments to families.

752

753 **Funding**

754 Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ
755 LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

756 This work was supported by the Czech Science Foundation grant 21-02929S (A.P) and German
757 Research Foundation grants EXC2048 (R.S.) and IRTG2466 (I.V.).

758

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

762

763 **Data availability**

764 All data supporting this study are included in the article and its Supplemental material. Sequence
765 data from this article can be found in the Gene Expression Omnibus under accession numbers
766 GSE233316 and GSE238237.

767

768

769 **References**

770 **Alexa, A. and Rahnenfuhrer, J.** (2021). topGO: Enrichment analysis for gene ontology.

771 **Alizadeh, M., Hoy, R., Lu, B., and Song, L.** (2021). Team effort: Combinatorial control
772 of seed maturation by transcription factors. *Curr Opin Plant Biol* **63**: 102091.

773 **Anders, S. and Huber, W.** (2010). Differential expression analysis for sequence count
774 data. *Genome Biol* **11**: 1–12.

775 **der Auwera, G.A. and O'Connor, B.D.** (2020). Genomics in the cloud: using Docker,
776 GATK, and WDL in Terra (O'Reilly Media).

777 **Baker, K., Dhillon, T., Colas, I., Cook, N., Milne, I., Milne, L., Bayer, M., and Flavell,**
778 **A.J.** (2015). Chromatin state analysis of the barley epigenome reveals a higher-order
779 structure defined by H3K27me1 and H3K27me3 abundance. *Plant J* **84**: 111–124.

780 **Bate, N.J., Niu, X., Wang, Y., Reimann, K.S., and Helentjaris, T.G.** (2004). An invertase
781 inhibitor from maize localizes to the embryo surrounding region during early kernel
782 development. *Plant Physiol* **134**: 246–254.

783 **Batista, R.A. and Köhler, C.** (2020). Genomic imprinting in plants-revisiting existing
784 models. *Genes Dev* **34**: 24–36.

785 **Bennet, M.D., Smith, J.B., and Barclay, I.** (1975). Early seed development in the
786 Triticeae. *Philos Trans R Soc Lond B Biol Sci* **272**: 199–227.

787 **Bian, J., Deng, P., Zhan, H., Wu, X., Nishantha, M., Yan, Z., Du, X., Nie, X., and Song,**
788 **W.** (2019). Transcriptional Dynamics of Grain Development in Barley (*Hordeum*
789 *vulgare* L.). *Int J Mol Sci* **20**: 962.

790 **Brown, R.C., Lemmon, B.E., and Nguyen, H.** (2003). Events during the first four rounds
791 of mitosis establish three developmental domains in the syncytial endosperm of
792 *Arabidopsis thaliana*. *Protoplasma* **222**: 167–174.

793 **Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and**
794 **Madden, T.L.** (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*
795 **10**: 1–9.

796 **De Camilli, P., Chen, H., Hyman, J., Panepucci, E., Bateman, A., and Brunger, A.T.**
797 (2002). The ENTH domain. *FEBS Lett* **513**: 11–18.

798 **Carena, M.J. ed** (2009). *Cereals* (Springer US: New York, NY).

799 **Castro-Mondragon, J.A., Jaeger, S., Thieffry, D., Thomas-Chollier, M., and van**
800 **Helden, J.** (2017). RSAT matrix-clustering: dynamic exploration and redundancy
801 reduction of transcription factor binding motif collections. *Nucleic Acids Res* **45**:
802 e119–e119.

803 **Charlton, W.L., Keen, C.L., Merriman, C., Lynch, P., Greenland, A.J., and Dickinson,**
804 **H.G.** (1995). Endosperm development in *Zea mays*; implication of gametic imprinting

805 and paternal excess in regulation of transfer layer development. *Development* **121**:
806 3089–3097.

807 **Chen, C. et al.** (2018). Characterization of imprinted genes in rice reveals conservation
808 of regulation and imprinting with other plant species. *Plant Physiol* **177**: 1754–1771.

809 **Chourey, P.S. and Nelson, O.E.** (1976). The enzymatic deficiency conditioned by the
810 shrunken-1 mutations in maize. *Biochem Genet* **14**: 1041–1055.

811 **Cook, F., Hughes, A., Nibau, C., Orman-Ligeza, B., Schatlowski, N., Uauy, C., and**
812 **Trafford, K.** (2018). Barley lys3 mutants are unique amongst shrunken-endosperm
813 mutants in having abnormally large embryos. *J Cereal Sci* **82**: 16–24.

814 **Costa, L.M., Gutierrez-Marcos, J.F., Brutnell, T.P., Greenland, A.J., and Dickinson,**
815 **H.G.** (2003). The globby1-1 (glo1-1) mutation disrupts nuclear and cell division in the
816 developing maize seed causing alterations in endosperm cell fate and tissue
817 differentiation. *Development* **130**: 5009–5017.

818 **Doan, D.N.R., Linnestad, C., and Olsen, A.** (1996). Isolation of molecular markers from
819 the barley endosperm coenocyte and the surrounding nucellus cell layers. *Plant Mol*
820 *Biol* **31**: 877–886.

821 **Druka, A. et al.** (2006). An atlas of gene expression from seed to seed through barley
822 development. *Funct Integr Genomics* **6**: 202–211.

823 **Du, L., Li, N., Chen, L., Xu, Y., Li, Y., Zhang, Y., Li, C., and Li, Y.** (2014). The ubiquitin
824 receptor DA1 regulates seed and organ size by modulating the stability of the
825 ubiquitin-specific protease UBP15/SOD2 in Arabidopsis. *Plant Cell* **26**: 665–677.

826 **Food and Agriculture Organization of the United Nations** (2023).

827 **Gendrel, A.V., Lippman, Z., Martienssen, R., and Colot, V.** (2005). Profiling histone
828 modification patterns in plants using genomic tiling microarrays. *Nat Methods* **2**: 213–
829 218.

830 **Gómez, E., Royo, J., Guo, Y., Thompson, R., and Hueros, G.** (2002). Establishment of
831 cereal endosperm expression domains identification and properties of a maize
832 transfer cell-specific transcription factor, ZmMRP-1. *Plant Cell* **14**: 599–610.

833 **Goyal, K., Walton, L.J., and Tunnacliffe, A.** (2005). LEA proteins prevent protein
834 aggregation due to water stress. *Biochem J* **388**: 151–157.

835 **Gu, B., Dong, H., Smith, C., Cui, G., Li, Y., and Bevan, M.W.** (2022). Modulation of
836 receptor-like transmembrane kinase 1 nuclear localization by DA1 peptidases in
837 Arabidopsis. *Proc Natl Acad Sci U S A* **119**: e2205757119.

838 **Gubatz, S., Dercksen, V.J., Brüß, C., Weschke, W., and Wobus, U.** (2007). Analysis of
839 barley (*Hordeum vulgare*) grain development using three-dimensional digital models.
840 *Plant J* **52**: 779–790.

841 **Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X.,
842 Murre, C., Singh, H., and Glass, C.K.** (2010). Simple combinations of lineage-
843 determining transcription factors prime cis-regulatory elements required for
844 macrophage and B cell identities. *Mol Cell* **38**: 576.

845 **Hertig, C., Melzer, M., Rutten, T., Erbe, S., Hensel, G., Kumlehn, J., Weschke, W.,
846 Weber, H., and Thiel, J.** (2020). Barley HISTIDINE KINASE 1 (HvHK1) coordinates
847 transfer cell specification in the young endosperm. *Plant Journal* **103**: 1869–1884.

848 **Hertig, C., Rutten, T., Melzer, M., Schippers, J.H.M., and Thiel, J.** (2023). Dissection
849 of Developmental Programs and Regulatory Modules Directing Endosperm Transfer
850 Cell and Aleurone Identity in the Syncytial Endosperm of Barley. *Plants* **12**.

851 **Hueros, G., Royo, J., Maitz, M., Salamini, F., and Thompson, R.D.** (1999). Evidence
852 for factors regulating transfer cell-specific expression in maize endosperm. *Plant Mol*
853 *Biol* **41**: 403–414.

854 **Jiang, D., Borg, M., Lorković, Z.J., Montgomery, S.A., Osakabe, A., Yelagandula, R.,
855 Axelsson, E., and Berger, F.** (2020). The evolution and functional divergence of the
856 histone H2B family in plants. *PLoS Genet* **16**: e1008964.

857 **Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G.** (2017).
858 PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory
859 interactions in plants. *Nucleic Acids Res* **45**: D1040–D1045.

860 **Kalla, R., Shimamoto, K., Potter, R., Nielsen, P.S., Linnestad, C., and Olsen, O. -A**
861 (1994). The promoter of the barley aleurone-specific gene encoding a putative 7 kDa
862 lipid transfer protein confers aleurone cell-specific expression in transgenic rice. *Plant*
863 *J* **6**: 849–860.

864 **Kanegae, H., Miyoshi, K., Hirose, T., Tsuchimoto, S., Mori, M., Nagato, Y., and**
865 **Takano, M.** (2005). Expressions of rice sucrose non-fermenting-1 related protein

866 kinase 1 genes are differently regulated during the caryopsis development. *Plant*
867 *Physiology and Biochemistry* **43**: 669–679.

868 **Kiesselbach, T.A. and Walker, E.R.** (1952). Structure of certain specialized tissue in the
869 kernel of corn. *Am J Bot* **39**: 561–569.

870 **Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L.** (2019). Graph-based
871 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat*
872 *Biotechnol* **37**: 907–915.

873 **Kovacik, M., Nowicka, A., and Pecinka, A.** (2020). Isolation of high purity tissues from
874 developing barley seeds. *J Vis Exp*: e61681.

875 **Langfelder, P. and Horvath, S.** (2012). Fast {R} functions for robust correlations and
876 hierarchical clustering. *J Stat Softw* **46**: 1–17.

877 **Langfelder, P. and Horvath, S.** (2008). WGCNA: An R package for weighted correlation
878 network analysis. *BMC Bioinformatics* **9**: 559.

879 **Langfelder, P., Zhang, B., and Horvath, S.** (2008). Defining clusters from a hierarchical
880 cluster tree: The dynamic tree cut package for R. *Bioinformatics* **24**: 719–720.

881 **Larsson, J.** (2020). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses.

882 **Le, B.H. et al.** (2010). Global analysis of gene activity during Arabidopsis seed
883 development and identification of seed-specific transcription factors. *Proc Natl Acad*
884 *Sci U S A* **107**: 8063–8070.

885 **Leah, R., Tommerup, H., Svendsen, I., and Mundy, J.** (1991). Biochemical and
886 molecular characterization of three barley seed proteins with antifungal properties. *J*
887 *Biol Chem* **266**: 1564–1573.

888 **Liao, Y., Smyth, G.K., and Shi, W.** (2013). The Subread aligner: Fast, accurate and
889 scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**: e108–e108.

890 **Love, M.I., Huber, W., and Anders, S.** (2014). Moderated estimation of fold change and
891 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

892 **Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and**
893 **Koltunow, A.** (2011). A genome-wide survey of imprinted genes in rice seeds reveals
894 imprinting primarily occurs in the endosperm. *PLoS Genet* **7**: e1002125.

895 **Magnard, J.L., Lehouque, G., Massonneau, A., Frangne, N., Heckel, T., Gutierrez-**
896 **Marcos, J.F., Perez, P., Dumas, C., and Rogowsky, P.M.** (2003). ZmEBE genes

897 show a novel, continuous expression pattern in the central cell before fertilization and
898 in specific domains of the resulting endosperm after fertilization. *Plant Mol Biol* **53**:
899 821–836.

900 **Martin, M.** (2011). Cutadapt removes adapter sequences from high-throughput
901 sequencing reads. *EMBnet J* **17**: 10–12.

902 **Mascher, M.** (2021). Pseudomolecules and annotation of the third version of the
903 reference genome sequence assembly of barley cv. Morex [Morex V3].

904 **Mena, M., Vicente-Carbajosa, J., Schmidt, R.J., and Carbonero, P.** (1998). An
905 endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and
906 activates transcription from the prolamin-box of a native B-hordein promoter in barley
907 endosperm. *Plant J* **16**: 53–62.

908 **Mozgova, I. and Hennig, L.** (2015). The polycomb group protein regulatory network.
909 *Annu Rev Plant Biol* **66**: 269–296.

910 **Nowicka, A., Ferková, L., Said, M., Kovacik, M., Zwyrtková, J., Baroux, C., and
911 Pecinka, A.** (2023). Non-Rabl chromosome organization in endoreduplicated nuclei
912 of barley embryo and endosperm tissues. *J Exp Bot* **74**: 2527–2541.

913 **Nowicka, A., Sahu, P.P., Kovacik, M., Weigt, D., Tokarz, B., Krugman, T., and
914 Pecinka, A.** (2021). Endopolyploidy variation in wild barley seeds across
915 environmental gradients in israel. *Genes (Basel)* **12**: 711.

916 **Olsen, O.-A.** (2001). ENDOSPERM DEVELOPMENT: Cellularization and cell fate
917 specification. *Annu Rev Plant Biol* **52**: 233–267.

918 **Opsahl-Ferstad, H.G., Deunff, E. Le, Dumas, C., and Rogowsky, P.M.** (1997). ZmEsr,
919 a novel endosperm-specific gene expressed in a restricted region around the maize
920 embryo. *Plant J* **12**: 235–246.

921 **Peirats-Llobet, M., Yi, C., Liew, L.C., Berkowitz, O., Narsai, R., Lewsey, M.G., and
922 Whelan, J.** (2023). Spatially resolved transcriptomic analysis of the germinating
923 barley grain. *Nucleic Acids Res* **51**: 7798–7819.

924 **Peng, Y., Chen, L., Lu, Y., Wu, Y., Dumenil, J., Zhu, Z., Bevan, M.W., and Lia, Y.**
925 (2015). The ubiquitin receptors DA1, DAR1, and DAR2 redundantly regulate
926 endoreduplication by modulating the stability of TCP14/15 in Arabidopsis. *Plant Cell*
927 **27**: 649–662.

928 **Probst, A. V, Desvoyes, B., and Gutierrez, C.** (2020). Similar yet critically different: the
929 distribution, dynamics and function of histone variants. *J Exp Bot* **71**: 5191–5204.

930 **R Core Team** (2020). R: A language and environment for statistical computing.

931 **Radchuk, V. V., Borisjuk, L., Sreenivasulu, N., Merx, K., Mock, H.P., Rolletschek, H.,**
932 **Wobus, U., and Weschke, W.** (2009). Spatiotemporal Profiling of Starch
933 Biosynthesis and Degradation in the Developing Barley Grain. *Plant Physiol* **150**:
934 190–204.

935 **Radchuk, V., Weier, D., Radchuk, R., Weschke, W., and Weber, H.** (2011).
936 Development of maternal seed tissue in barley is mediated by regulated cell
937 expansion and cell disintegration and coordinated with endosperm growth. *J Exp Bot*
938 **62**: 1217–1227.

939 **Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J.**
940 (2019). g:Profiler: a web server for functional enrichment analysis and conversions of
941 gene lists (2019 update). *Nucleic Acids Res* **47**: W191–W198.

942 **Roberts, I.N., Caputo, C., Criado, M.V., and Funk, C.** (2012). Senescence-associated
943 proteases in plants. *Physiol Plant* **145**: 130–139.

944 **Sano, N., Rajjou, L., and North, H.M.** (2020). Lost in translation: Physiological roles of
945 stored mRNAs in seed germination. *Plants (Basel)* **9**: 347–361.

946 **Serna, A., Maitz, M., O’Connell, T., Santandrea, G., Thevissen, K., Tienens, K.,**
947 **Hueros, G., Faleri, C., Cai, G., Lottspeich, F., and Thompson, R.D.** (2001). Maize
948 endosperm secretes a novel antifungal protein into adjacent maternal tissue. *Plant J*
949 **25**: 687–698.

950 **Sosso, D., Wisniewski, J.P., Khaled, A.S., Hueros, G., Gerentes, D., Paul, W., and**
951 **Rogowsky, P.M.** (2010). The Vpp1, Esr6a, Esr6b and OCL4 promoters are active in
952 distinct domains of maize endosperm. *Plant Science* **179**: 86–96.

953 **Sreenivasulu, N., Usadel, B., Winter, A., Radchuk, V., Scholz, U., Stein, N., Weschke,**
954 **W., Strickert, M., Close, T.J., Stitt, M., Graner, A., and Wobus, U.** (2008). Barley
955 grain maturation and germination: Metabolic pathway and regulatory network
956 commonalities and differences highlighted by new MapMan/PageMan profiling tools.
957 *Plant Physiol* **146**: 1738–1758.

958 **Stewart Gillmor, C., Silva-Ortega, C.O., Willmann, M.R., Buendía-Monreal, M., and**
959 **Poethig, R.S.** (2014). The Arabidopsis Mediator CDK8 module genes CCT (MED12)
960 and GCT (MED13) are global regulators of developmental phase transitions.
961 *Development* **141**: 4580–4589.

962 **Strejčková, B., Čegan, R., Pecinka, A., Milec, Z., and Šafář, J.** (2020). Identification of
963 polycomb repressive complex 1 and 2 core components in hexaploid bread wheat.
964 *BMC Plant Biol* **20**: 1–13.

965 **Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T.** (2011). REVIGO Summarizes and
966 Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**: e21800.

967 **The International Barley Genome Sequencing Consortium** (2012). A physical, genetic
968 and functional sequence assembly of the barley genome. *Nature* **491**: 711–716.

969 **Thiel, J. et al.** (2021). Transcriptional landscapes of floral meristems in barley. *Sci Adv* **7**:
970 eabf0832.

971 **Tibshirani, R., Walther, G., and Hastie, T.** (2001). Estimating the number of clusters in
972 a data set via the gap statistic. *J R Stat Soc Series B Stat Methodol* **63**: 411–423.

973 **Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J.** (2005). The Botany Array
974 Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J* **43**: 153–
975 163.

976 **Tu, S., Li, M., Chen, H., Tan, F., Xu, J., Waxman, D.J., Zhang, Y., and Shao, Z.** (2021).
977 MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Res*
978 **31**: 131–145.

979 **Vishal, B. and Kumar, P.P.** (2018). Regulation of seed germination and abiotic stresses
980 by gibberellins and abscisic acid. *Front Plant Sci* **9**: 838.

981 **Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu,**
982 **W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M.** (2011).
983 Parent-of-origin effects on gene expression and DNA methylation in the maize
984 endosperm. *Plant Cell* **23**: 4221–4233.

985 **Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H., and Wobus,**
986 **U.** (2003). The role of invertases and hexose transporters in controlling sugar ratios
987 in maternal and filial tissues of barley caryopses during early development. *Plant J*
988 **33**: 395–411.

989 **Yanagisawa, S. and Sheen, J.** (1998). Involvement of Maize Dof Zinc Finger Proteins in
990 Tissue-Specific and Light-Regulated Gene Expression. *Plant Cell* **10**: 75–89.

991 **Yang, G., Liu, Z., Gao, L., Yu, K., Feng, M., Yao, Y., Peng, H., Hu, Z., Sun, Q., Ni, Z.,**
992 **and Xin, M.** (2018). Genomic imprinting was evolutionarily conserved during wheat
993 polyploidization. *Plant Cell* **30**: 37–47.

994 **Yon Rhee, S., Wood, V., Dolinski, K., and Draghici, S.** (2008). Use and misuse of the
995 gene ontology annotations. *Nat Rev Genet* **9**: 509–515.

996 **Zhang, H. et al.** (2004). Large-scale analysis of the barley transcriptome based on
997 expressed sequence tags. *Plant J* **40**: 276–290.

998 **Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Zhao, H., Guan, H., Hu, X.,**
999 **Jiao, Y., Song, W., and Lai, J.** (2011). Extensive, clustered parental imprinting of
1000 protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad*
1001 *Sci U S A* **108**: 20042–20047.

1002 **Zhang, S., Laurie, A.E., Ae, W., Meng, L., and Lemaux, P.G.** (2002). Similarity of
1003 expression patterns of knotted1 and ZmLEC1 during somatic and zygotic
1004 embryogenesis in maize (*Zea mays* L.). *Planta* **215**: 191–194.

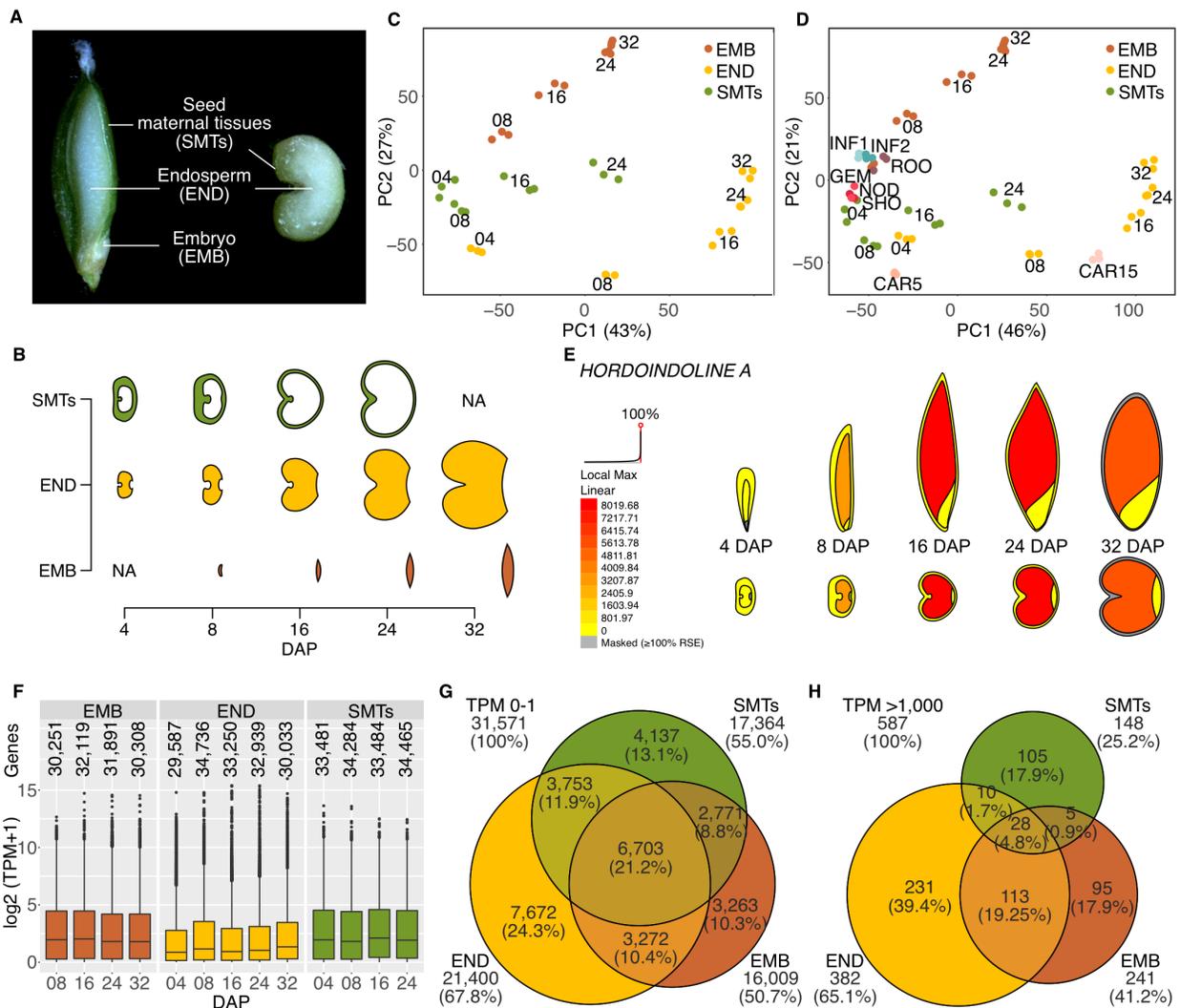
1005 **Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E.,**
1006 **Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Shirley, X.S.** (2008). Model-
1007 based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: 1–9.

1008 **Zöllner, N.R., Bezruczyk, M., Laureyns, R., Nelissen, H., Simon, R., and Frommer,**
1009 **W.B.** (2021). An RNA in situ hybridization protocol optimized for monocot tissue.
1010 *STAR Protoc* **2**: 100398.

1011

1012

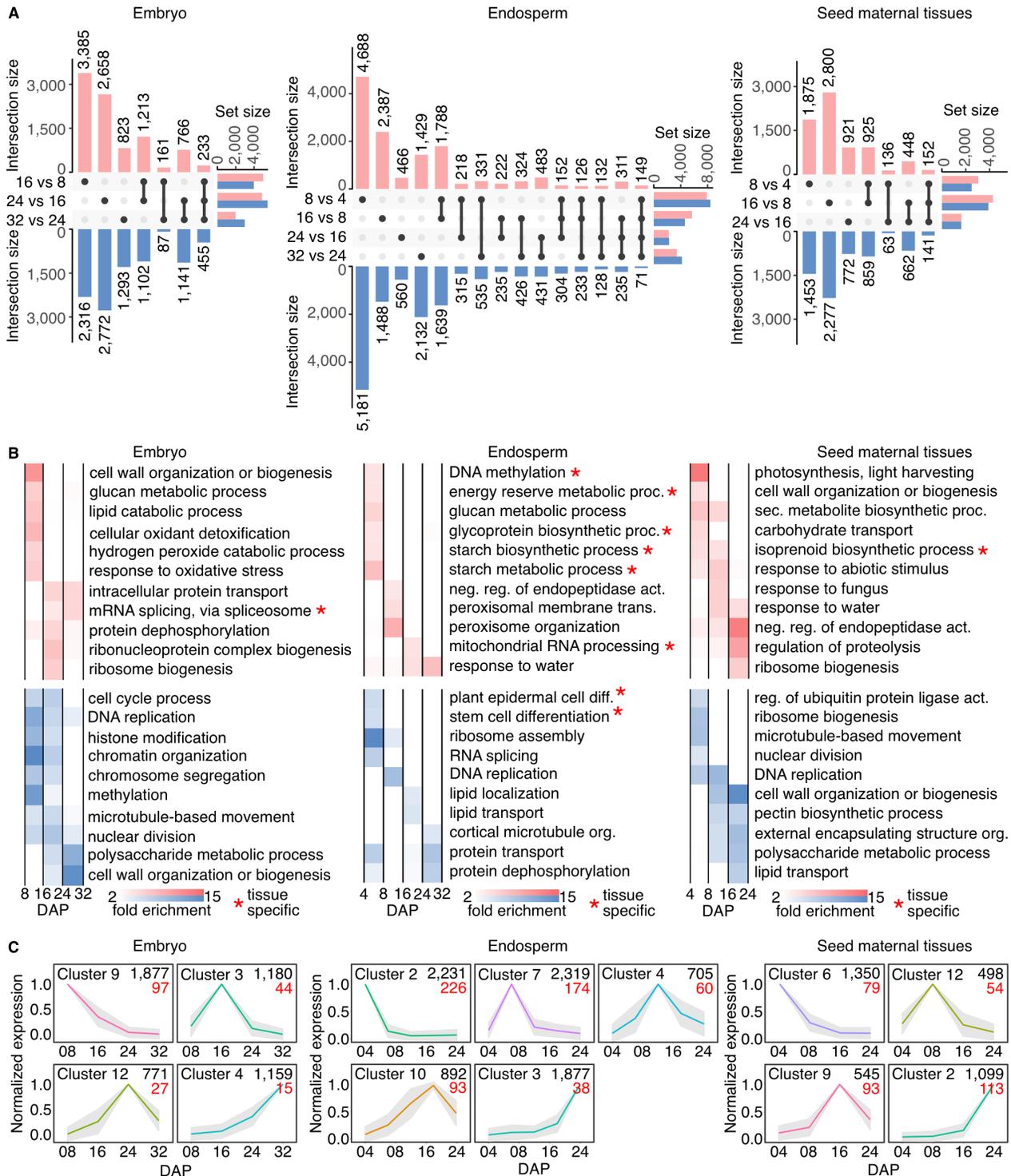
1013 **Figures and Tables**



1014
 1015
 1016 **Figure 1.** Transcriptomes of developing barley grain. **A-B)** Overview of the analyzed tissues and
 1017 time points used for transcriptomic analysis. DAP - Days after pollination; NA - not analyzed. **C)**
 1018 The variance of the 39 samples represented by principal component (PC) analysis for the embryo
 1019 (orange), endosperm (yellow), and seed maternal tissues (green). The numbers within the graph
 1020 indicate DAP and three close spots represent biological triplicates. The number next to PC
 1021 indicates variance. **D)** Principal component analysis of seed tissues used in C) in combination with
 1022 other barley tissues (The International Barley Genome Sequencing Consortium, 2012). ROO –
 1023 root, GEM – germinating embryo, NOD – nodule, SHO – shoot, INF1 and INF2 – developing
 1024 inflorescence of 5- and 10-mm length, CAR5 and CAR15 – caryopsis 5 and 15 DAP. **E)**
 1025 Visualization of interactive heatmap in the Barley ePlant. Heatmap in transverse (upper) and
 1026 sagittal (lower) sections shows TPM expression levels of *HORDOINDOLINE A* in different tissues

1027 of developing barley grain. **F)** Boxplots of expression for genes with non-zero expression (source
1028 values in Supplemental Data Set S1). **G)** Venn diagram showing numbers of lowly expressed
1029 genes (0-1 TPM during entire grain development). **H)** Venn diagram of the highest expressed
1030 genes (>1,000 TPM).

1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060



1062

1063

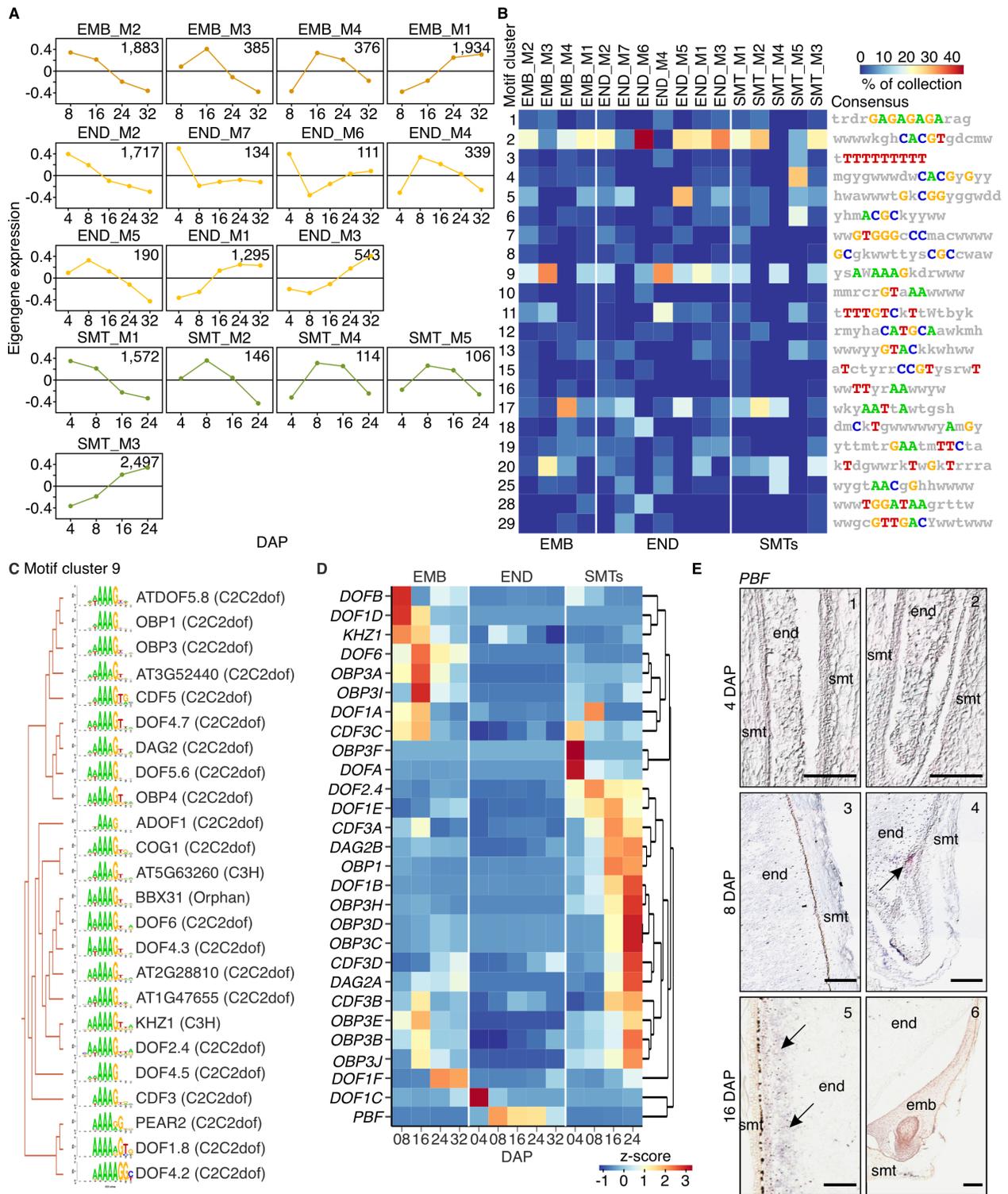
1064

1065

1066

Figure 2. Tissue and timepoint specificity of gene expression during barley grain development. **A)** The UpSet plots show up-regulated (red) and down-regulated (blue) differentially expressed genes (DEGs) in embryo, endosperm, and seed maternal tissues. The set size on the x-axis defines the total number of DEGs between two subsequent experimental time points. The y-axis

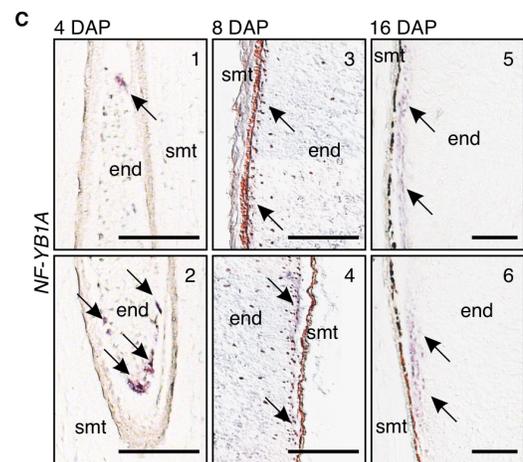
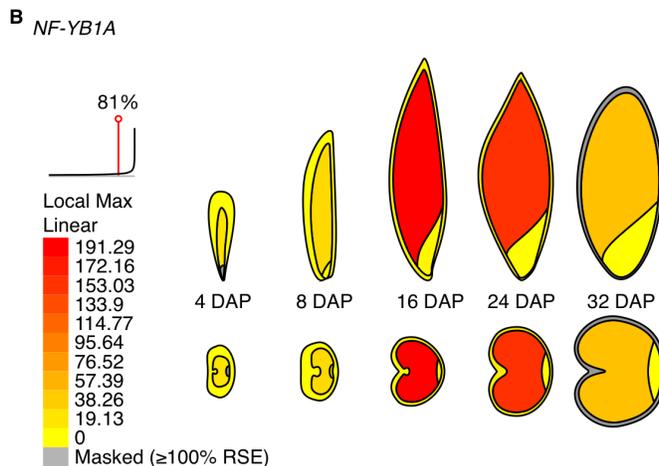
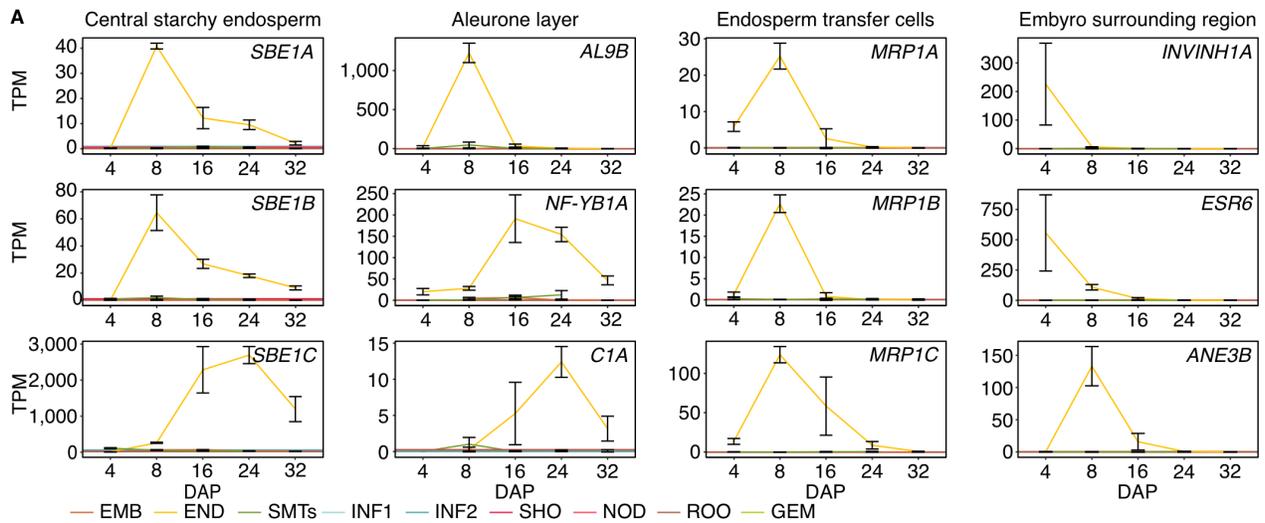
1067 shows the number of DEGs (intersection size) in single stages (single black dots) and their
1068 combinations (connected dots; source data in Supplemental Data Set S2). **B)** Gene ontology (GO)
1069 term enrichment of DEGs (FDR-adjusted $P < 0.05$) between developmental time points in embryo,
1070 endosperm, and seed maternal tissues. The top representative enriched GO categories among
1071 up-regulated (red) or down-regulated (blue) genes are shown. Color saturation corresponds to
1072 fold enrichment (full list in Supplemental Tables S2 and S3). Tissue-specific GO categories are
1073 indicated by a red asterisk. **C)** K-means co-expression clusters of DEGs, peaking in single
1074 consecutive developmental time points in embryo, endosperm, and seed maternal tissues. The
1075 black numbers in the upper right corner indicate gene count in individual clusters. The red numbers
1076 indicate a subset of tissue-specific marker genes. DAP – days after pollination (the full list in
1077 Supplemental Fig. S5 and Supplemental Data Set S3).
1078



1079
1080
1081
1082
1083

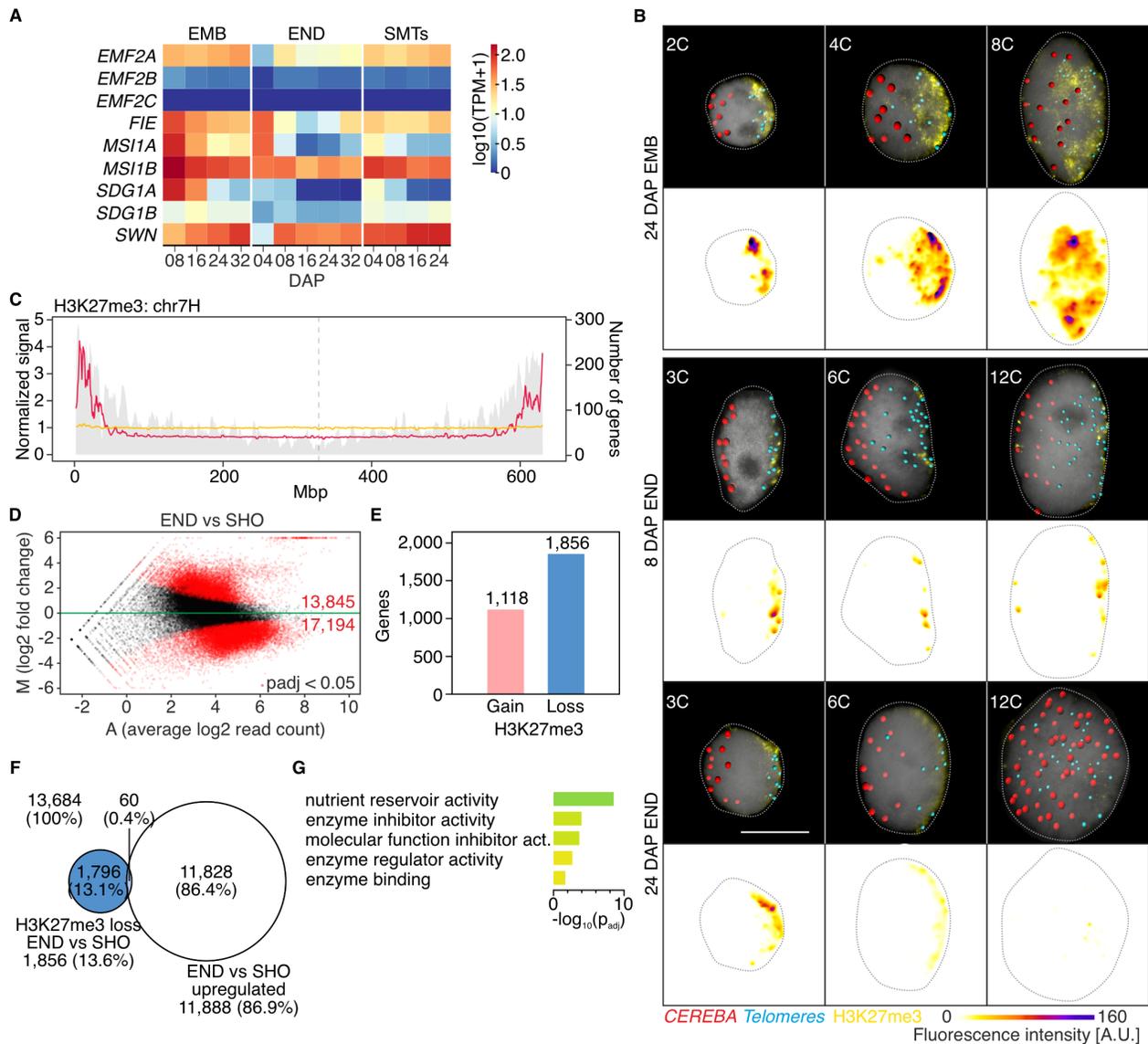
Figure 3. Gene co-expression network analysis and promoter motif enrichment. **A)** Display of selected weighted gene co-expression network analysis (WGCNA) modules (the full list of modules is provided in Supplemental Data Set S4 and Supplemental Fig. S6). The graphs show

1084 eigengene expression in each module. The numbers in the upper right corner indicate gene count
1085 in individual clusters. EMB – embryo, END – endosperm, SMTs – seed maternal tissues, DAP –
1086 days after pollination. **B)** Analysis of transcription factor (TF) binding motifs in individual WGCNA
1087 modules shown in (A) by screening promoters of barley genes against the Arabidopsis TF binding
1088 motif database and clustering. The consensual motif in each motif cluster (MC, rows) is shown on
1089 the right side and its representation (%) in the WGCNA module is indicated by the color scale.
1090 The heatmap depicts some motifs from individual clusters identified in each WGCNA module
1091 containing at least 10 motifs (full list in Supplemental Table S5). **C)** Hierarchical clustering of TF
1092 binding sequence motifs in MC9 with the *in silico*-predicted binding Arabidopsis TFs and their
1093 families (in parentheses). **D)** Heatmap of hierarchically clustered expression for barley orthologs
1094 of Arabidopsis TFs from (C) in EMB, END, and SMTs (source data in Supplemental Table S7). **E)**
1095 RNA *in situ* hybridization of *PBF* with antisense probe in barley grains at 4, 8 and 16 DAP. emb –
1096 embryo, end – endosperm, smt – seed maternal tissues. Arrows indicate signal deposition in the
1097 endosperm. The full-scale images are presented in Supplemental Figure S8 and the numbers in
1098 the top right corners correspond to the inset numbers. Scale bars 200 μm .
1099



1100
 1101
 1102 **Figure 4.** Expression of marker genes in endosperm domains in barley. **A)** Expression profiles of
 1103 barley orthologs of selected endosperm marker genes in other cereals grouped according to the
 1104 domain of expression - central starchy endosperm, aleurone layer, endosperm transfer cells, and
 1105 embryo surrounding region across different barley tissues (The International Barley Genome
 1106 Sequencing Consortium, 2012). EMB – embryo, END – endosperm, SMTs – seed maternal
 1107 tissues, ROO – root, GEM – germinating embryo, NOD – nodule, SHO – shoot, INF1 and INF2 –
 1108 developing inflorescence 5 and 10 mm, CAR5 and CAR15 – caryopsis 5 and 15 days after
 1109 pollination (DAP). Error bars indicate standard deviation (the full list in Supplemental Table S8 and
 1110 Supplemental Fig. S9). **B)** Visualization *NF-YB1A* expression (transcripts per million) in different
 1111 tissues of developing barley grain in the Barley ePlant. **C)** RNA *in situ* hybridization of *NF-YB1A*
 1112 in barley grains at 4, 8, and 16 days after pollination (DAP) using an antisense probe. emb –
 1113 embryo, end – endosperm, smt – seed maternal tissues. The full-scale images are presented in

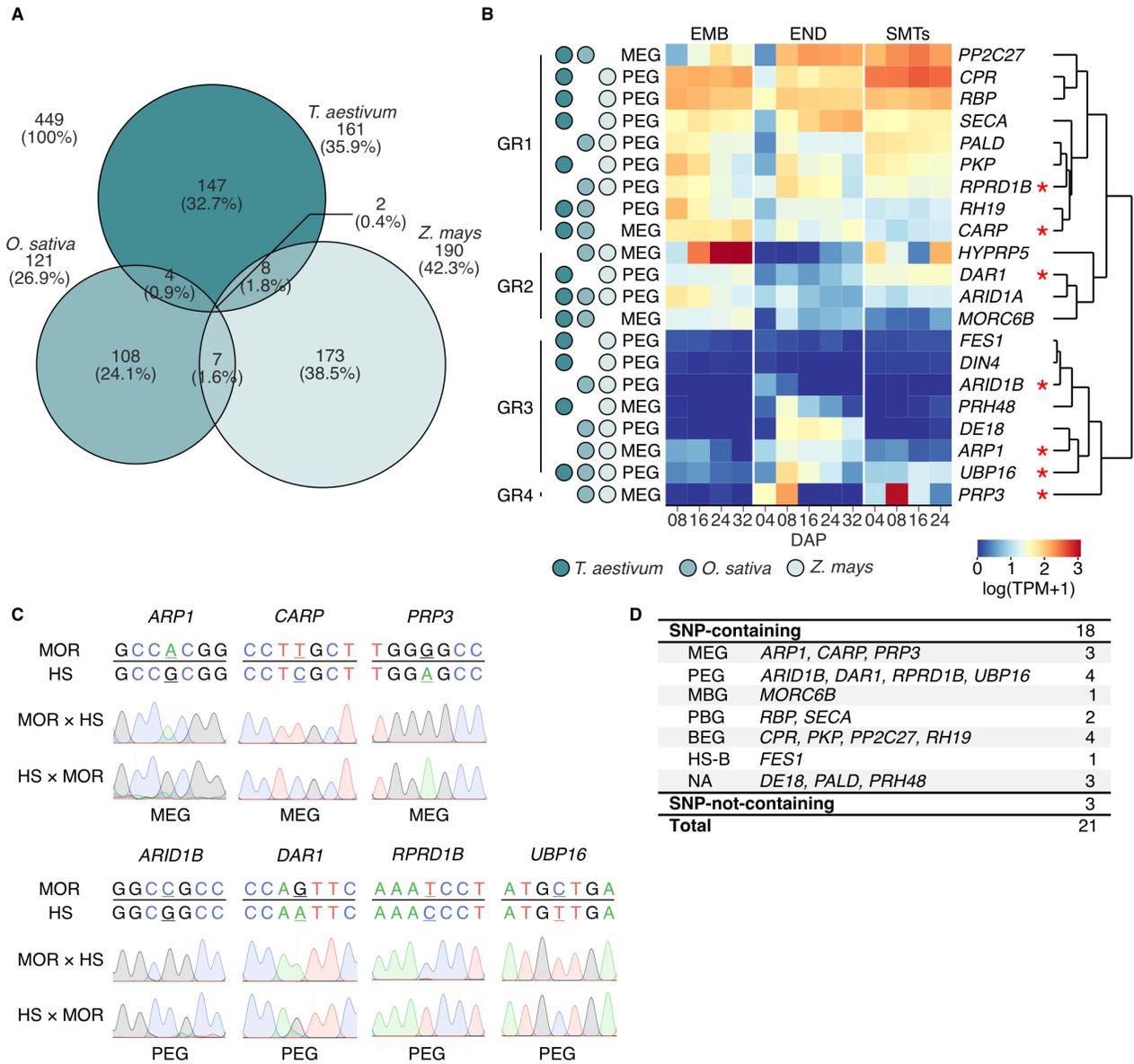
- 1114 Supplemental Figure S9 and the numbers in the top right corners correspond to the inset numbers.
- 1115 Arrows indicate signal deposition in the aleurone region. Scale bars, 200 μm .



1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127

Figure 5. Dynamics of expression from Polycomb repressive complex 2 (PRC2) genes in barley grain tissues. **A)** Heatmap of expression from genes coding PRC2 complex subunits in the embryo (EMB), endosperm (END), and seed maternal tissues (SMTs) at different days after pollination (DAP, source data in Supplemental Table S10). **B)** Black background images show representative embryo and endosperm nuclei of different C-values collected at 8 and 24 DAP. DNA was stained with DAPI (grey), H3K27me3 was immunostained (yellow), and *CEREBAs* centromeric (red) and telomeric (blue) repeats were visualized by fluorescence *in situ* hybridization and signal segmentation. White background images show H3K27me3 immunostaining fluorescence signal intensities in arbitrary units (A.U.). Scale bar = 10 μ m. Raw images and 3D image segmentation pictures of the nucleus surface, and immunostaining and FISH signals are presented in

1128 Supplemental Fig. S11. **C)** Normalized signal abundance of H3K27me3 in the endosperm (yellow)
1129 and 10-cm whole seedling (Baker et al., 2015) (magenta). The gray background signal is gene
1130 density (secondary y-axis; the full list is provided in Supplemental Fig. S12) across chromosome
1131 7H. **D)** MAplot showing genomic intervals with differential signal intensities between endosperm
1132 and 10-cm seedling (SHO). Intervals passing the threshold (Benjamini-Hochberg FDR-adjusted
1133 p-value < 0.05) are in red. The red numbers indicate genomic intervals with significant gain ($M >$
1134 0) or loss of H3K27me3 ($M < 0$). Source data are provided in Supplemental Data Set S7. **E)**
1135 Number of genes corresponding to genomic regions with significant gain and loss of H3K27me3
1136 (full list provided in Supplemental Data Set S8). **F)** Venn diagram showing the number of genes
1137 with loss of H3K27me3 in 10-cm seedlings (SHO; based on ChIP-seq data from Baker et al., 2015)
1138 and genes up-regulated at 8, 16, 24, or 32 DAP endosperm. **G)** Gene ontology (GO) term
1139 enrichment of genes with loss of H3K27me3 and significant up-regulation in the endosperm
1140 (source data in Supplemental Table S13).



1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

Figure 6. Identification and validation of imprinted genes in barley endosperm. **A)** Venn diagram showing the number of genes imprinted in single and multiple cereal species (source data in Supplemental Table S15). **B)** Heatmap of expression of barley orthologs of 21 genes imprinted in multiple cereal species in the embryo (EMB), endosperm (END), and seed maternal tissues (SMTs) at different days after pollination (DAP). The heatmap is partitioned into groups based on different expression profiles in seed tissues. MEG – maternally expressed gene, PEG – paternally expressed gene, * - imprinted in barley (source data in Supplemental Table S16). **C)** Validation of SNPs for selected imprinted genes in 8 DAP endosperm tissues by Sanger sequencing. F1

1152 reciprocal hybrids were obtained by crossing genetically distant cultivars Morex (MOR) and wild
1153 barley (HS). The maternal genotype is mentioned first in the crosses. The informative SNPs are
1154 underlined. **D)** Summary of validation of imprinted genes in barley. MBG – maternally biased gene,
1155 PBG – paternally biased gene, BEG – biallelic expressed gene, HS-B – wild barley biased, NA –
1156 information not available.

Appendix III

Dynamics of endoreduplication in developing barley seeds

Nowicka, A., **Kovacik, M.**, Tokarz, B., Vrána, J., Zhang, Y., Weigt, D., Doležel, J.,
and Pecinka, A.

Journal of Experimental Botany

DOI: 10.1093/jxb/eraa453

IF (2021): 7.298

RESEARCH PAPER

Dynamics of endoreduplication in developing barley seeds

Anna Nowicka^{1,2}, Martin Kovacik¹, Barbara Tokarz³, Jan Vrána¹, Yueqi Zhang⁴, Dorota Weigt⁵, Jaroslav Doležel¹ and Ales Pecinka^{1,*}

¹ Institute of Experimental Botany of the Czech Academy of Sciences, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-779 00 Olomouc, Czech Republic

² The Polish Academy of Sciences, The Franciszek Górski Institute of Plant Physiology, Niezapominajek 21, PL 30-239 Krakow, Poland

³ Department of Botany, Physiology and Plant Protection, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow, Al. 29 Listopada 54, PL 31-425 Krakow, Poland

⁴ Research School Biology (RSB), University of Western Australia (UWA), 35 Stirling Highway, 6009, Crawley, Perth, Australia

⁵ Department of Genetics and Plant Breeding, Poznan University of Life Sciences, 11 Dojazd St., PL-60 632 Poznan, Poland

* Correspondence: pecinka@ueb.cas.cz

Received 25 May 2020; Editorial decision 23 September 2020; Accepted 29 September 2020

Editor: Peter Bozhkov, Swedish University of Agricultural Sciences, Sweden

Abstract

Seeds are complex biological systems comprising three genetically distinct tissues: embryo, endosperm, and maternal tissues (including seed coats and pericarp) nested inside one another. Cereal grains represent a special type of seeds, with the largest part formed by the endosperm, a specialized triploid tissue ensuring embryo protection and nourishment. We investigated dynamic changes in DNA content in three of the major seed tissues from the time of pollination up to the dry seed. We show that the cell cycle is under strict developmental control in different seed compartments. After an initial wave of active cell division, cells switch to endocycle and most endoreduplication events are observed in the endosperm and seed maternal tissues. Using different barley cultivars, we show that there is natural variation in the kinetics of this process. During the terminal stages of seed development, specific and selective loss of endoreduplicated nuclei occurs in the endosperm. This is accompanied by reduced stability of the nuclear genome, progressive loss of cell viability, and finally programmed cell death. In summary, our study shows that endopolyploidization and cell death are linked phenomena that frame barley grain development.

Keywords: Cell cycle, embryo, endoreduplication, endosperm, *Hordeum vulgare*, cell death, seed development, super cycle value.

Introduction

Seeds represent an encapsulated embryonic stage unique to Angiosperms and Gymnosperms. Seed functions include protection of the embryo, plant dispersal, survival in harsh conditions, and feeding of the embryo during germination (Bewley *et al.*, 2006). High energy content and long storability make seeds one of the most valuable plant products. Cereal

Abbreviations: AL, aleurone layer; BETL, basal endosperm transfer layer; CSE, central starchy endosperm; DSB, double-strand breaks; DAG, days after germination; DAP, days after pollination; ESR, embryo-surrounding region; FDA, fluorescein diacetate; MI, mitotic index; PCA, principal component analysis; RAM, root apical meristem; SCV, super cycle value; SMT, seed maternal, tissue; TUNEL, terminal deoxynucleotidyl transferase dUTP nick end labeling.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Experimental Biology. All rights reserved.

For permissions, please email: journals.permissions@oup.com

crops provide ~60% of the global food resources in the form of seeds (Food and Agriculture Organization of the United Nations, May 2020, <http://www.fao.org/>). Knowledge of molecular processes governing seed development and maturation is thus important to support the development of crops with higher yield and quality to secure enough food for the growing population. Here, we studied the dynamics of the cell cycle during seed formation in cultivated barley (*Hordeum vulgare*, $2n=2x=14$, 5.1 Gbp/1C). The species ranks the fourth cereal in global production (Langridge, 2018), and its grains are used for feed (70%), production of alcoholic beverages (21%), and food (6%) (reviewed in Tricase *et al.*, 2018). Furthermore, there is a growing interest in using barley for biofuel, cosmetics, and molecular farming (3%) (Hicks *et al.*, 2014; Holásková *et al.*, 2018; Langridge, 2018; Tricase *et al.*, 2018). Finally, the reference genome, extensive genetic resources, and economic importance make barley an ideal temperate zone cereal model species (Mascher *et al.*, 2017).

Cereal grain development includes three major stages (Sabelli and Larkins, 2009; Sreenivasulu *et al.*, 2010; Dante *et al.*, 2014). Stage I is initiated by double fertilization and characterized by cell proliferation and a slight weight gain; stage II involves differentiation of the main tissue types and a large weight increase accompanied by accumulation of storage compounds; stage III corresponds to seed maturation, weight reduction by desiccation, and finally physiological maturation and dormancy. The three phases partially overlap with water, milk, and dough Caryopsis growth stages, respectively (Tottman *et al.*, 1979).

The diploid embryo tissues of maternal and paternal origin proliferate and differentiate into the embryonic root, shoot apical meristem, cotyledon, and plumule (primary leaf) (Dante *et al.*, 2014; Rodríguez *et al.*, 2015). Triploid ($3x$) endosperm originates from fertilization of the diploid central cell by a haploid sperm nucleus. Initially, endosperm forms a syncytium, with nuclei being pushed to the cell periphery by a central vacuole. Later, microtubules form a radial network around nuclei and the anticlinal cell wall formation marks the onset of cellularization and the beginning of differentiation into the starchy endosperm and aleurone layer (Olsen, 2001). The mature barley endosperm comprises: the central starchy endosperm (CSE), aleurone layer (AL), the subaleurone layer, the basal endosperm transfer layer (BETL), and the embryo-surrounding region (ESR) (Olsen, 2001; Sabelli and Larkins, 2009). The barley grain is covered by seed coats of maternal origin and pericarp [seed maternal tissues (SMTs)], which contain a high amount of starch, serve as feeding and protective structures, and participate in photosynthesis (Sreenivasulu *et al.*, 2010; Sabelli *et al.*, 2013; Radchuk and Borisjuk, 2014). The entire barley seed is protected by diploid hulls (the lemma and palea) of maternal origin, which remain tightly attached to the grain even after ripening (Rodríguez *et al.*, 2015).

Rapid grain development requires strict regulation of gene expression, cell cycle dynamics, and accumulation of storage molecules (Sabelli and Larkins, 2009; Dante *et al.*,

2014). Here, we estimated the dynamics of endoreduplication during grain development by measuring nuclear DNA content in the major seed tissues. Special attention was paid to endopolyploidization, a modified mitotic cycle during which G_2 -phase nuclei undergo one or more additional rounds of DNA replication—endoreduplication—leading to chromosomes with four or more chromatids (reviewed in D'amato, 1964; Nagl, 1976). Endoreduplication is common in plants. Endopolyploid cells are larger, perform specialized functions, and/or are highly metabolically active (reviewed in Chevalier *et al.*, 2011; Sabelli, 2012). In some species, endoreduplication is partially stress inducible, which may represent a bypass mechanism for tissue growth by cell expansion in cells with potentially unstable chromosomes (Adachi *et al.*, 2011; Liu *et al.*, 2015; Bhosale *et al.*, 2018).

Endoreduplication is a developmentally controlled process. The occurrence of endoreduplication in fruits or seeds of many cultivated plants suggests that it might have been positively selected during plant domestication and breeding. However, the phenotypic and molecular consequences of endoreduplication in these tissues remain unclear and are very likely to be species dependent. In cultivated tomato, the amount of DNA in nuclei of some pericarp cells reaches up to 256C (reviewed in Chevalier *et al.*, 2014) where 1C corresponds to the DNA content of an unreplicated haploid chromosome set. Suppression of two major positive regulators of endopolyploidization, *WEE1* kinase and *CELL CYCLE SWITCH PROTEIN 52 A (CCS52A)*, resulted in a reduction in fruit size (Gonzalez *et al.*, 2007; Mathieu-Rivet *et al.*, 2010), suggesting a direct role for endoreduplication in cell and organ size determination in tomato. In contrast, RNAi-induced down-regulation of S phase and the DNA replication repressor *RETINOBLASTOMA-RELATED GENE 1 (RBR1)* resulted in more endoreduplicated but, surprisingly, smaller nuclei and cells (Sabelli *et al.*, 2013). However, the total seed size was not affected, possibly due to an increased number of endosperm cells in *RBR1* knockdown maize plants. This shows that in maize grains, endoreduplication can be decoupled from endosperm growth. This still leaves an open question on the functional significance of endoreduplication during seed development. However, to what extent endoreduplication is involved in barley seed development remains unknown.

An important factor playing a role during cereal grain development is programmed cell death. At stage I, several SMTs and the nucellar projection cells die (An and You, 2004; Domínguez and Cejudo, 2014; Tran *et al.*, 2014; Radchuk *et al.*, 2018). At later stages of seed development, ESR and CSE tissues experience cell death, but remain intact in the mature grain and their content will not be mobilized until germination. Finally, a mature grain consists mainly of dead material, where only the embryo, BETL, and AL remain alive (Young and Gallie, 2000; Sreenivasulu *et al.*, 2010; Yifang *et al.*, 2012; Kobayashi *et al.*, 2013; Wu *et al.*, 2016).

Even though barley is a cereal model species, the cell cycle and endoreduplication dynamics are poorly described during its seed development. The seed develops inside hulls of maternal origin (see [Supplementary Fig. S1](#) at *JXB* online). Barley seed consists of the embryo with one maternal and one paternal genome; the endosperm with two maternal genomes and one paternal genome; and SMTs with two maternal genomes. Therefore, we investigated the dynamics of the mitotic cycle and endoreduplication in individual seed tissues during 7 weeks of barley grain development. We found a high degree of endoreduplication in endosperm and preferential elimination of endopolyploid nuclei in terminally developed tissues during the second half of the seed growth period. Collectively, this study provides comprehensive characteristics of cellular processes during the entire period of barley grain development.

Materials and methods

Plant materials and growth conditions

Six cultivars (cv.) of two-rowed spring barley (*Hordeum vulgare* subsp. *vulgare*) were used in this study: Betzes (PI 129430), Compana (PI 539111), Golden Promise (GP; PI 343079), Ingrid (PI 263574), Klages (CIho 15478), and Mars (PI 599629). The seeds were obtained from the National Small Grains Collection of the National Plant Germplasm System of the United States Department of Agriculture-Agricultural Research Service. Seeds were germinated for 3 d on wet filter paper at 25 °C in the dark. Germinating kernels were planted into 12×12 cm pots filled with a mixture of soil and sand (3/1; v/v) and grown in a climatic chamber under controlled long-day conditions (16 h day with 200 μmol m⁻² s⁻¹ light intensity and 20 °C; 8 h night with 16 °C) with 60% humidity. The day of pollination was defined using the morphology of the stigma and anthers according to the Waddington scale (W10) ([Waddington et al., 1983](#)) in the center of the spike ([Weschke et al., 2000](#)). The spikelets on the day of pollination were characterized by extended hulls, extended and widely branched stigma, and the presence of pollen grains on stigmatic hairs. Five seeds from each row corresponding to the center of the spikelet were collected at 2 d and 4 d intervals starting from anthesis until 48 days after pollination (DAP). Dry seeds were analyzed after at least 3 months of storage.

Estimation of nuclear DNA content and calculations of the super cycle value (SCV)

Nuclear DNA content was estimated using flow cytometry. Leaves collected from growing plants (from 10 d old till ~2 months old), coleoptile (a protective sheath covering the emerging shoot), and root tips dissected from seedlings at 4 days after germination (DAG) served as somatic tissue controls and were analyzed in ≥10 replicates. Each seed tissue was measured from ≥5 seeds, and the measurements were repeated three or more times on different days. Embryos were dissected using a needle and forceps under an SZX16 binocular microscope (Olympus). Five embryos were collected and used as one sample at 4 and 6 DAP. Cell nuclei were isolated from 4–8 DAP embryos by homogenizing them with a pestle in a 1.5 ml Eppendorf tube containing 300 μl of Otto I solution (0.1 M citric acid, 0.5% Tween-20). The crude suspension was filtered through a 50 μm nylon mesh (Sysmex-Partec) and stained after adding 600 μl of Otto II solution (0.4 M Na₂HPO₄·12H₂O) containing 2 μg ml⁻¹ DAPI. At 12 DAP, embryos and the other tissues were homogenized with a razor blade in a Petri dish containing 500 μl of Otto I solution and stained with

1 ml of Otto II solution containing DAPI. Nuclear samples were analyzed using a CyFlow Space flow cytometer (Sysmex-Partec) equipped with a UV-LED diode array. At least 5000 particles were acquired per sample, using a log₃ scale. Histograms were evaluated by the FloMax program (Sysmex-Partec) and interpreted as shown in [Supplementary Fig. S4](#).

The percentages of nuclei in the embryo, endosperm, and SMTs were calculated based on the number of measured particles (counts) in (i) whole seeds and (ii) seeds containing only endosperm and SMTs (embryos were removed). Based on (ii), we estimated the proportion of endosperm and SMT nuclei, and next we also used (i) values to calculate the percentage of embryo nuclei according to the formula: embryo (%) = 100 – endosperm (%) – SMTs (%).

To estimate the amount of endoreduplication in individual samples, we introduce a new formula termed the super cycle value (SCV). Compared with other existing approaches, this formula estimates the frequency of endoreduplicated nuclei more conservatively (see the Results for more details). For the SCV, 8C in the diploid and 12C in the triploid tissues were considered as the first endoreduplicated levels. For diploid tissues, the SCV = [(n 2C×0) + (n 4C×0) + (n 8C×1) + (n 16C×2)] / (n 2C + n 4C + n 8C + n 16C) was calculated; n = number of counts per given C-value content. For triploid endosperm, 3C and 6C received a value of 0, 12C received a value of 1, and 24C received a value of 2.

Isolation of nuclei and TUNEL assays

Around 100 embryos were manually dissected from 8 DAP seeds using an SZX16 binocular microscope (Olympus). Precisely 80 roots of seedlings at 4 DAG were cut ~1 cm from the apex. Both types of tissues were rinsed immediately in 10 mM Tris buffer pH 7.0, fixed in 2% (v/v) formaldehyde/Tris buffer for 30 min at 4 °C, and then washed 3× 5 min in Tris buffer at 4 °C. Embryos were homogenized with a pellet pestle in 1.5 ml Eppendorf tubes in LB01 buffer ([Doležel et al., 1989](#)). Root apices were excised (~1 mm from the tip) and homogenized in LB01 buffer for 13 s at 15 000 rpm using a Polytron PT1300D homogenizer (Kinematica AG). The crude homogenates were filtered through a 50 μm pore size mesh. From at least 30 dissected embryos older than 8 DAP, peeled seeds without the embryo were rinsed in Tris buffer, pre-fixed by vacuum infiltration in 4% formaldehyde/Tris buffer for 20–30 min at 4 °C, followed by fixation and washing as above. Tissues were chopped with a razor blade in LB01 buffer on a Petri dish and filtered through a 50 μm nylon mesh. Suspensions of nuclei were stained with 2 μg ml⁻¹ DAPI.

Approximately 500 nuclei for each determination of DNA content were sorted into a drop of LB01 buffer on microscopic slides using a FACSAria II SORP flow cytometer and sorter (BD Biosciences, Santa Clara, CA, USA), air-dried, and stored at –20 °C until use. Double-strand breaks (DSBs) were detected by TUNEL (terminal deoxynucleotidyl transferase dUTP nick end labeling) assay using the *in situ* cell death detection kit with a fluorescein-dUTP label following the manufacturer's instructions (11684795910, Roche). Nuclei which had been flow-sorted from root tips were used as controls. The TUNEL negative (–) background noise control was prepared by omitting terminal deoxynucleotidyl transferase treatment, while the high signal control nuclei [TUNEL positive (+)] were digested with 100 U ml⁻¹ DNase I for 10 min at 21 °C. Nuclei were counterstained with DAPI (1 μg ml⁻¹) in Vectashield (H-1000, Vector Laboratories). Analysis of fluorescence signals was performed with an AxioImager Z2 (Zeiss) epifluorescence microscope equipped with a DSD2 spinning disk confocal imaging module and monochromatic Zyla 4.2 camera (both Andor). Z-stacks were captured separately for each fluorochrome using the appropriate excitation and emission filters with the IQ3 (Andor) system. At least 50 nuclei were evaluated per individual sample, each in three biological replicates. TUNEL (–) nuclei were characterized by pale-green fluorescence signals, while TUNEL (+) nuclei had bright green signals ([Ghasemzadeh et al., 2015](#); [Palermo et al., 2017](#)).

Fluorescein diacetate and Evans blue staining assays

Hulls were removed from fresh seeds with forceps. Peeled seeds were cut to half in phosphate-buffered saline (PBS) along the longitudinal and transverse axes with a razor blade. At least 20 seeds were tested per experimental point. One half of the sample was stained with 2 mg l^{-1} fluorescein diacetate (FDA, F7378, Sigma-Aldrich) in PBS for 15 min in the dark, and then washed twice for 10 min with PBS (Kobayashi *et al.*, 2013). The other half of the samples were stained in 0.1% (w/v) Evans blue (314-13-6, Sigma-Aldrich) for 2 min. Stained sections were washed twice for 10 min with distilled water (Wu *et al.*, 2016). Transverse and sagittal sections of samples were observed with an SZX16 binocular microscope (Olympus) equipped with a GFP (green fluorescent protein) filter. Images were captured with a Regita 1300 QImaging camera and QCapture $\times 64$ software (Olympus) using the same exposure times. The sagittal section area of whole peeled seed, endosperm, and embryo was measured in ImageJ (Schneider *et al.*, 2012).

Light microscopy and determination of seed growth parameters

For phenotypic analysis, seeds from 0 (ovary) to 48 DAP and dry seeds were peeled off, weighed with an analytical scale (Sartorius), and then photographed using an SZX16 binocular microscope (Olympus) equipped with a Regita 1300 QImaging camera and QCapture $\times 64$ software (Olympus). A minimum of 50 seeds were used per individual replicate. To analyze transverse and sagittal plans, the seeds were cut with a razor blade along the longitudinal and transverse axis, respectively. Seed length, width, and sagittal section area were measured using ImageJ. The growth rate (GR) was calculated according to the formula: $\text{GR} (\%) = (x_2 - x_1) / x_2 \times 100$, where x_1 was the previous time point measurement (e.g. at 0 DAP) and x_2 was the next time point measurement of seed growth (e.g. at 2 DAP).

Confocal microscopy, morphology of the nuclei, and calculation of the mitotic index

Whole seeds were fixed in acetic acid/alcohol (1:3) for 2–4 h and stored in 70% ethanol at -20°C until use. The transverse slides were cut with a razor blade and stained with $2 \mu\text{g ml}^{-1}$ DAPI in Vectashield. The images were captured with an AxioImager Z2 epifluorescence microscope (Zeiss) equipped with a DSD2 confocal module, monochromatic camera Zyla 4.2, and IQ3 program (all Andor). The photos were processed with Imapris 9.2 software (Bitplane, Oxford Instruments). Three to five slides, each corresponding to one endosperm, were evaluated per experimental point. In total, ~ 600 cells per slide were scored for quantification of mitotic divisions. The mitotic index (MI) was calculated according to the formula: $\text{MI} (\%) = \text{cells in mitosis} / \text{total number of cells} \times 100$. For calculation of the proportion of deformed nuclei, flow-sorted nuclei were used. Three to four slides were evaluated per time point, each with at least 250 nuclei per individual C-value.

Statistical analysis

Statistical significance was examined by one-way ANOVA and post-hoc comparison by Tukey's multiple range test ($P \leq 0.05$) using Minitab v. 18 (Minitab, LLC) software. Principal component analysis (PCA) and correlation analysis were performed with Statistica v. 13 (StatSoft) and RStudio programs, respectively.

Results

Endoreduplication occurs in barley endosperm and SMTs

To assess the dynamics of barley seed growth, we monitored grain weight, length, and width of peeled (after manual hull

removal) growing seeds of the cv. Compana from 0 (ovary) to 48 DAP and at 3 months after harvest (Fig. 1; Supplementary Fig. S2). From 0 to 4 DAP, the seeds increased their size and weight mainly due to the growth of SMTs. From 6 DAP onwards, sagittal sections revealed endosperm expansion and acceleration of seed growth (Supplementary Fig. S2C). At 12 DAP, seeds reached the maximum length of ~ 9 mm, and from 12 to 28 DAP, the weight of the caryopsis increased mostly because of seed widening, presumably due to CSE expansion. During this period, the embryo also grew rapidly (Supplementary Fig. S2B, C). At 32 DAP, the seed reached maximum weight, length, and width of 0.076 g, 9.37 mm, and 4.07 mm per grain, respectively (Fig. 1B). Subsequently, the seeds started desiccating, leading to their reduction in size and weight.

To test the relationship between seed parameters and the degree of endopolyploidy, we estimated nuclear C-values using flow cytometry. Somatic tissue controls represented by the root apical meristem (RAM), coleoptile, and young leaves contained 96–97% 2C and 4C nuclei, representing G_1 - and G_2 -phase nuclei, respectively, and only $\leq 3.5\%$ 8C endopolyploid nuclei (Supplementary Fig. S3). The RAM tissues had an almost equal ratio of 2C (46.7%) and 4C (48.8%) nuclei, indicating a high cell division activity. In contrast, the leaves contained 80% of 2C nuclei, suggesting a lower cell division rate. However, the proportion of endoreduplicated nuclei remained low in somatic tissues, as indicated by a very low SCV (0.04–0.06).

We measured C-values of nuclei isolated from (i) dissected embryos and (ii) a mixture of SMTs and endosperm (Fig. 2A; Supplementary Figs S4, S5). This allowed us to estimate the relative proportion of embryo, SMTs, and endosperm nuclei during grain development (Fig. 2B). Directly after pollination, SMTs represented the majority of nuclei, but the proportion of endosperm nuclei quickly increased and reached the majority (58%) of all seed nuclei at 8 DAP. Afterwards, the proportion of endosperm nuclei started decreasing, and the percentage of embryo nuclei continuously increased up to 34% at 44 DAP. Finally, the dry barley grain contained 31% of nuclei from the embryo, 41% from SMTs, and 28% from the endosperm. Interestingly, the low proportion of endosperm nuclei contrasts with the fact that this tissue makes most of the seed mass (Fig. 1A).

Next, we estimated C-values of nuclei in individual tissues (Fig. 2C; Supplementary Fig. S5). Embryos contained $\sim 75\%$ of 2C nuclei during the first 12 DAP and subsequently their proportion decreased to $\sim 60\%$. Reverse dynamics were observed for 4C nuclei. The endoreduplicated 8C nuclei reached 9% at 16 DAP and remained at this level. SMTs showed large dynamics of 2C and 4C nuclei starting at 80% (2C) and then oscillating between 25% and 50%. The amount of endoreduplicated 8C and 16C SMT nuclei reached 20% at 44 DAP, but then strongly decreased during seed desiccation. Endosperm contained the highest percentage of endoreduplicated nuclei (up to 40%), and 2-day-old syncytium already had 8% of endoreduplicated 12C and 24C nuclei (Fig. 2C; Supplementary Fig. S5). The frequency of endoreduplicated (12C and 24C) nuclei continuously

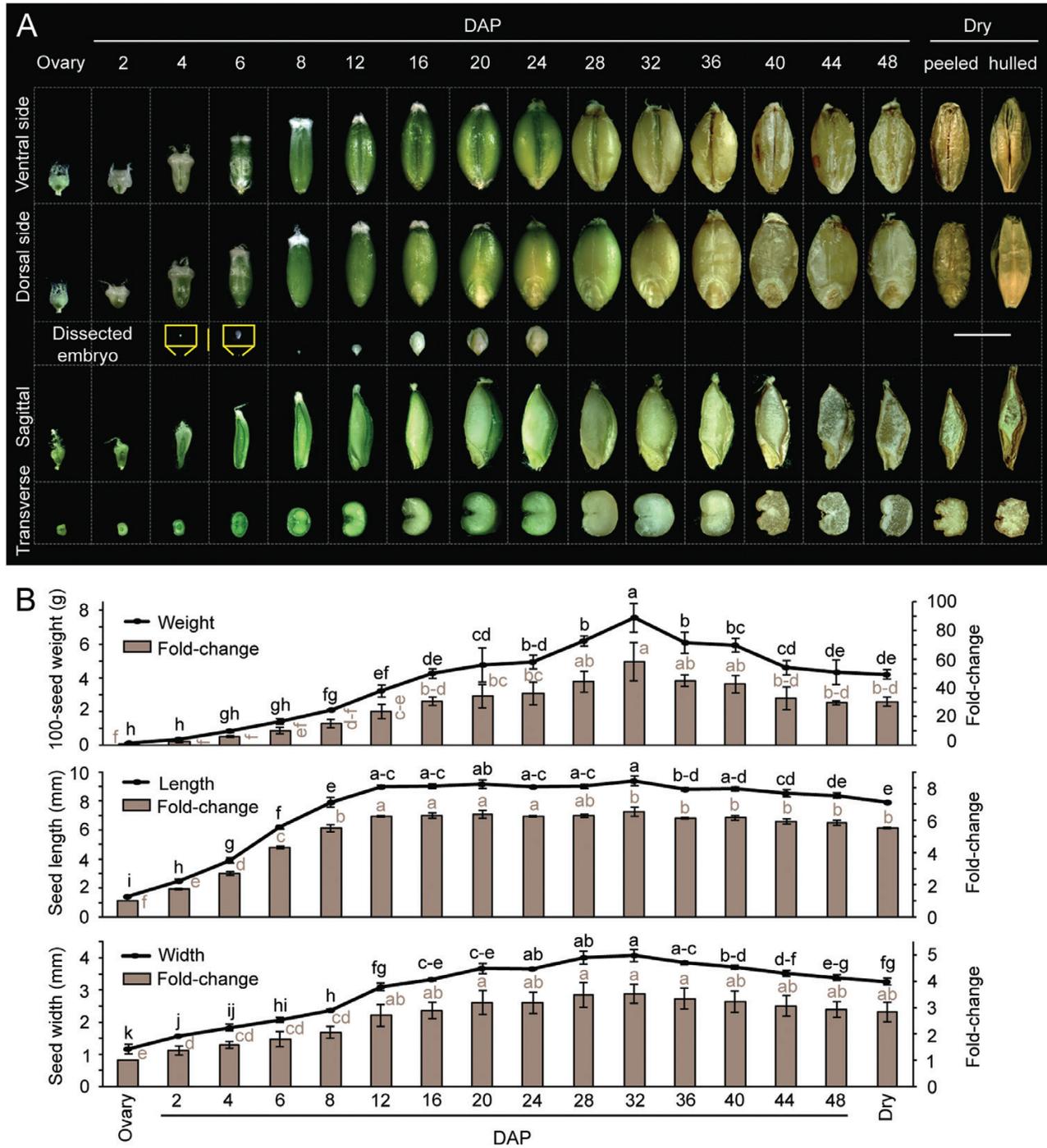


Fig. 1. Phenotype of developing barley seeds of cv. Compana. (A) Grain developmental series from 0 (ovary) to 48 DAP and in dry seeds. For analysis of sagittal and transverse planes, the seeds were cut in half. Scale bar=5 mm. The yellow insets show early-stage embryos with scale bar=500 μ m. (B) Quantitative data for 100-seed weight, seed length, and seed width calculated as absolute values (black line, left y-axis) and relative to the ovary (columns, right y-axis). Data represent the means (\pm SD) from three biological replicates, each with at least 50 seeds. ANOVA was carried out separately for each tested parameter. Values marked with the same letter do not differ according to Tukey’s multiple range test ($P \leq 0.05$).

increased up to ~50% from 20 to 28 DAP. Subsequently, the frequency of endoreduplicated nuclei rapidly decreased, and endosperm tissues of dry seeds contained only 7% of such nuclei. This indicated a programmed and preferential loss of endopolyploid

nuclei during the terminal stages of barley endosperm development. Finally, we estimated C-values of nuclei from hulls, but this tissue showed little variation and its profile strongly resembled that of leaves [Supplementary Figs S3, S6](#)).

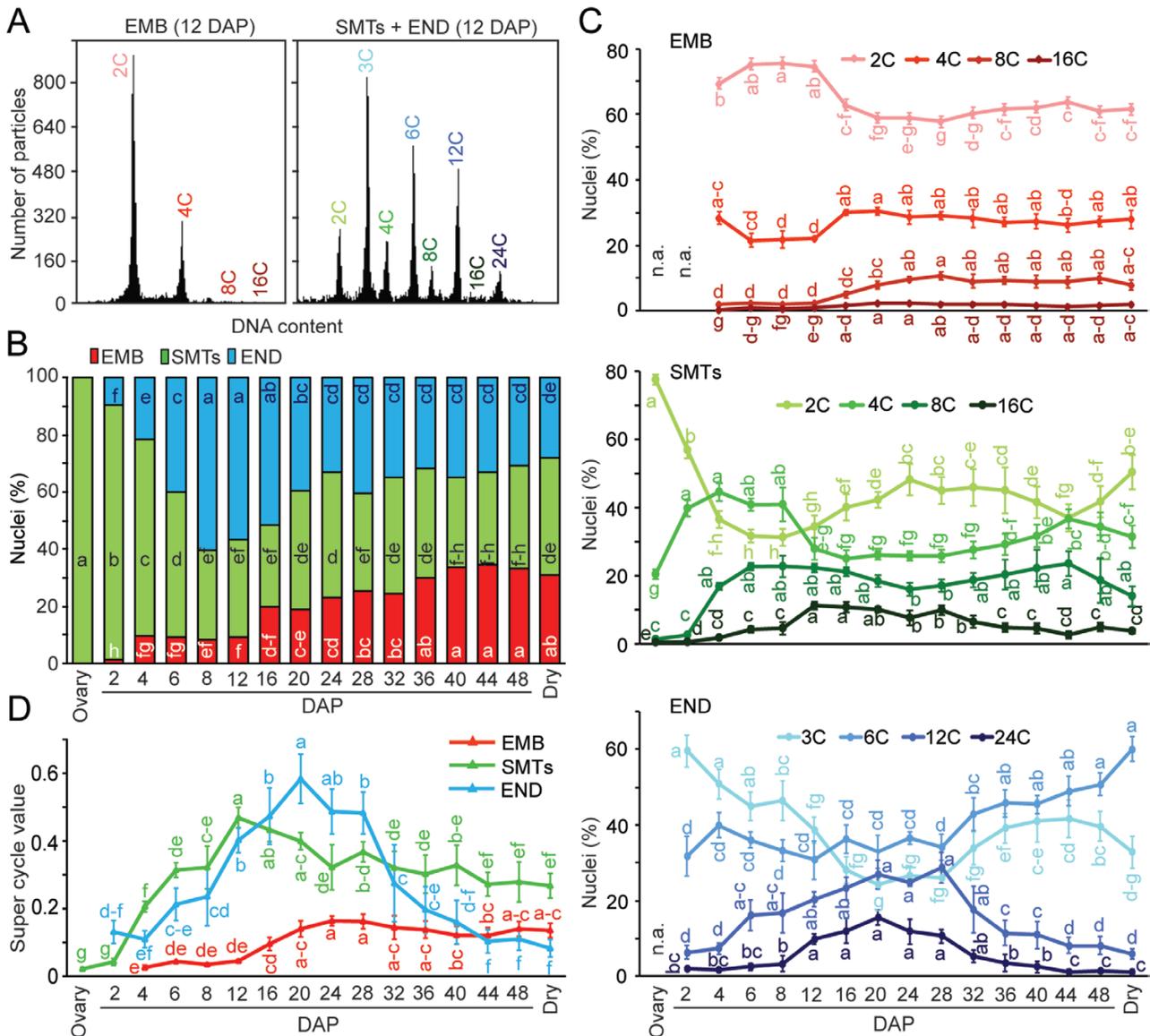


Fig. 2. Nuclear C-values in developing barley seed tissues. (A) Representative histograms of nuclear DNA contents obtained from the dissected embryo (EMB; left graph) and seed maternal tissues and endosperm (SMTs and END, respectively; right graph) at 12 DAP. Representative histograms of nuclear DNA contents obtained at 0–24 DAP and dry seeds are shown in [Supplementary Fig. S4](#). (B) Cumulative percentage of nuclei from the major seed tissues. Stacks marked with the same letter do not differ within the respective group according to Tukey's test ($P \leq 0.05$). (C) Percentage of nuclei with different C-values in major seed tissues at different DAP. Data are the means (\pm SD) from three biological replicates, each with at least five individual measurements. ANOVA was carried out separately for each C-value. Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$). n.a. indicates that samples were not analyzed. (D) Super cell cycle values during seed development based on the data from (C). ANOVA was carried out as described for (C).

There are multiple methods for quantifying nuclear C-values. This includes the mean C-value, which is an indicator describing the average C-value of all measured nuclei ([Rewers *et al.*, 2009](#)). However, the mean C-value does not allow a direct comparison of the tissues with different ploidy such as the diploid embryo and triploid endosperm. Here, the cell cycle value (i.e. endoreduplication index), calculated and averaged from the number of endoreduplication cycles for all nuclei, is a better solution ([Barow and Meister, 2003](#); [Parker](#)

et al., 2018). Nevertheless, the existing cell cycle value formulae consider 4C nuclei as already endoreduplicated once. Although some 4C nuclei might already be programmed for endoreduplication, others will not be. The latter may be particularly true for embryonic and RAM tissues used in this study, where many cells with 4C nuclei will be regularly cycling. Therefore, we introduced a new formula called the SCV (see the Materials and methods for details), which conservatively compares the degree of endoreduplication in individual tissues

irrespective of their basic ploidy (Fig. 2D; Supplementary Figs S3C, S6C). For diploid tissues, SCV=0 means that all nuclei are 2C or 4C while SCV=1 means that the nuclei are on average 8C. The SCV of 4 DAP embryos was 0.03 and remained stable until 12 DAP, then gradually increased to ~0.15 and reached a plateau. The SCV of SMTs started from 0.02, constantly increased to 0.47 at 12 DAP, and then gradually decreased to 0.27 in dry seeds. The endosperm SCV curve had a profile similar to that of SMTs, but started from 0.13 at 2 DAP, peaked at 0.58 at 20 DAP, and then decreased to 0.08 in dry seeds.

The dynamics of endoreduplication in developing barley seeds prompted us to analyze what the degree of correlation is between endoreduplication (i.e. SCV) and seed growth parameters (i.e. seed weight, length, and width). We performed Pearson correlation coefficient analysis and visualized it using a heat map (see Supplementary Fig. S7). When we looked at the endocycle during embryo development, we observed a strong positive correlation of SCV with seed weight and width ($r=0.78$ and 0.83 , respectively). For SMTs, the strongest correlation existed between SCV and seed length ($r=0.83$). Surprisingly, in developing endosperm, only a moderate correlation appeared between compared parameters ($r=0.29-0.46$). These correlations have an indicative value (e.g. the phase when seeds greatly gain weight and width is the time when the embryo undergoes most of its endoreduplication), but do not describe causality.

To summarize, barley embryo shows a moderate degree, while SMTs and endosperm show high degrees of endoreduplication. Surprisingly, the amount of SMTs and endosperm endopolyploid nuclei becomes strongly reduced during seed maturation and drying.

Endoreduplicated nuclei accumulate DNA damage

The loss of endoreduplicated nuclei at later stages of seed development raised the question of their fate. We hypothesized that they may have reduced genome stability and are removed. Therefore, we performed TUNEL assays using flow-sorted nuclei from the embryo, SMTs, and endosperm at 4, 8, 16, and 24 DAP (Fig. 3). In the TUNEL assay, 3' termini of DNA are labeled by tagged nucleotides using terminal deoxynucleotidyl transferase, and then detected. The signals indicate the presence of DNA DSBs. The 2C and 4C nuclei from the RAM at 4 DAG were used as somatic tissue controls. The background negative signal controls were represented by 2C RAM nuclei without the terminal deoxynucleotidyl transferase treatment. In contrast, damaging the DNA of the same nuclei using 100 U ml^{-1} DNase I for 10 min resulted in 80% of TUNEL (+) nuclei. For embryos, in 2C and 4C nuclei we observed ~15% TUNEL (+) cells. However, in 8C nuclei isolated from 24 DAP embryos, the accumulation of DSBs increased with increasing amounts of DNA [49% TUNEL (+) nuclei]. The nuclei of SMTs showed ~40% TUNEL (+) signal in 2C and 4C up to 16 DAP. However, for all C-values at 24 DAP, amounts of DSBs

increased and ~60% of TUNEL (+) nuclei appeared. For endosperm, 3C, 6C, and endoreduplicated 12C nuclei accumulated DSBs in increasing C-value- and age-dependent manners.

This demonstrates that endoreduplication is associated with reduced genome stability in barley seed tissues and the greatest damage was observed in the late SMTs and endosperm nuclei.

Seed maternal and endosperm tissues undergo cell death

Studies in maize (Young and Gallie, 2000), rice (Kobayashi *et al.*, 2013; Wu *et al.*, 2016), bread wheat (Yifang *et al.*, 2012), and triticale (Li *et al.*, 2010) identified that cell death is an essential process during cereal seed development. To detect viable and dead cells in developing barley grains, we performed FDA and Evans blue staining (Fig. 4A, B; Supplementary Figs S8, S9). FDA is hydrolyzed in living cells into the green fluorescent fluorescein, which indicates viable cells (Schnürer and Rosswall, 1982). In turn, Evans blue points to a loss of membrane integrity by dyeing the intracellular space of non-viable cells.

Cell death followed a specific pattern in developing barley seeds (Fig. 4A, B; Supplementary Fig. S8). During early seed development (0–8 DAP), we observed fluorescein signals in the middle part of SMTs (pericarp, and seed coats), but not in the embryo sac or in the developing endosperm. From 12 DAP, the fluorescence appeared in the endosperm and its intensity increased during CSE formation until 32 DAP, suggesting high metabolic activity of this tissue. From 36 DAP, the fluorescence intensity started decreasing from the periphery towards the central zone of the CSE, indicating reduced viability of cells in this outer region. During maturation, the area of fluorescein-labeled cells further shrank. This pointed to a progressive loss of cell viability during seed desiccation (Fig. 4A; Supplementary Fig. S8). Evans blue staining revealed a complementary pattern. During early seed development, we detected increasing regions of blue staining in the top (in the region surrounding the brush) and bottom parts of maternal tissues, but not in the longitudinal elongation zone. Staining in the endosperm was first detected in a few dispersed cells at 8 DAP, and the number of stained cells increased over time. The AL was free of stain until 48 DAP, but some signal could be detected in AL of dry seeds (Supplementary Fig. S9). No staining was observed within the embryo at any stage of seed development (Fig. 4B; Supplementary Fig. S9).

To verify a potential link between cell death and the loss of endoreduplicated nuclei, we investigated the morphology of CSE nuclei after DAPI staining of sliced seeds (Fig. 4C; Supplementary Fig. S10). The nuclei frequently divided at 6 and 8 DAP (MI ~20%) but the frequency of mitoses decreased over time and no cell divisions were observed at 20 DAP or after this time point (Fig. 4D). From 16 DAP, the density of nuclei in the central part of the endosperm progressively decreased (indicating their elimination by cell death) and many larger nuclei were progressively degenerating, as indicated by

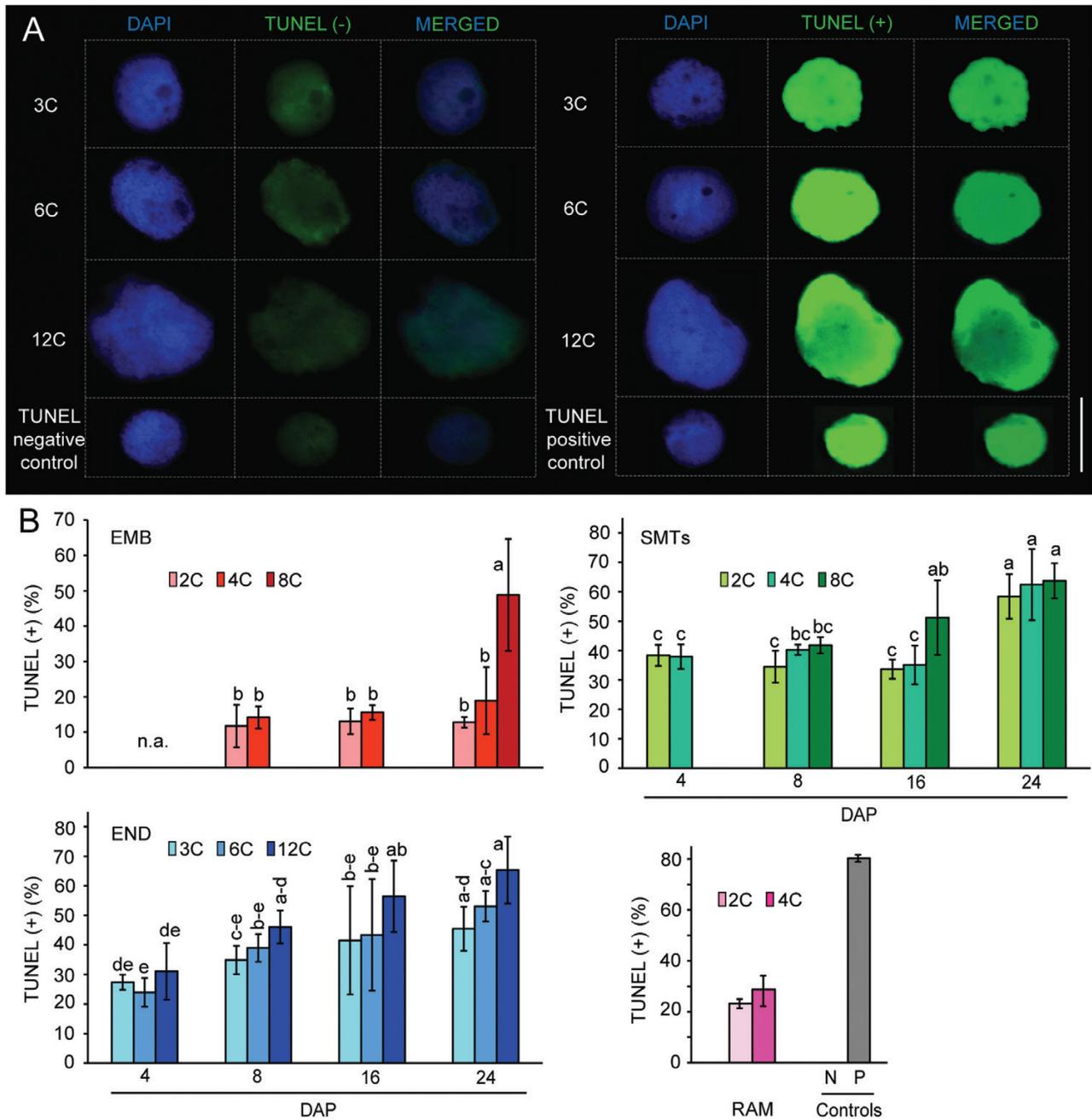


Fig. 3. Analysis of genome integrity using TUNEL assay. (A) Representative TUNEL negative (-) and positive (+) nuclei isolated from 24 DAP endosperm tissues. Nuclei were counterstained with DAPI (blue), and DNA double-strand breaks (DSBs; green) were detected using TUNEL assay. All nuclei were photographed with the same settings. Scale bar=10 μ m. (B) Quantification of DNA damage in nuclei of different C-values and from different tissues. The TUNEL assay control was represented by RAM nuclei. The negative control (N) was prepared without terminal deoxynucleotidyl transferase and its fluorescence was considered as the background signal. Positive control (P) nuclei were treated with DNase I. Four DAP embryos were not analyzed (n.a.) for technical reasons. Values are the means (\pm SD) from three biological replicates, each with at least 50 individual nuclei. ANOVA was carried out for all C-values, time points, and tissues within the plot. Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).

their deformed shape (Fig. 4E, F), and finally only remnants of nuclei (pieces of chromatin) were observed. Conversely, the nuclei in the three layers of AL cells started to be clearly visible at ~12 DAP and persisted until maturation.

This showed that some SMTs and endosperm cells undergo cell death from as early as 2 and 8 DAP, respectively; endoreduplicated CSE nuclei are the first to be degraded during seed maturation, while AL nuclei remain alive.

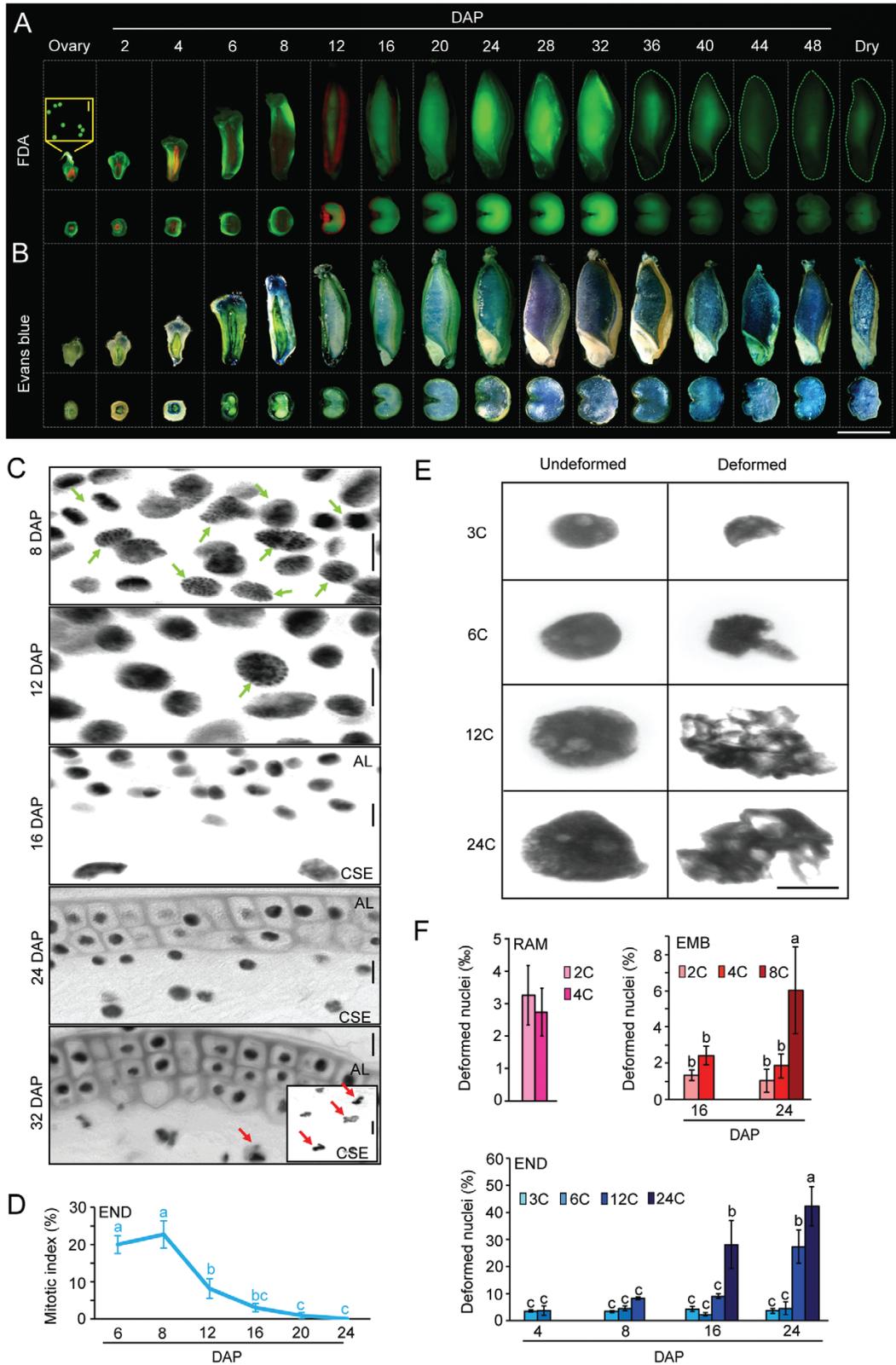


Fig. 4. Time-course study of cell viability in developing seeds. (A) Representative sagittal and transverse sections of barley seeds showing fluorescein signals (green fluorescence) after staining with 2 mg l⁻¹ fluorescein diacetate (FDA). The inset shows FDA-stained pollen attached to the stigma, with scale bar=50 μm. Dashed white lines show the whole sagittal section area of the seed. A negative control without FDA staining is shown in

Dynamics of endoreduplication in six two-row barley cultivars

To challenge the validity of our findings in a broader context, we compared the data from the US cv. Compana with five European two-row spring barley cultivars: Mars (Czech Republic), Betzes, Klages (both Germany), Ingrid (Sweden), and GP (UK) (Fig. 5A). The cultivars differed as to their dry seed weight and morphology (Fig. 5B; Supplementary Fig. S11).

First, we estimated whether there was a statistically significant difference between individual C-values at different DAPs for diploid tissues (embryo and SMTs) and triploid endosperm for all cultivars from 4 DAP until 24 DAP and in dry seeds, and plotted the *P*-values to reveal the major trends (Fig. 5C; Supplementary Figs S12 and S13 show the source data for the statistics). There were few differences until 8 DAP for the diploid tissues. From 12 DAP, the intercultivar variation increased for 4C, 8C, and 16C nuclei and continued mainly with the 16C nuclei at 16–24 DAP. From 24 DAP, the intercultivar differences shifted towards the nuclei with lower C-values (2C and 4C) and also persisted in dry seeds. However, it should be noted that the differences were often caused by one or two cultivars (Supplementary Fig. S12); nevertheless, these data point to an existing variation in endoreduplication during barley embryo and/or SMT development.

During endosperm growth, the frequencies of almost all C-values varied in the cultivars from 12 to 24 DAP (Fig. 5C; Supplementary Fig. S13). The frequency of 3C nuclei differed from 24% to 48% at 12 DAP (Compana and GP versus Ingrid, respectively) and from 28% to 39% at 24 DAP (Mars versus Klages, respectively). The variation in 6C nuclei was mainly due to the cv. Mars, which had more 6C nuclei than other genotypes (e.g. 49% versus ~34% in other cultivars at 16 DAP). For the endoreduplicated endosperm nuclei (12C and 24C), the major differences occurred from 12 to 20 DAP, corresponding to the time of rapid endosperm expansion, the endoreduplication maximum, and then its decrease. Interestingly, there were no significant differences in C-values in the endosperm of dry seeds between the cultivars, suggesting that it is rather the dynamics of the whole process that were affected. To provide an average picture, we calculated the mean percentage of nuclei of individual C-values for all genotypes,

which summarizes the trends in DNA content changes during barley seed development (Supplementary Figs S12C, S13C).

Next, we calculated the SCV for the cultivars (Fig. 5D; Supplementary Fig. S14A). For embryos/SMTs, it varied significantly from 12 DAP to 24 DAP, and in dry seeds. For instance, at 16 DAP, the SCV ranged from 0.19 (Klages) to 0.32 (GP). In endosperm tissues, the SCV differed from 12 to 20 DAP. To gain insight into the relationships among the cultivars, we performed PCA using SCV data (Supplementary Fig. S14B). The first component (PC1) separated samples based on DAP and revealed which time points of diploid tissues and endosperm development were more similar. The second component (PC2) showed the relationships between the cultivars. In the mixture of diploid tissues, the PCA revealed the relationship between cultivars; Mars, Klages, and Compana formed one group, and Betzes, GP, and Ingrid were far from them. Subsequently, the SCV analysis in endosperm revealed two groups; the first group included the interval from 4 to 8 DAP (the mean SCV ranged from 0.10 to 0.29), and the second group contained the period from 16 to 24 DAP (the mean SCV ranged from 0.26 to 0.18), excluding 12 DAP (mean SCV=0.33) and dry seed (mean SCV=0.17) (Fig. 5D; Supplementary Fig. S14B).

Collectively, these results demonstrate genetic variation in the endosperm endocycle kinetics during barley seed development.

Discussion

Barley caryopsis development has been explored extensively from the biochemical and metabolic perspective (Radchuk *et al.*, 2009; Sreenivasulu *et al.*, 2010; Peukert *et al.*, 2014). In contrast, cell cycle dynamics and endoreduplication during barley seed formation have received much less attention. This motivated us to investigate the spatiotemporal dynamics of endoreduplication and loss of nuclei in the embryo, endosperm, and SMTs from pollination until dry seeds (Fig. 6).

Endoreduplication is receiving increasing attention as one of the mechanisms that leads to cell differentiation and specialization (Bhosale *et al.*, 2018). To quantify endoreduplication frequency in diploid and triploid tissues, we introduced the new concept of SCV. Compared with previous studies, the formula for the SCV is more conservative and considers as endoreduplicated only from 8C nuclei in diploid and from

Supplementary Fig. S8. (B) Representative sagittal and transverse sections of developing barley seeds stained with 0.1% Evans blue. Scale bar for (A) and (B)=5 mm. The seeds shown are representative of at least 20 individual stained seed. (C) DAPI staining of nuclei and cell membranes in barley endosperm tissues. Dark spots indicate nuclei and dark squares indicate the area of cells, as DAPI is a membrane-permeant dye. Green arrows mark dividing nuclei. Red arrows mark degenerating nuclei and nuclear residues. CSE=central starchy endosperm, AL=aleurone layer. Scale bar=20 μ m. More time points are presented in Supplementary Fig. S10. (D) The frequency of dividing nuclei in developing endosperm. Every time point represents the mean of at least three independent measurements (slides), each with at least 600 counted nuclei. Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$). (E) Representative undeformed and deformed nuclei isolated from 24 DAP endosperm tissues. Nuclei were counterstained with DAPI (gray). Scale bar=20 μ m. (F) Quantification of deformed nuclei of different C-values and from different tissues. Values are the means (\pm SD) from three biological replicates, each with at least 250 nuclei per individual C-value. ANOVA was carried out separately for each tissue. There were no significant differences for the RAM ($P=0.486$). Values marked with the same letter do not differ according to Tukey's test ($P \leq 0.05$).

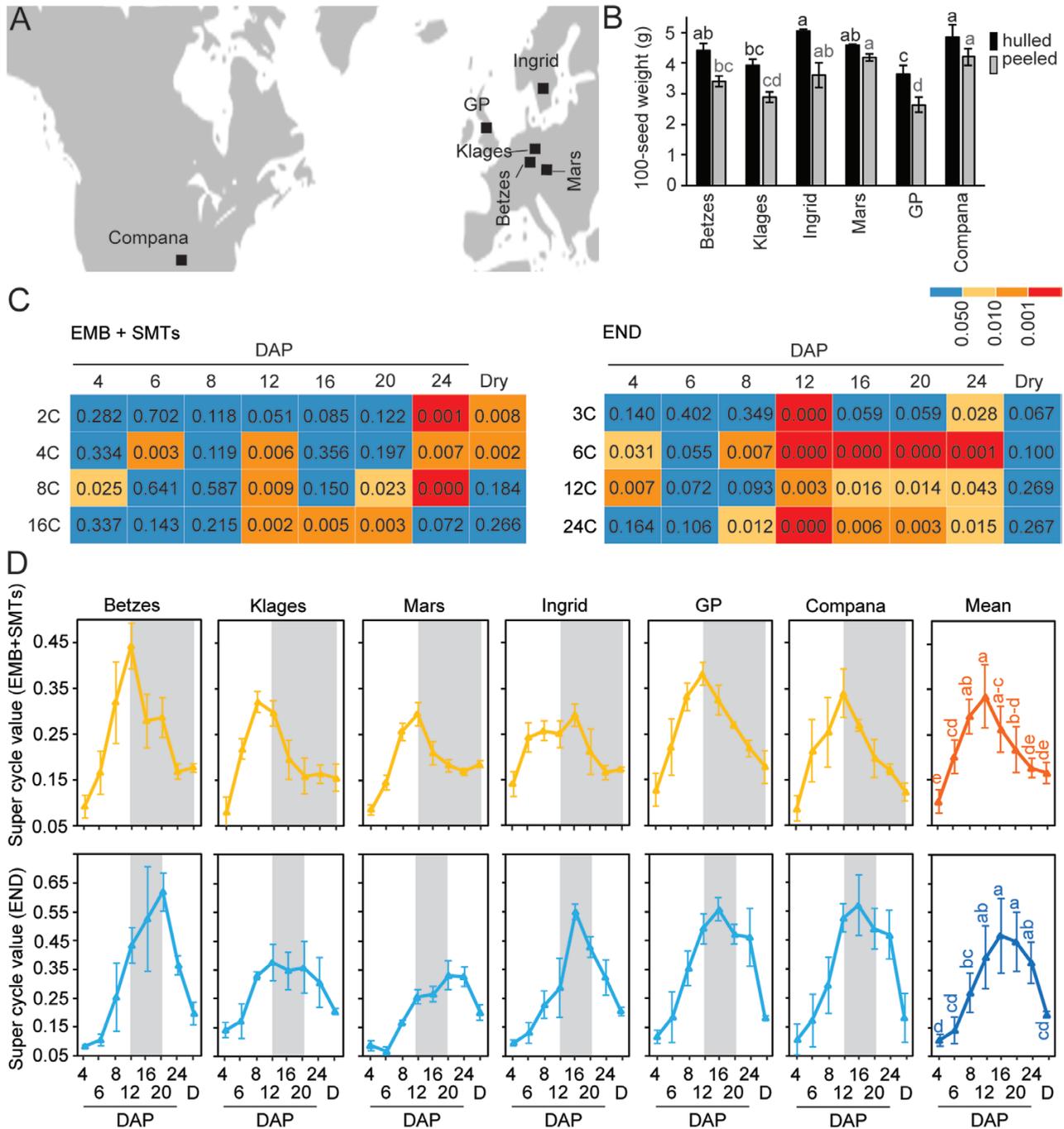


Fig. 5. Estimation of nuclear DNA contents in seed tissues of six two-row barley cultivars. (A) Geographical origin of the used cultivars. GP=Golden Promise. (B) 100-seed weight of hulled and peeled dry kernels. Data are the means (\pm SD) from three biological replicates, each with at least 100 seeds. ANOVA was carried out separately for hulled and peeled seeds. Values marked with the same letter do not differ according to Tukey's test ($P < 0.05$). (C) A heat map of the P -values of AVOVA between cultivars for individual C-values in diploid and triploid tissues at a given DAP. Source data are shown in [Supplementary Figs S12 and S13](#). (D) Super cycle values of the embryo and seed maternal tissues (EMB+SMTs) versus endosperm (END) at a given DAP. Values are the means (\pm SD) from three to five biological replicates, each with at least three individual measurements. ANOVA was carried out separately for each time point between cultivars, and gray shading shows the periods of significant differences between cultivars. P -values are shown in [Supplementary Fig. S14](#). Mean=mean percentages of super cycle value calculated between all cultivars. Data are the averages (\pm SD) between cultivars. Points marked with the same letter do not differ according to Tukey's test ($P < 0.05$).

12C nuclei in triploid tissues. We found that barley endosperm tissues underwent one to two rounds of endoreduplication resulting in 12C and 24C nuclei, respectively. This is similar to wheat (Chojecki *et al.*, 1986) and rice (Kobayashi, 2019). More endoreduplication cycles were found in the endosperm of sorghum with four (Kladnik *et al.*, 2006) and in maize with up to seven rounds of endocycles (Engelen-Egles *et al.*, 2000; Sabelli and Larkins, 2009). In barley endosperm, the major wave of endoreduplication started from ~6 DAP and increased linearly to 20 DAP, which corresponds to seed developmental stage II characterized by production of storage components (Dante *et al.*, 2014). This is similar to sorghum with the onset of endosperm endoreduplication at 5 DAP (Kladnik *et al.*, 2006), but earlier than that at 10 DAP in maize (Kowles *et al.*, 1990; Leiva-Neto *et al.*, 2004; Sabelli and Larkins, 2009) or 12 DAP in wheat (Chojecki *et al.*, 1986). Some differences also occur in the peak of endosperm endoreduplication between cereals. Endoreduplication peaked at 20 DAP under our experimental conditions, while it peaked at 15–18 DAP in maize and at 24 DAP in wheat (Chojecki *et al.*, 1986; Sabelli and Larkins, 2009).

How much the dynamics are affected by environment or basal ploidy remains unknown. A common observation for cereals is that endoreduplication is initiated in the endosperm central region from where it spreads toward the periphery, excluding the AL (Kladnik *et al.*, 2006; Sabelli, 2012; Kobayashi, 2019). Here, barley is an exception as it contains endoreduplicated cells also in the AL (Keown *et al.*, 1977). We detected one to two rounds of endocycles (leading to 8C and 16C nuclei) also in SMTs and embryos, and their endoreduplication peaked at ~12 and 20 DAP, respectively. Hence, our data support the notion that endoreduplication is associated specifically with the reproductive organs in grasses (Sabelli, 2012). Barley somatic tissues generally have a low degree of endopolyploidy and most cells are found at the G₁ phase of the cell cycle (Trafford *et al.*, 2013).

Despite the fact that endoreduplication appears to be ubiquitous among cereals, its role in seed development remains poorly understood (Sabelli, 2012). We detected a moderate to high positive correlation between the SCV of the embryo and SMTs and the biomass parameters of the growing grain. Surprisingly, only a weak correlation appeared between an increasing endoreduplication in endosperm and seed biomass parameters. This is in agreement with the observations in maize, where modification of the cell cycle did not affect the final seed size (Sabelli *et al.*, 2013). The pericarp is the major site of starch deposition during the first days after pollination in barley (Radchuk and Borisjuk, 2014). At later time points, the main tissues accumulating starch, proteins, and lipids are the CSE and AL (Radchuk *et al.*, 2009; Sreenivasulu *et al.*, 2010; Peukert *et al.*, 2014). By combining these metabolomics studies with our morphometric analysis and SCV, we conclude that endopolyploidization in the pericarp and endosperm will be associated with the accumulation of storage components. This is further supported by the observation of a correlation between

high transcriptional activity of storage protein and starch genes and increasing amounts of nuclear DNA during barley endosperm development (Giese, 1992; Sreenivasulu *et al.*, 2004). In turn, endopolyploidization in barley embryo coincides in time with the differentiation of embryo tissues, which is evidence for its a role in sustaining cell fate (Bramsiepe *et al.*, 2010).

We found that many SMTs and endosperm nuclei were lost during the later stages of seed development. The results of TUNEL assays suggested that the degeneration initiated with highly endopolyploid nuclei. However, nuclear DNA content analysis of dry seeds revealed the presence of 3C and 6C (major ploidy level) nuclei and only ~6% of endoreduplicated 12C nuclei. Microscopic observations confirmed that these populations of nuclei originated from the AL (Keown *et al.*, 1977; Sreenivasulu *et al.*, 2010). Before degeneration, we observed accumulation of DNA damage in endoreduplicated nuclei. Data from Arabidopsis showed that endopolyploid nuclei in somatic cells and endosperm possess less condensed heterochromatin, which could be (i) a consequence of or (ii) the reason for lower genome stability (Schubert *et al.*, 2006; Baroux *et al.*, 2007). Hence, another potential function or a consequence of endoreduplication could be disabling specific nuclear functions and finally marking the cells for cell death (Sabelli, 2012). This is supported by the overlap of the spatial patterns of endoreduplication and cell death in the CSE in maize and rice (Young and Gallie, 2000; Kobayashi *et al.*, 2013; Kobayashi, 2019).

The experiments performed in this study, including the analysis of cell viability and microscopic analysis of seed tissues, showed that both barley SMTs and endosperm underwent cell death. Although the pericarp tissues were highly metabolically active during the first 8 DAP, Evans blue staining revealed a loss of viability already at 2 DAP in some regions. The previous study focusing on cell death in barley SMTs showed that cell divisions already decreased at 2 DAP in the pericarp, and the expansion of the tissue occurred by cell elongation in longitudinal directions (Radchuk *et al.*, 2011). We observed that cell death first appeared in the top and the bottom parts of the pericarp, and later (as the pericarp became very thin) expanded to the elongation zone. This pattern is consistent with the results described in bread wheat (Young and Gallie, 1999) and triticale (Li *et al.*, 2010). To summarize, SMTs first synthesize, temporarily store, and finally transport the nutrients to the developing embryo and endosperm (Li *et al.*, 2010; Radchuk *et al.*, 2011), and this is accompanied by the elimination of the tissues by cell death.

In barley, wheat, and triticale, cell death proceeded stochastically throughout the CSE (Young and Gallie, 1999; Li *et al.*, 2010). Conversely, in maize and rice, dying cells first appear in the upper CSE and expand outwards (Young and Gallie, 2000). These results suggest that the cell death patterns vary greatly depending on the species. However, there seems to be a link between endoreduplication, cell death, and storage compound deposition (Fig. 6). At 24 DAP, when the whole CSE

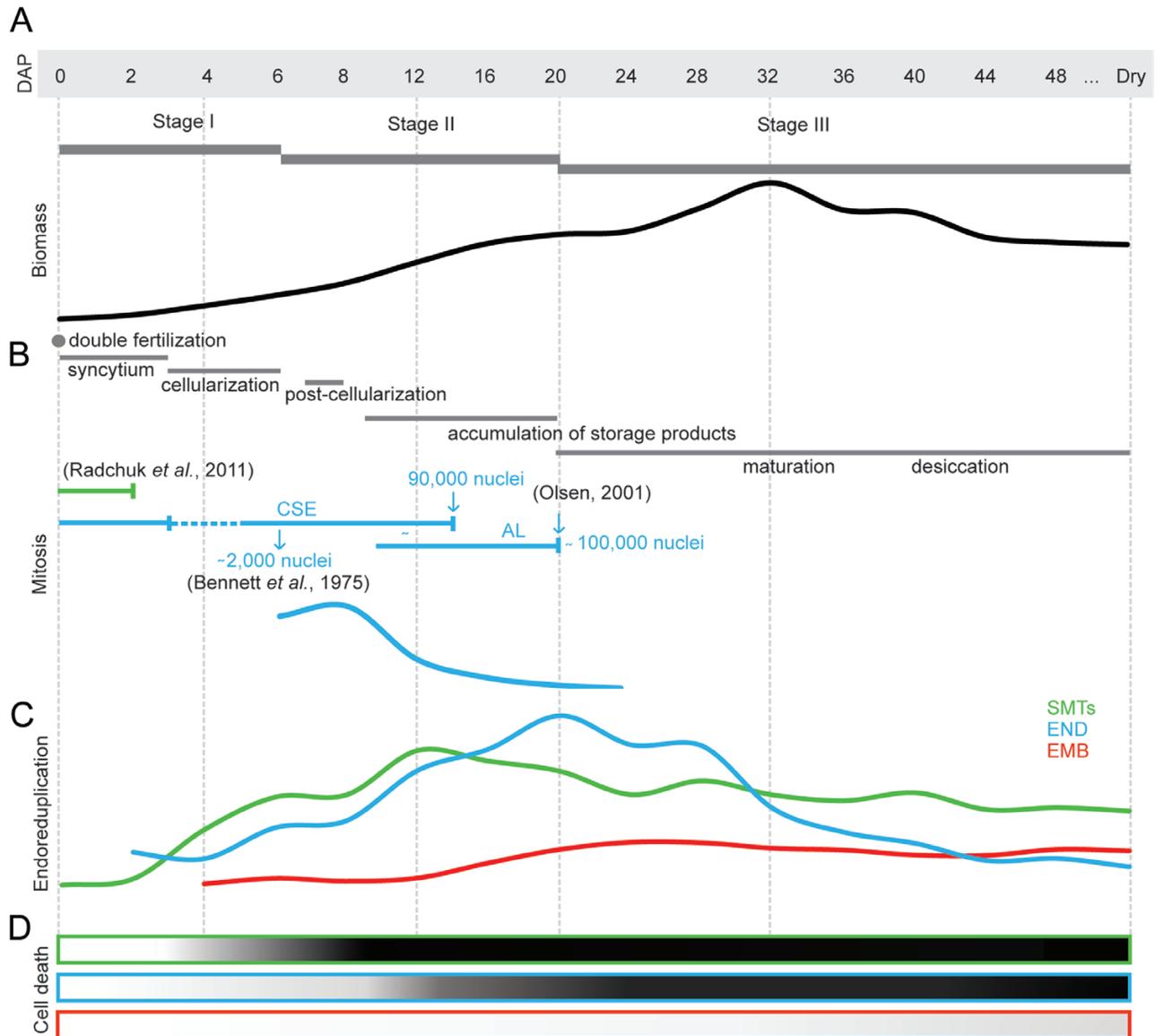


Fig. 6. The model of endoreduplication and cell death dynamics during barley seed development. (A) Phases of barley seed development. Thick gray horizontal lines mark three previously described general cereal developmental stages (Sabelli and Larkins, 2009; Sreenivasulu *et al.*, 2010; Dante *et al.*, 2014). The black curved line shows the dynamics of fresh weight changes. (B) Overview of developmental events in barley seed development. Data regarding nuclei numbers have been published (Bennett *et al.*, 1975; Olsen, 2001, and references therein; Radchuk *et al.*, 2011). (C) Schematic dynamics of endoreduplication in maternal tissues (SMTs; green), endosperm (END; blue), and embryo (EMB; red). (D) The dynamics of programmed cell death in barley seed tissues.

has undergone cell death, there was still detectable mRNA and the starch accumulation continued (Radchuk *et al.*, 2009). This indicates that specific precursors, regulatory molecules, and/or enzymes are synthesized in excess before degradation of the nuclei.

By analyzing six elite barley cultivars, we found variation in the kinetics of endosperm endoploidization. On the one hand, some varieties reached the peak earlier (the variation was between 16 and 20 DAP) and, on the other hand, there were varieties with a relatively low SCV throughout the entire endosperm development. Similar variability was observed in

rice (Kobayashi, 2019). This indicates the presence of genetic variation in the endopolyploidy dynamics during barley endosperm development. In our data set, we did not detect a link between the level of endoreduplication and the geographical origin of the cultivars. However, it has to be noted that the cultivars used in our study represent only a limited genetic diversity and that the variation is probably much greater in genetically distant landraces (Milner *et al.*, 2019). Furthermore, our study suggests that cell cycle dynamics during barley development are a genetically strictly controlled process. Although endoreduplication plays an important role in determining

cell size in cereal endosperm (Kladnik *et al.*, 2006; Kobayashi, 2019), there was no significant correlation between the SCV in developing endosperm and weight of the mature caryopsis. Because organ size is determined not only by cell size, but also by cell number (Robinson *et al.*, 2018), we speculate that the variation in cell size associated with endoreduplication is probably cancelled out by the diversity in cell number in the examined barley cultivars.

Supplementary data

The following supplementary data are available at *JXB* online.

Fig. S1. Morphology of barley seeds 16 days after pollination (DAP).

Fig. S2. Growth ratio of developing barley seed.

Fig. S3. Flow cytometric estimation of nuclear C-values in barley somatic tissues.

Fig. S4. An example showing the interpretation of histograms of nuclear DNA content.

Fig. S5. Representative gated flow cytometric histogram of barley seed tissues of cv. Compana.

Fig. S6. Estimation of nuclear C-values in barley hulls.

Fig. S7. Analysis of correlation between seed endopolyploidy and biomass parameters.

Fig. S8. Negative control of fluorescein diacetate assay.

Fig. S9. Analysis the aleurone layer viability.

Fig. S10. DAPI staining of nuclei and cell membranes in barley endosperm tissues.

Fig. S11. Phenotypic analysis of dry seeds from six two-row barley cultivars.

Fig. S12. Comparison of nuclear C-values in diploid seed tissues from two-row barley cultivars.

Fig. S13. Comparison of endosperm individual C-values from two-row barley cultivars.

Fig. S14. Super cell cycle value analysis of embryo and seed maternal tissues versus endosperm.

Acknowledgments

We thank Eva Jahnová and for technical assistance and Zdenka Bursová for plant care. AN, MK, and AP were supported primarily by The Czech Science Foundation grant 18-12197S. AP was further supported by a Purkyně Fellowship from the Czech Academy of Sciences. Multiple co-authors were funded from the European Regional Development Fund project 'Plants as a tool for sustainable global development'. (no. CZ.02.1.01/0.0/0.0/16_019/0000827).

Author contributions

AP and AN designed the research. AN, MK, BT, YZ, and DW performed experiments. JV flow-sorted the nuclei. AN analyzed the data, and designed and prepared the figures. JD discussed the data. AN and AP wrote the manuscript with contributions from all authors. All authors approved the final version of this article.

Data availability

All data supporting the findings of this study are available within the paper and within its supplementary data published online.

References

- Adachi S, Minamisawa K, Okushima Y, *et al.* 2011. Programmed induction of endoreduplication by DNA double-strand breaks in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **108**, 10004–10009.
- An LH, You RL. 2004. Studies on nuclear degeneration during programmed cell death of synergid and antipodal cells in *Triticum aestivum*. *Sexual Plant Reproduction* **17**, 195–201.
- Baroux C, Pecinka A, Fuchs J, Schubert I, Grossniklaus U. 2007. The triploid endosperm genome of Arabidopsis adopts a peculiar, parental-dosage-dependent chromatin organization. *The Plant Cell* **19**, 1782–1794.
- Barow M, Meister A. 2003. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. *Plant, Cell & Environment* **26**, 571–584.
- Bennett M, Smith J, Barclay I. 1975. Early seed development in the Triticeae. *Philosophical Transactions of the Royal Society B: Biological Sciences* **272**, 199–227.
- Bewley JD, Black M, Halmer P. 2006. *The encyclopaedia of seeds: science, technology and uses*. Wallingford, UK: CABI Publishing.
- Bhosale R, Boudolf V, Cuevas F, *et al.* 2018. A spatiotemporal DNA endoploidy map of the Arabidopsis root reveals roles for the endocycle in root development and stress adaptation. *The Plant Cell* **30**, 2330–2351.
- Bramsiepe J, Wester K, Weini C, Roodbarkelari F, Kasili R, Larkin JC, Hülskamp M, Schnittger A. 2010. Endoreplication controls cell fate maintenance. *PLoS Genetics* **6**, e1000996.
- Chevalier C, Bourdon M, Pirrello J, Cheniclet C, Gévaudant F, Frangne N. 2014. Endoreduplication and fruit growth in tomato: evidence in favour of the karyoplasmic ratio theory. *Journal of Experimental Botany* **65**, 2731–2746.
- Chevalier C, Nafati M, Mathieu-Rivet E, Bourdon M, Frangne N, Cheniclet C, Renaudin JP, Gévaudant F, Hernould M. 2011. Elucidating the functional role of endoreduplication in tomato fruit development. *Annals of Botany* **107**, 1159–1169.
- Chojecki A, Gale M, Bayliss M. 1986. The number and sizes of starch granules in the wheat endosperm, and their association with grain weight. *Annals of Botany* **58**, 819–831.
- D'amato F. 1964. Endopolyploidy as a factor in plant tissue development. *Caryologia* **17**, 41–52.
- Dante RA, Larkins BA, Sabelli PA. 2014. Cell cycle control and seed development. *Frontiers in Plant Science* **5**, 493.
- Doležel J, Binarová P, Lcretti S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* **31**, 113–120.
- Dominguez F, Cejudo FJ. 2014. Programmed cell death (PCD): an essential process of cereal seed development and germination. *Frontiers in Plant Science* **5**, 366.
- Engelen-Eigles G, Jones RJ, Phillips RL. 2000. DNA endoreduplication in maize endosperm cells: the effect of exposure to short-term high temperature. *Plant, Cell & Environment* **23**, 657–663.
- Ghasemzadeh J, Talebi AR, Khalili MA, Fesahat F, Halvaei I, Nabi A, Ashourzadeh S. 2015. Sperm parameters, protamine deficiency, and apoptosis in total globozoospermia. *Iranian Journal of Reproductive Medicine* **13**, 495–502.
- Giese H. 1992. Replication of DNA during barley endosperm development. *Canadian Journal of Botany* **70**, 313–318.
- Gonzalez N, Gévaudant F, Hernould M, Chevalier C, Mouras A. 2007. The cell cycle-associated protein kinase WEE1 regulates cell size in relation to endoreduplication in developing tomato fruit. *The Plant Journal* **51**, 642–655.

- Hicks KB, Montanti J, Nghiem NP.** 2014. Use of barley grain and straw for biofuels and other industrial uses. Elsevier Ltd.
- Holásková E, Galuszka P, Mičúchová A, Šebela M, Öz MT, Frébort I.** 2018. Molecular farming in barley: development of a novel production platform to produce human antimicrobial peptide LL-37. *Biotechnology Journal* **13**, e1700628.
- Keown AC, Taiz L, Jones RL.** 1977. The nuclear DNA content of developing barley aleurone cells. *American Journal of Botany* **64**, 1248–1253.
- Kladnik A, Chourey PS, Pring DR, Dermastia M.** 2006. Development of the endosperm of *Sorghum bicolor* during the endoreduplication-associated growth phase. *Journal of Cereal Science* **43**, 209–215.
- Kobayashi H.** 2019. Variations of endoreduplication and its potential contribution to endosperm development in rice (*Oryza sativa* L.). *Plant Production Science* **22**, 227–241.
- Kobayashi H, Ikeda TM, Nagata K.** 2013. Spatial and temporal progress of programmed cell death in the developing starchy endosperm of rice. *Planta* **237**, 1393–1400.
- Kowles RV, Srienc F, Phillips RL.** 1990. Endoreduplication of nuclear DNA in the developing maize endosperm. *Developmental Genetics* **11**, 125–132.
- Langridge P.** 2018. Economic and academic importance of barley. In: Stein N, Muehlbauer GJ, eds. *The barley genome*. Cham: Springer, 1–10.
- Leiva-Neto JT, Grafi G, Sabelli PA, Dante RA, Woo YM, Maddock S, Gordon-Kamm WJ, Larkins BA.** 2004. A dominant negative mutant of cyclin-dependent kinase A reduces endoreduplication but not cell size or gene expression in maize endosperm. *The Plant Cell* **16**, 1854–1869.
- Li CY, Li WH, Li C, Gaudet DA, Laroche A, Cao LP, Lu ZX.** 2010. Starch synthesis and programmed cell death during endosperm development in triticale (\times *Triticosecale* Wittmack). *Journal of Integrative Plant Biology* **52**, 602–615.
- Liu CH, Finke A, Díaz M, Rozhon W, Poppenger B, Baubec T, Pecinka A.** 2015. Repair of DNA damage induced by the cytidine analog zebularine requires ATR and ATM in Arabidopsis. *The Plant Cell* **27**, 1788–1800.
- Mascher M, Gundlach H, Himmelbach A, et al.** 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433.
- Mathieu-Rivet E, Gévaudant F, Sicard A, et al.** 2010. Functional analysis of the anaphase promoting complex activator CCS52A highlights the crucial role of endo-reduplication for fruit growth in tomato. *The Plant Journal* **62**, 727–741.
- Milner SG, Jost M, Taketa S, et al.** 2019. Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics* **51**, 319–326.
- Nagl W.** 1976. DNA endoreduplication and polyteny understood as evolutionary strategies. *Nature* **261**, 614–615.
- Olsen OA.** 2001. Endosperm development: cellularization and cell fate specification. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**, 233–267.
- Palermo GD, O'Neill CL, Chow S, Cheung S, Parrella A, Pereira N, Rosenwaks Z.** 2017. Intracytoplasmic sperm injection: state of the art in humans. *Reproduction* **154**, F93–F110.
- Parker F, Laimbeer E, Markis M, Veilleux RE.** 2018. Measuring endoreduplication by flow cytometry of isolated tuber protoplasts. *Journal of Visualized Experiments* **133**, e57134.
- Peukert M, Thiel J, Peshev D, Weschke W, Van den Ende W, Mock HP, Matros A.** 2014. Spatio-temporal dynamics of fructan metabolism in developing barley grains. *The Plant Cell* **26**, 3728–3744.
- Radchuk V, Borisjuk L.** 2014. Physical, metabolic and developmental functions of the seed coat. *Frontiers in Plant Science* **5**, 510.
- Radchuk V, Borisjuk L, Sreenivasulu N, Merx K, Mock HP, Rolletschek H, Wobus U, Weschke W.** 2009. Spatiotemporal profiling of starch biosynthesis and degradation in the developing barley grain. *Plant Physiology* **150**, 190–204.
- Radchuk V, Tran V, Radchuk R, et al.** 2018. Vacuolar processing enzyme 4 contributes to maternal control of grain size in barley by executing programmed cell death in the pericarp. *New Phytologist* **218**, 1127–1142.
- Radchuk V, Weier D, Radchuk R, Weschke W, Weber H.** 2011. Development of maternal seed tissue in barley is mediated by regulated cell expansion and cell disintegration and coordinated with endosperm growth. *Journal of Experimental Botany* **62**, 1217–1227.
- Rewers M, Sadowski J, Sliwinska E.** 2009. Endoreduplication in cucumber (*Cucumis sativus*) seeds during development, after processing and storage, and during germination. *Annals of Applied Biology* **155**, 431–438.
- Robinson DO, Coate JE, Singh A, Hong L, Bush M, Doyle JJ, Roeder AHK.** 2018. Ploidy and size at multiple scales in the Arabidopsis sepal. *The Plant Cell* **30**, 2308–2329.
- Rodríguez MV, Barrero JM, Corbineau F, Gubler F, Benech-Arnold RL.** 2015. Dormancy in cereals (not too much, not so little): about the mechanisms behind this trait. *Seed Science Research* **25**, 99–119.
- Sabelli PA.** 2012. Replicate and die for your own good: endoreduplication and cell death in the cereal endosperm. *Journal of Cereal Science* **56**, 9–20.
- Sabelli PA, Larkins BA.** 2009. The development of endosperm in grasses. *Plant Physiology* **149**, 14–26.
- Sabelli PA, Liu Y, Dante RA, et al.** 2013. Control of cell proliferation, endoreduplication, cell size, and cell death by the retinoblastoma-related pathway in maize endosperm. *Proceedings of the National Academy of Sciences, USA* **110**, E1827–E1836.
- Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675.
- Schnürer J, Rosswall T.** 1982. Fluorescein diacetate hydrolysis as a measure of total microbial activity in soil and litter. *Applied and Environmental Microbiology* **43**, 1256–1261.
- Schubert V, Klatt M, Pecinka A, Meister A, Jasencakova Z, Schubert I.** 2006. Sister chromatids are often incompletely aligned in meristematic and endopolyploid interphase nuclei of *Arabidopsis thaliana*. *Genetics* **172**, 467–475.
- Sreenivasulu N, Altschmied L, Radchuk V, Gubatz S, Wobus U, Weschke W.** 2004. Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *The Plant Journal* **37**, 539–553.
- Sreenivasulu N, Borisjuk L, Junker BH, Mock H-P, Rolletschek H, Seiffert U, Weschke W, Wobus U.** 2010. Barley grain development: toward an integrative view. In: Jeon K, ed. *International review of cell and molecular biology*. Elsevier Inc., 49–89.
- Tottman DR, Makepeace RJ, Broad H.** 1979. An explanation of the decimal code for the growth stages of cereals, with illustrations. *Annals of Applied Biology* **93**, 221–234.
- Trafford K, Haleux P, Henderson M, Parker M, Shirley NJ, Tucker MR, Fincher GB, Burton RA.** 2013. Grain development in *Brachypodium* and other grasses: possible interactions between cell expansion, starch deposition, and cell-wall synthesis. *Journal of Experimental Botany* **64**, 5033–5047.
- Tran V, Weier D, Radchuk R, Thiel J, Radchuk V.** 2014. Caspase-like activities accompany programmed cell death events in developing barley grains. *PLoS One* **9**, e109426.
- Tricase C, Amicarelli V, Lamonaca E, Leonardo Rana R.** 2018. Economic analysis of the barley market and related uses. In: Tadele Z, ed. *Grasses as food and feed*. London: IntechOpen, 26–46.
- Waddington SR, Cartwright PM, Wall PC.** 1983. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany* **51**, 119–130.
- Weschke W, Panitz R, Sauer N, Wang Q, Neubohn B, Weber H, Wobus U.** 2000. Sucrose transport into barley seeds: molecular characterization of two transporters and implications for seed development and starch accumulation. *The Plant Journal* **21**, 455–467.
- Wu X, Liu J, Li D, Liu CM.** 2016. Rice caryopsis development II: dynamic changes in the endosperm. *Journal of Integrative Plant Biology* **58**, 786–798.
- Yifang C, Jun Z, Peisong X, Weidong Z, Jianmin C, Cunxu W.** 2012. Programmed cell death in wheat starchy endosperm during kernel development. *African Journal of Agricultural Research* **7**, 6533–6540.
- Young TE, Gallie DR.** 1999. Analysis of programmed cell death in wheat endosperm reveals differences in endosperm development between cereals. *Plant Molecular Biology* **39**, 915–926.
- Young TE, Gallie DR.** 2000. Programmed cell death during endosperm development. *Plant Molecular Biology* **44**, 283–301.L

Appendix IV

Endopolyploidy variation in wild barley seeds across environmental gradients in Israel

Nowicka, A., Sahu, P., P., **Kovacik, M.**, Weigt, D., Tokarz, B., Krugman, T.,
and Pecinka, A.

Genes (Basel)

DOI: 10.3390/genes12050711

IF (2021): 4.141

Article

Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel

Anna Nowicka ^{1,2,†}, Pranav Pankaj Sahu ^{1,3,†}, Martin Kovacik ¹, Dorota Weigt ⁴, Barbara Tokarz ⁵, Tamar Krugman ^{6,*} and Ales Pecinka ^{1,*}

- ¹ Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Czech Academy of Sciences, Šlechtitelů 31, 779 00 Olomouc, Czech Republic; nowicka@ueb.cas.cz (A.N.); sahu@ueb.cas.cz (P.P.S.); kovacik@ueb.cas.cz (M.K.)
- ² The Franciszek Górski Institute of Plant Physiology, The Polish Academy of Sciences, Niezapominajek 21, 30-239 Krakow, Poland
- ³ Global Change Research Institute of the Czech Academy of Sciences, Bělidla 986/4a, 603 00 Brno, Czech Republic
- ⁴ Department of Genetics and Plant Breeding, Poznan University of Life Sciences, 11 Dojazd St., 60 632 Poznan, Poland; dorota.weigt@up.poznan.pl
- ⁵ Department of Botany, Physiology and Plant Protection, Faculty of Biotechnology and Horticulture, University of Agriculture in Krakow, Al. 29 Listopada 54, 31-425 Krakow, Poland; barbara.tokarz@urk.edu.pl
- ⁶ Institute of Evolution, University of Haifa, Abba Khoushy Ave. 199, Haifa 3498838, Israel
- * Correspondence: tkrugman@evo.haifa.ac.il (T.K.); pecinka@ueb.cas.cz (A.P.); Tel.: +972-4-824-0783 (T.K.); +420-585-238-709 (A.P.)
- † These authors contributed equally to this work.

Citation: Nowicka, A.; Sahu, P.P.; Kovacik, M.; Weigt, D.; Tokarz, B.; Krugman, T.; Pecinka, A. Endopolyploidy Variation in Wild Barley Seeds across Environmental Gradients in Israel. *Genes* **2021**, *12*, 711. <https://doi.org/10.3390/genes12050711>

Academic Editors: Elwira Sliwinska, Monika Rewers and Iwona Jedrzejczyk

Received: 16 March 2021
Accepted: 7 May 2021
Published: 10 May 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 (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: Wild barley is abundant, occupying large diversity of sites, ranging from the northern mesic Mediterranean meadows to the southern xeric deserts in Israel. This is also reflected in its wide phenotypic heterogeneity. We investigated the dynamics of DNA content changes in seed tissues in ten wild barley accessions that originated from an environmental gradient in Israel. The flow cytometric measurements were done from the time shortly after pollination up to the dry seeds. We show variation in mitotic cell cycle and endoreduplication dynamics in both diploid seed tissues (represented by seed maternal tissues and embryo) and in the triploid endosperm. We found that wild barley accessions collected at harsher xeric environmental conditions produce higher proportion of endoreduplicated nuclei in endosperm tissues. Also, a comparison of wild and cultivated barley strains revealed a higher endopolyploidy level in the endosperm of wild barley, that is accompanied by temporal changes in the timing of the major developmental phases. In summary, we present a new direction of research focusing on connecting spatiotemporal patterns of endoreduplication in barley seeds and possibly buffering for stress conditions.

Keywords: endoreduplication; endosperm; *Hordeum vulgare* subsp. *spontaneum*; seed development; super cycle value

1. Introduction

Cultivated barley (*Hordeum vulgare* subsp. *vulgare*, $2n = 2x = 14$) was domesticated about 10,000 years ago from its progenitor wild barley (*H. vulgare* subsp. *spontaneum*, $2n = 2x = 14$). Barley belongs to the group of “Neolithic founder crops”, and was one of the first cereals that became a pillar of food and feed for ancient societies [1]. The Fertile Crescent is the center of barley domestication, distribution, and diversity [1]. A recent arche-genomic study performed on ancient DNA of 6000 years-old barley grains excavated at a cave in the Judean Desert in Israel, narrowed its domestication region to the Upper Jordan Valley [2]. Subspecies *spontaneum* is distributed from eastern North Africa, through the

Middle East to India and west China [1]. It constitutes an important annual element of open herbaceous and park-like vegetation [3]. Wild barley natural habitats are characterized by wide ecogeographical diversity caused mostly by contrasting climatic and topographic conditions within the East Mediterranean region. This is reflected by its phenotypic and genetic heterogeneity [4]. During the last century, wild barley was collected all over its distribution area and seed samples are stored and maintained in ex situ genebanks [3,4]. While domestication and modern plant breeding have reduced the genetic diversity of cultivated barleys, the stocks of subsp. *spontaneum* form a major source for variability, novel genes, and alleles for barley breeding [5,6]. For example, wild barley was found to be an important source of resistance to biotic and abiotic stresses, including multiple diseases [7], tolerance to cold [8], drought [9], and salt [10].

Cereal grain development includes three major phases, characterized by different cellular and physiological events (Figure 6a–c quoted from [11]). Phase I starts with double fertilization and passes smoothly into the cells/nuclei proliferation; phase II comprises differentiation of embryo and endosperm tissues, and seed mass gain by the accumulation of storage compounds; phase III corresponds to seed maturation, weight reduction by desiccation, and an onset of dormancy. These phases partially overlap with three morphological caryopsis growth stages named water, milk, and dough, respectively [12].

Cereal grain consists of three major compartments: multilayered seed maternal tissues (SMTs; nucellar projection, pericarp plus seed coats), endosperm, and embryo. The pericarp (diploid, 2x) is derived from the ovary wall and adheres strongly to the seed coats of the ovule [13]. Within the first days after pollination (DAP), the pericarp serves to protect and support the growing endosperm and embryo by starch deposition and photosynthesis in cultivated barley [14,15]. During double fertilization, one sperm nucleus fuses with the egg cell nucleus and gives rise to the diploid embryo (2x), while the second sperm cell nucleus fuses with the diploid central cell to form a triploid endosperm (3x) with the peculiar genetic constitution of one paternal and two maternal genomes. Endosperm nuclei first form syncytium (a.k.a. coenocyte) and later endosperm cellularizes and differentiates into five specialized tissues: the central starchy endosperm (CSE), the sub-aleurone layer (SAL), the aleurone layer (AL), the basal endosperm transfer layer (BETL), and the embryo-surrounding region (ERS) [16]. Endosperm protects and nourishes the embryo. It is the main caryopsis part accumulating primarily sugars and proteins [13,16]. The cereal kernel is covered by hulls that consist of the lemma, palea, and glumes of maternal origin and which remain tightly attached to the grain even after ripening [17].

Both SMTs and endosperm tissues undergo genetically controlled endoreduplication during seed development in cultivated barley [11]. Endoreduplication (a.k.a. endopolyploidization) occurs via the endocycle and is a variant of the cell cycle, in which cell nuclei increase their ploidy through repeated rounds of replication without cell divisions (reviewed, e.g., in [18] and [19]). To unravel the mechanism involved in the switch from a mitotic cell cycle to an endocycle many cyclin-dependent kinases (CDKs), their cyclin partners, CDK inhibitors (e.g., WEE1), and retinoblastoma-related (RBR) proteins have been studied [20]. Despite many efforts, the knowledge about the molecular control of endoreduplication is fragmentary. In most Angiosperms, endoreduplication is common in specialized cells producing secondary metabolites and/or as a means to accelerate cell expansion of specific tissues [20]. Also, various abiotic and biotic factors affect the endopolyploidy level of cells and tissues [21,22]. For instance, salinity or the absence of light stimulates extra endocycles in different Arabidopsis organs [23,24]. Endoreduplication can also be triggered upon symbiotic [25] and also pathogenic [26] plant-microbe interactions. In contrast, endopolyploidization can be repressed by both very high and very low temperatures [22] or drought [27].

The development of cereal seeds would not be possible without programmed cell death (PCD). In phase I, maternal tissues, i.e., components of the embryo sac, nucellus, nucellar projection, seed coats, and pericarp undergo a progressive degeneration by PCD [28,29]. During phases II and III, mainly two endosperm parts: ESR and CSE undergo cell

death, but the cells remain intact in the mature grain and their contents will not be remobilized until germination. Finally, the mature grain contains mainly dead material, where only the embryo, BETL, and AL tissues remain alive [30–32].

Wild barley is a generalist abundant across diverse habitats ranging from the mesic Mediterranean meadows to the xeric southern habitats and even penetrating the central Negev desert in Israel. Such environmental heterogeneity can be a direct driving force for adaptation [4]. The main objective of this study was to investigate the dynamics of endoreduplication in seed tissues of wild barley originating from mesic, semi-mesic, semi-xeric, and xeric ecogeographic sites of Israel. For this purpose, we measured the DNA contents in diploid seed tissues (embryo and maternal tissues) and triploid endosperm using flow cytometry. We calculated the proportion of nuclei with different DNA contents and estimated the level of endoreduplication with a new formula called the super cycle value (SCV) [11]. For a better understanding of the dynamics of processes associated with wild barley grain development, we also monitored the morphology of developing seeds and performed Evans blue cell death assay. We found that wild barley accessions originating from the xeric environments have on average higher proportion of endoreduplicated nuclei in seed tissues, and tend to have a higher SCV index. This indicates the impact of harsh conditions on endopolyploidization. A comparison of wild and cultivated barleys reveals a higher endopolyploidy level in the endosperm of wild barley that is accompanied by temporal changes in the timing of the major developmental phases.

2. Materials and Methods

2.1. Plant Materials and Growth Conditions

Ten wild barley (*H. vulgare* subsp. *spontaneum*) accessions originating from Israel were used in this study. Seeds were obtained from the Institute of Evolution Wild Cereal Gene Bank (ICGB) at the University of Haifa, Israel, and Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany. The ICGB accessions were named based on the seed collection sites and the type of environment (Figure 1; Table 1). Three accessions originated from typical xeric (x) environments: Machtsh Gadol (MGx), Mehola (MHx), and Wadi Qilt (WQx); three from mesic (m) environments: Rosh Pinna (RPm), Tel Hai (THm), and Zefat (ZFm), and one accession from Bar Giyyora represented semi-mesic environment (BGsm). Two accessions originated from Nahal Oren (NO) Canyon, also named “Evolution Canyon” [32]. The first NO accession was collected from the North-facing slope (NFS) representing a mesic environment (NOm), and the second from the South-facing slope (SFS) belonging to the semi-xeric group (NOsx). In brief, the xeric environment is characterized by low annual rainfall and high temperatures, and mesic by high annual rainfall and lower temperatures. Differences between environments are mainly reflected at seed development time during March to April. The environmental conditions at BG are regarded as semi-mesic due to the high rainfall and dry environment in the Judean mountains. The SFS is regarded as semi-xeric due to higher solar radiation as compared with the NFS Nahal Oren. The IPK accession HS584 carries the gene bank name HOR 12560, and the exact site of the collection is unknown.

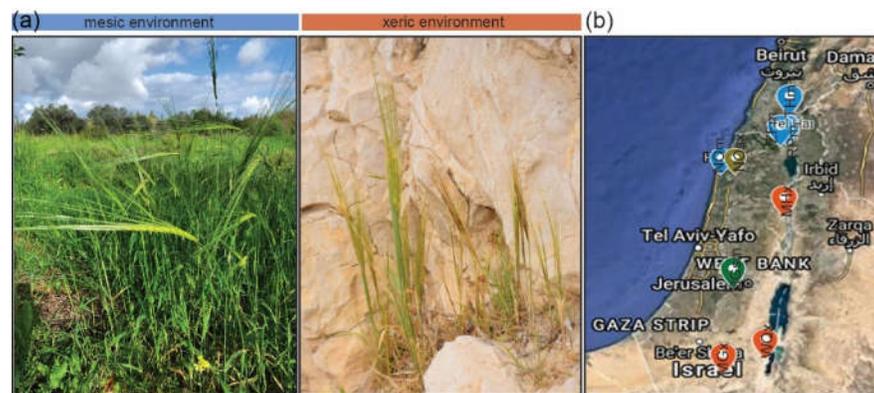


Figure 1. Geographic origin of wild barley (*H. vulgare* subsp. *spontaneum*) accessions. (a) Examples of wild barley from the Galilee (mesic) and Judean desert (xeric) in Israel. (b) Collection sites in Israel. Blue points = mesic sites (THm = Tel Hai, ZFm = Zefat, RPs = Rosh Pinna, NOm = Nahal Oren northern facing slope); green point = semi-mesic site (BGsm = Bar Giyyora); green-brown point = semi-xeric site (NOsx = Nahal Oren southern facing slope); orange points = xeric sites (MHx = Mehola, MGx = Machtsh Gadol, WQx = Wadi Qilt). The map was generated using Google Maps. Detailed ecogeographical data are presented in Table 1.

Also, published data [11] from six cultivars (cv.) of two-rowed spring barley (*H. vulgare* subsp. *vulgare*): Betzes (PI 129430), Compana (PI 539111), Golden Promise (GP; PI 343079), Ingrid (PI 263574), Klages (CIho 15478) and Mars (PI 599629) and three additional cv. of six-rowed spring barley: Glacier (CIho 6976), Mars (CIho 7015) and Morex (BCC 906) were used for comparison.

Grains were stratified in the dark at 4 °C for 48 h, evenly spread on wet filter paper in a Petri dish, covered with a lid, and germinated at 25 °C for 3 days in the dark. Germinating kernels were planted into 5 cm × 5 cm peat pots with a mixture of soil and sand (2:1, v/v) and grown in an air-conditioned phytochamber with a long day regime (16 h day with 20 °C and 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity; 8 h night with 16 °C; 60% humidity). After 10 days, wild barley plants were placed into the vernalization chamber (short-day regime; 8 h day with 4 °C, light intensity 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$; 16 h night with 4 °C; humidity 85%) for three weeks. Ten-day-old cultivated barley plants and 31-day-old wild barley plants were transferred into the 12 cm × 12 cm pots filled with the above-described soil mixture and grown under long-day conditions. For each accession five plants were grown. Day of pollination (DOP) was monitored using the morphology of stigma and anthers according to the Waddington scale (W10) [33] as we described previously [11,34]. In brief, the spikelets at DOP were characterized by extended hulls, widely branched stigma, and the presence of pollen grains on stigmatic hairs. Seeds were collected from the center of the spikelet at two- and four-day intervals, starting from 4 until 24 days after pollination (DAP). For this experimental setup, in total seven-time points were examined (i.e., 4, 6, 8, 12, 16, 20, 24 DAP). For three accessions, HS584, RPs, and BGsm collecting the seeds were extended up to 48 DAP (additional six collection points: 28, 32, 36, 40, 44, 48 DAP). Mature dry seeds (called ‘dry seeds’ latter in the text) were harvested around 60–65 DAP from fully dried mother plants, cleaned, and stored first ~30 days at 20 °C, then ~60 days at 4 °C, both in darkness. The analysis was performed after 90 ± 5 days after harvesting the seeds. During collecting the seed from mother plants, kernels were at the hard-dough phase of barley grain development (87–89 stages according to [12]). It means that grains were dry and cannot be squeezed out. The maximum dry seed section area was reduced by approximately 30–40% as compared to 20–28 DAP seeds (Supplementary Figure S5). Hulls had yellow color.

Table 1. Sampling sites and ecogeographical data of the analyzed wild barley accessions.

Locality/Name (All Within Israel)	Gene- bank	Acronym	Type of Envi- ronment	Longitude (N)	Latitude (E)	Altitude (a.s.l.)	Maximum Temperature in April (°C)	Rainfall in April (mm)	Annual Rainfall (mm)	Average Annual Humidity at 14:00 Mean ± SD	Annual Evaporation (cm)	Soil Type
Bar Giyyora	ICGB	BCsm	Semi-mesic	35.083333	31.716667	760	22	18	535	47.1 ± 10.8	215	T
HS584	IPK	HS584	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Machtesh Gadol	ICGB	MGx	Xeric	35.000000	30.950000	n.a.	25	3	70	n.a.	n.a.	n.a.
Mehola	ICGB	MHx	Xeric	35.533333	32.350000	-150	30	6	<200	37.5	240	A
Nahal Oren	ICGB	NOM	Mesic	34.966667	32.716667	n. a.	24	13	584	n.a.	n.a.	n.a.
Nahal Oren	ICGB	NOsx	Semi- xeric	34.966667	32.716667	n. a.	24	13	584	n.a.	n.a.	n.a.
Rosh Pinna	ICGB	RPm	Mesic	35.550000	32.983333	700	25	20	535	43.6 ± 10.5	220	T
Tel Hai	ICGB	THm	Mesic	35.573979	33.234719	400	26	23	768	46.9 ± 7.6	220	T
Wadi Qilt	ICGB	WQx	Xeric	35.44565	31.859	50	30	6	<200	34.7 ± 9.3	330	A
Zefat	ICGB	ZFm	Mesic	35.496001	32.969206	800	20	27	670	50.4 ± 13.1	220	R

This table was partially prepared based on [35] and <https://ims.gov.il/he/ClimateAtlas>, accessed on 15 March 2021. ICGB = Institute of Evolution Wild Cereal Gene Bank at the University of Haifa, Israel; IPK = Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. Annual rainfall = average for the period 1981–2010. Average maximal temperature and rainfall in April were recorded in 1995–2009. a.s.l. — above sea level; n.a. — not available; SD — standard deviation over mean monthly data; Soil type: A = alluvium, R = rendzina, T = terra rossa.

2.2. Analysis of Nuclear DNA Content and Calculation of the Super Cycle Value (SCV)

Nuclear DNA contents were estimated using flow cytometry (FCM). For each time point, five to six individual seeds, freshly collected from one spike were analyzed. The measurements were repeated three times on different days using seeds harvested from different mother plants and keeping the same time of day for analysis. The isolation of nuclei and estimation of nuclear DNA content was performed as previously described [11]. Briefly, seeds directly after harvesting were cleaned by removing hulls using tweezers. Then, single seeds were immediately homogenized with a razor blade in a Petri dish containing 500 μL of Otto I solution (0.1 M citric acid, 0.5% Tween 20). The crude suspension was filtered through 50 μm nylon mesh (Sysmex-Partec) and stained around 15 min with 1 mL of Otto II solution (0.4 M $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$) supplemented with 2 $\mu\text{g mL}^{-1}$ DAPI (4',6-diamidino-2-phenylindole). Nuclei samples were analyzed using either a CyFlow Space or a Partec PAS I flow cytometers (Sysmex-Partec, Muenster, Germany), both equipped with UV-led diode lamps. For calibration of the cytometers, the optics were adjusted using calibration beads (A7304, Invitrogen, Carlsbad, CA, USA) until the coefficient of variation (CV) reached <2%. At least 5000 particles were acquired per sample, using a log3 scale. Histograms were evaluated by the FloMax software (Sysmex-Partec, Muenster, Germany).

To estimate amount of endoreduplication, we used super cycle value (SCV) [11]. In SCV, 8C in the diploid and 12C in the triploid tissues were considered as the first levels of endopolyploid nuclei. Our rationale is, that it is not possible to unambiguously distinguish by FCM whether a given 4C (or 6C nucleus in endosperm) nucleus just entered endoreduplication or will mitotically divide [36]. For diploid tissues $\text{SCV} = ((n \ 2\text{C} \times 0) + (n \ 4\text{C} \times 0) + (n \ 8\text{C} \times 1) + (n \ 16\text{C} \times 2)) / (n \ 2\text{C} + n \ 4\text{C} + n \ 8\text{C} + n \ 16\text{C})$, and for triploid endosperm $\text{SCV} = ((n \ 3\text{C} \times 0) + (n \ 6\text{C} \times 0) + (n \ 12\text{C} \times 1) + (n \ 24\text{C} \times 2)) / (n \ 3\text{C} + n \ 6\text{C} + n \ 12\text{C} + n \ 24\text{C})$, n = number of counts per given C-value content.

2.3. Determination of Seed Morphology Parameters

Analysis of dry seed morphology parameters was performed in three biological replicates, each with at least 20 seeds collected from four to five spikes of different plants. Dry kernels were peeled off, weighed with an analytical scale (Sartorius, Göttingen, Germany), and photographed using a SZX16 binocular microscope (Olympus, Tokyo, Japan) bonded with a Regita 1300 QImaging camera and QCapture $\times 64$ software (Olympus). Seed length and width were measured using ImageJ calibrated with internal size control. Seeds from 4 to 48 DAP and dry seeds were peeled off and cut with a razor blade along the longitudinal and transverse axis. At least 20 individual seeds were photographed as described above using a binocular microscope. Hulled seeds that possessed awns were photographed with a D5600 (Nikon, Tokyo, Japan) digital camera equipped with an 80 mm Nikkor objective. All photo-matrix were composed of separately taken photos of individual seeds and merged in Adobe Photoshop CS5 (Adobe Inc., San Jose, CA, USA).

2.4. Cell Death Assay by Evans Blue Staining

Seeds from 4 to 48 DAP and dry seeds were peeled off and cut with a razor blade along the longitudinal and transverse axis. At least 20 individual seeds bulked from four to five spikes of different plants were stained in 0.1% (w/v) Evans blue (314-13-6, Sigma-Aldrich, St. Louis, MO, USA) for 2 min. Stained sections were washed twice for 10 min with distilled water [31]. Transverse and sagittal sections of samples were analyzed with an SZX16 binocular microscope (Olympus). Images were captured with a Regita 1300 QImaging camera and QCapture $\times 64$ software (Olympus) using the same settings and proceeded in Adobe Photoshop CS5 (Adobe Inc.).

2.5. Statistical Analysis

All data after testing for normal distribution were examined by one- or two-way analysis of variance (ANOVA), after which post hoc comparison was performed using Duncan's multiple ranges ($p \leq 0.05$) test. Data expressed as percentages were first transformed using arcsine transformation. Principal component (PC) analysis was used to analyze relations between variables. Statistical analyses were performed in Statistica v. 12 (Stat Soft Inc., Tulsa, OK, USA), Minitab v. 18 (Minitab, LLC, State College, PA, USA) or RStudio programs.

3. Results

3.1. Variation in Mature Dry Seed Morphology of Wild Barley

We used samples from nine geographically distant sites from North to South Israel along with the aridity gradient (Figure 1b; Table 1). Also, we included a commonly used gene-bank accession of wild barley named HS584 from an unknown origin in Israel. The wild barley accessions varied as to their mature dry seed weight, length, width, and awn length (Figure 2a–d; Supplementary Figure S1). For example, the TKW in wild barleys ranged from 14.2 g (MHx) to 40.5 g (NOM). We noted that seeds of wild barley were longer than those of cultivars (wild barley seed length ≥ 9 mm, cultivars seed length ~ 7 –8 mm) (Supplementary Figure S1a). However, only slight differences between wild barleys appeared for seed width (Supplementary Figure S1b). Seeds of wild barley accessions had longer awn. We also noted an intraspecific variation with xeric accessions having shorter awns than the mesic ones (Supplementary Figure S1c). ANOVA results showed that the values of observed variables between wild barley depended on the accession (genotype), and except for seed length, also from the type of environment (Figure 2d). Using these seed phenotypic data, we performed principal component (PC) analysis (Figure 2c; Supplementary Figure S2). However, this analysis did not reveal any specific group. We noted that one xeric accession MHx varied from the rest of the wild barleys. In addition, barley cultivars were separated from wild accessions.

Collectively, these data show a phenotypic variation of wild barley seeds. The shortened seed awn length is the most pronounced feature differentiating xeric barley accessions.

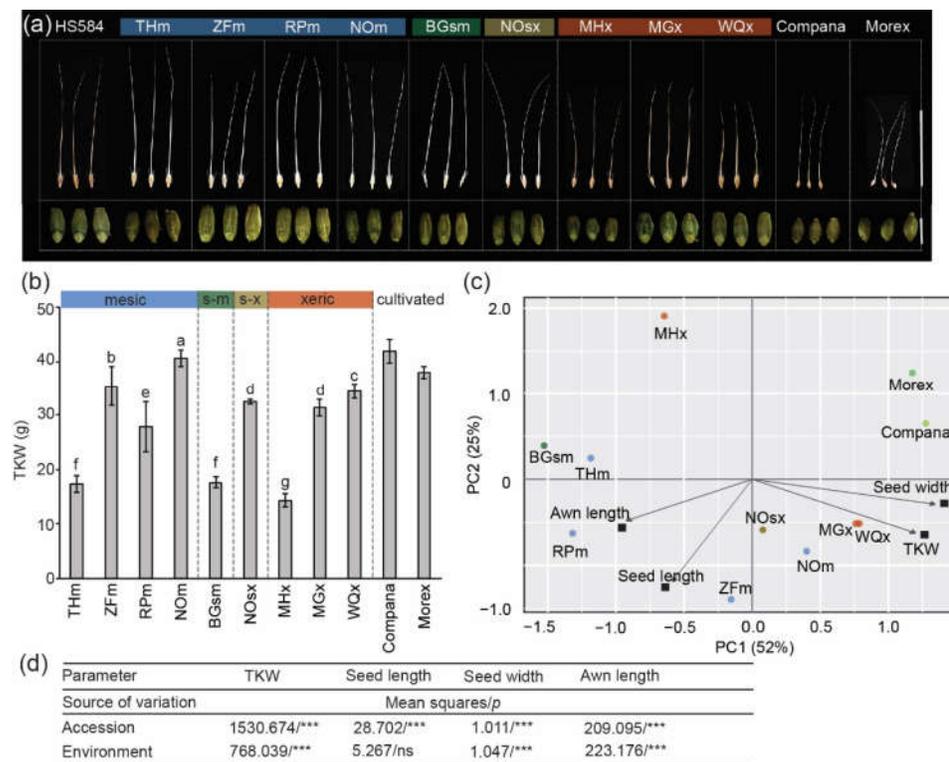


Figure 2. Dry seed phenotypes of wild barley accessions. (a) Dorsal views of hulled dry seeds with awns (upper panel; Scale bar = 10 cm) and peeled dry seeds (lower panel; scale bar = 10 mm). Mesic accessions: THm = Tel Hai, ZFm = Zefat, RPm = Rosh Pinna, NOm = Nahal Oren NSF; semi-mesic accession: BGsm = Bar Giyyora; semi-xeric accession: NOsx = Nahal Oren SFS; xeric accessions: MHx = Mehola, MGx = Machtsh Gadol, WQx = Wadi Qilt. All genotypes were grown in phytochamber under the same conditions. (b) Quantitative data for the thousand-kernel weight (TKW). Data are the means (\pm SD) from three biological replicates. Values marked with the same letter do not differ according to Duncan multiple range tests ($p \leq 0.05$). (c) Principal component (PC) analysis of TKW, seed length, and width for peeled seeds, and awn length. The positions represent contribution rates of the two PCs (Source data are shown in Supplementary Figure S1, other combinations of PCs are presented in Supplementary Figure S2). The ecological conditions at the sampling site of HS584 are unknown. Compana and Morex represent two- and six-rowed cultivated barley controls, respectively. (d) Summary of ANOVA performed for seed traits. The sources of variance were as follows: nine accessions and four environment types. *** Significant at $p \leq 0.001$; ns—not significant.

3.2. Variation in Endoreduplication Dynamics in Developing Wild Barley Seeds

3.2.1. Diploid Seed Tissues

We used whole peeled seeds (hulls were manually removed) to study the degree of endopolyploidy in the seeds of wild barley. We measured C-values of diploid nuclei from the embryo (EMB) and seed maternal tissues (SMTs, containing: nucellar projection, pericarp, and seed coats) and of triploid nuclei fraction represented by endosperm (END) (Supplementary Figure S3). These measurements were performed for a period 4–24 DAP and then in dry seeds (Results for endosperm are presented in the next subsection numbered 3.2.2.). Diploid seed tissues contained 2C and 4C nuclei representing G1 and G2 phases of the cell cycle, and 8C and 16C endoreduplicated nuclei originating from one and two endocycles, respectively (Supplementary Figure S3).

We found that all wild barleys contained similar amounts of 2C and 4C nuclei, each oscillating between 40 to 50%, with a minimal amount $\leq 10\%$ of endoreduplicated nuclei at 4 DAP. Up to 12 DAP the number of endopolyploid nuclei increased to reach the max-

imum, i.e., 10–24% for 8C and 5–11% for 16C. After 20 DAP, the 2C nuclei fraction increased, while the proportion of 4C, 8C, and 16C was gradually reduced. Finally, in mature dry seeds, 2C nuclei amounted to around half (50–60%), 4C nuclei around 30%, and 8C and 16C nuclei < 20% (Figure 3a; Supplementary Tables S1). AVOVA results showed that the values of these variables depended on both the type of environment and DAP and the interaction between these two factors (Figure 3c).

To estimate the degree of endoreduplication, we calculated the SCV parameter (Figure 3b; Supplementary Tables S2). At 4 DAP, a very low SCV of ≤ 0.09 was observed for all accessions. From 6 DAP onwards, the SCV increased to reach the peak at 12–24 DAP depending on the genotype. The highest SCV of 0.42, appeared in the two xeric accessions MG_x and WQ_x, at 16 and 24 DAP, respectively. Both accessions originate from the most southern collection sites (Figure 1). Similar to the previous observation, the values of these variables depended on the environment type and DAP, as well as the interaction between these two components (Figure 3c). The SCV curve for HS584 had a very smooth profile without any abrupt changes between neighborhood time points, and resembled the THm SCV line (Figure 3b).

Taken together, these data show endoreduplication variation in developing embryos and/or SMTs of wild barley seeds. The most southern xeric accessions show a tendency for a higher endopolyploidy level.

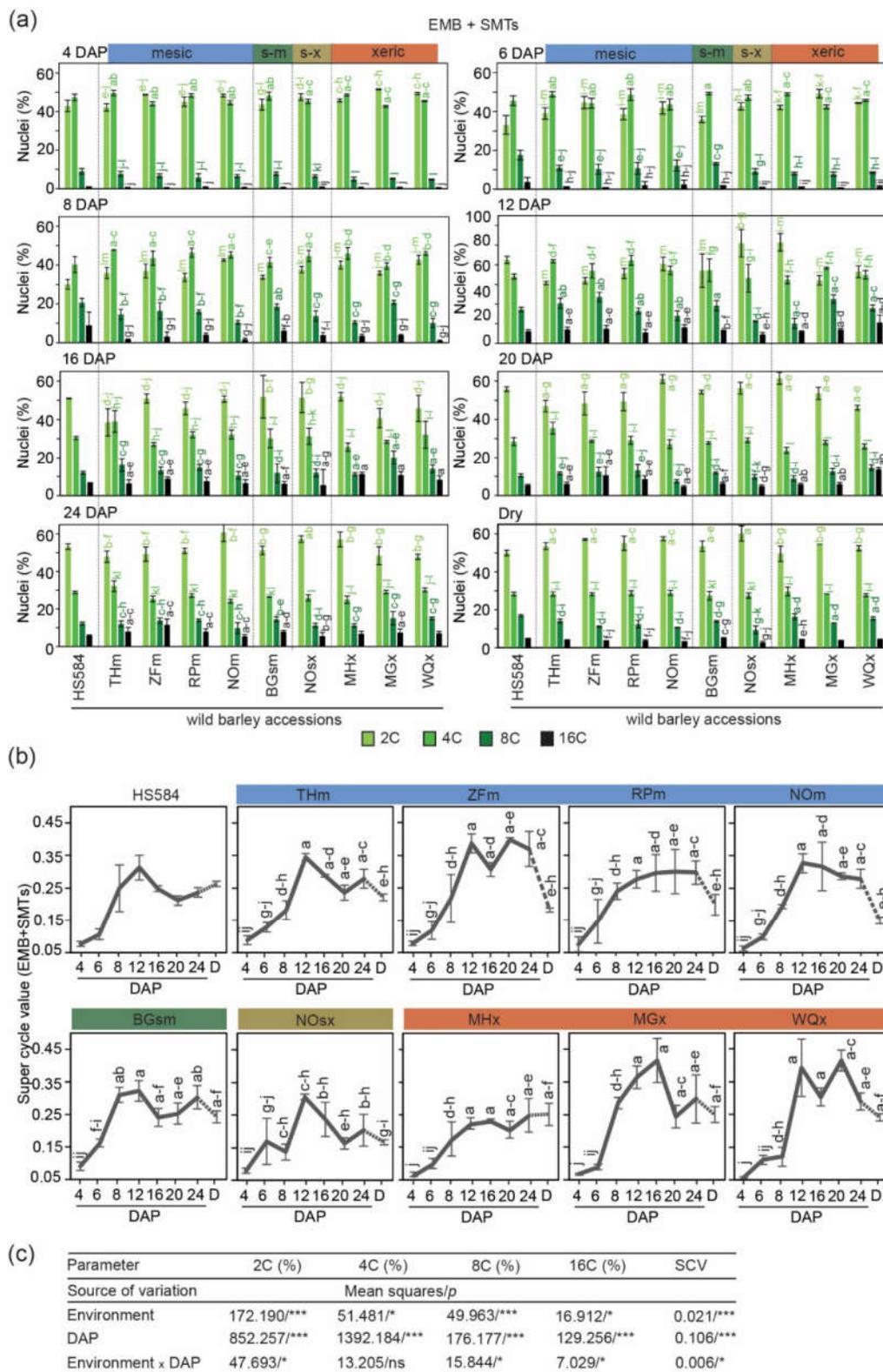


Figure 3. Estimation of C-values in diploid seed tissues represented by the embryo (EMB) and seed maternal tissues (SMTs) of ten wild barley accessions originating from Israel. (a) Percentage of 2C, 4C, 8C and 16C nuclei at a given day after pollination (DAP) and in dry seeds. Mesic accessions: THm = Tel Hai, ZFm = Zefat, RPm = Rosh Pinna, NOm = Nahal Oren NSF; semi-mesic accession: BGsm = Bar Giyyora; semi-xeric accession: NOsx = Nahal Oren SF; xeric accessions: MHx = Mehola, MGx = Machtsh Gadol, WQx = Wadi Qilt. The ecological conditions at the sampling site of HS584 are

unknown. Data are the means (\pm SD) from three biological replicates, each with at least 5 individual measurements (seeds). Data marked with the same letter do not differ according to the Duncan test ($p \leq 0.05$) (Source data are shown in Supplementary Tables S1 (b) Super cycle values at a given DAP calculated based on the data from (a), D = dry seed. The dashed line between 24 DAP and dry seed samples indicates further seed development after 24 DAP (Source data are shown in Supplementary Tables S2) (c) Summary of ANOVA performed for (a) and (b). The sources of variance were as follows: four environment types, eight-time point (DAP), and interaction between environment and DAP. *, *** Significant at $p \leq 0.05$, 0.001, respectively; ns—not significant.

3.2.2. Triploid Endosperm Tissues

Endosperm seed tissues contained four populations of nuclei, where 3C and 6C values reflected G1 and G2 phases of the mitotic cell cycle, and 12C and 24C nuclei resulted from one and two endocycles, respectively (Supplementary Figure S3). We calculated the frequencies of individual C-values in all ten wild barley accessions up to 24 DAP, and then in dry seeds (Figure 4a; Supplementary Tables S3). The inter-accession differences in endosperm C-values were striking already from the beginning of seed development. For instance, the frequency of 3C nuclei ranged from 50 to 80% (MHx vs. MGx, respectively), and 6C nuclei from 14 to 40% (inversely MGx vs. MHx, respectively) at 4 DAP. Only at this time point, all accessions contained a similar amount of endoreduplicated nuclei ($\leq 9\%$). From 6 to 24 DAP, the amount of 3C decreased approximately two times (from $\sim 60\%$ to $\sim 25\%$), the fraction of 6C nuclei maintained a constant level (around $\sim 30\%$), and the amount of 12C and 24C nuclei continuously increased up to 50% for MHx, MGx and WQx (all xeric accessions). In dry seeds, the fraction of 3C nuclei ranged from 31 to 47% (HS584 and NOm, respectively), 6C from 37 to 57% (NOsx vs. MHx, respectively), and endoreduplicated nuclei from 12 to 22% (NOsx vs. HS584, respectively). ANOVA results showed that the values of these variables depended on the environment type or DAP, but not the interaction between these two factors (Figure 4c).

At 4 and 6 DAP, the SCV corresponded to ~ 0.10 . From 8 DAP, the SCV started to increase to reach the peak at 12–24 DAP. During this period, xeric accessions showed generally higher SCV. For example, it was 0.78 for MHx (16 DAP), 0.61 for MGx (16–24 DAP), and 0.66 for WQx (20 DAP). In turn, accessions from the mesic environments showed a slightly lower SCV peak, ranging from 0.47 to 0.63 for RPm (20–24 DAP) and ZFm (16 DAP), respectively. The semi-mesic BGsm and semi-xeric NOsx accessions reached the maximum SCV of 0.60 (20–24 DAP) and 0.54 (6 DAP), respectively (Figure 4b; Supplementary Tables S4). ANOVA analysis revealed that the values of these variables were both environment- and DAP-, but not additively, dependent (Figure 4c). For HS584, the endosperm SCV profile was the most similar to THm and NO (Figure 4b).

Collectively, these results demonstrated that wild barley accessions reached the peak of endosperm endoreduplication at 12–24 DAP, and endopolyploidy level tended to be higher in xeric accessions.

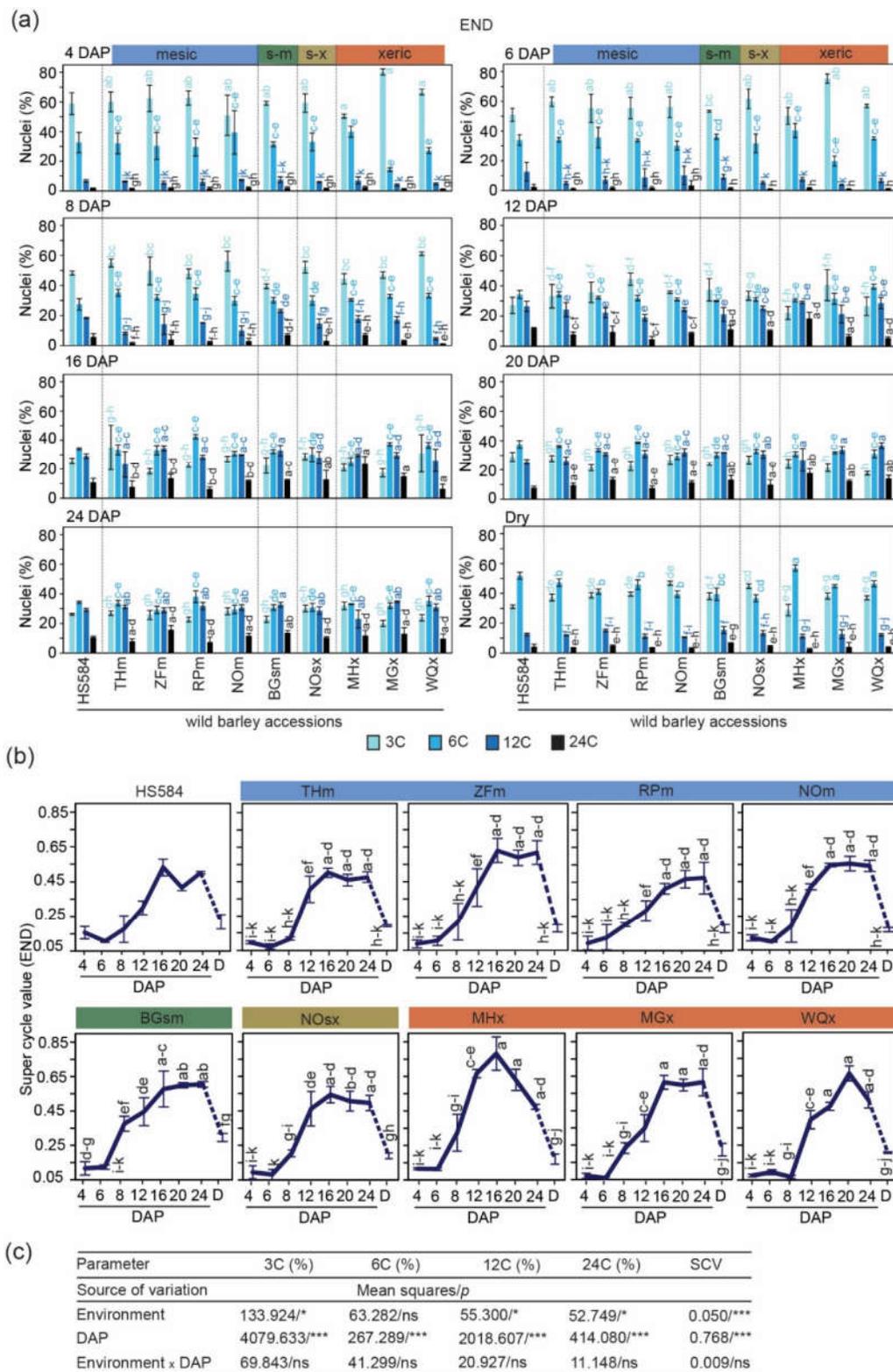


Figure 4. Estimation of C-values in triploid endosperm (END) tissues of ten analyzed wild barley accessions. (a) Percentage of 3C, 6C, 12C and 32C nuclei at a given day after pollination (DAP) and in dry seeds. Mesic accessions: THm = Tel Hai, ZFm = Zefat, RPm = Rosh Pinna, NOM = Nahal Oren NSF; semi-mesic accession: BGsm = Bar Giyyora; semi-xeric accession: NOsx = Nahal Oren SFS; xeric accessions: MHx = Mehola, MGx = Machtsh Gadol, WQx = Wadi Qilt. The ecological conditions at the sampling site of HS584 are unknown. Data are the means (\pm SD) from three biological replicates,

each with at least 5 individual measurements (seeds). Data marked with the same letter do not differ according to the Duncan test ($p \leq 0.05$) (Source data are shown in Supplementary Tables S3) (b) Super cycle values at a given DAP calculated based on the data from (a), D = dry seed. The dashed line between 24 DAP and dry seed samples indicates further seed development after 24 DAP (Source data are shown in Supplementary Tables S4). (c) Summary of ANOVA performed for (a) and (b). The sources of variance were as follows: four environment types, eight-time point (DAP), and interaction between environment and DAP. *, *** Significant at $p \leq 0.05$, 0.001, respectively; ns— not significant.

3.3. Comparison of Endoreduplication Dynamics in Developing Seeds of Wild and Cultivated Barley

Finding the differences between wild barley accessions, raised the question of whether it differs from cultivated barley. Therefore, we compared the data from the wild and the cultivated barley [11].

To provide a representative picture, we calculated the mean SCV at different DAP for diploid tissues (embryo and SMTs) and triploid endosperm for all ten wild barley accessions and nine barley cultivars (Figure 5a; Supplementary Tables S5). The cultivars were represented by six two-rowed [11] and three six-rowed genotypes (Supplementary Figure S4). We performed ANOVA to investigate the influence of two parameters: type of the sample (wild barley vs. cultivars) and age of the seeds (Figure 5a). For diploid tissues, ANOVA revealed differences depending on DAP and on the interaction of the sample type and DAP. Both wild and cultivated barleys achieved the highest mean SCV at 12 DAP (0.32—cultivars; 0.33—wild barley). The mean SCV for these two types varied significantly at 4–8 DAP, 20–24 DAP and in dry seeds.

For endosperm tissues, ANOVA revealed dependency of the SCV values on the sample type and DAP, and the interactions between these two factors. Mean SCV for endosperm tissues was higher for wild barley. In wild barley, SCV peaked at 16–20 DAP reaching the value ~0.55. Cultivated barley reached a sharp peak at 16 DAP with SCV ~0.46. The mean SCV for these two types varied significantly at 16–24 DAP and in dry seeds (Figure 5a).

To gain insight into the SCV relationships among the wild and cultivated barley, we performed PC analysis (Figure 5b). The first component (PC1) grouped samples based on DAP and showed similarity between individual experimental points of diploid tissues and endosperm development. For the mix of embryo/SMT nuclei, the SCV analysis revealed two groups: (i) 4 to 8 DAP and (ii) after 16 DAP. Similar sample distribution occurred for endosperm, excluding dry seed sample separated from the rest of time points. The second component (PC2) displayed the associations between the genotypes. The SCV data revealed two groups, the first formed by wild barley accessions and the second by barley cultivars.

Taken together, these data highlight the large inherent variation between wild and cultivated barley. Interestingly, the level of endoreduplication in endosperm tissues is higher in wild barley.

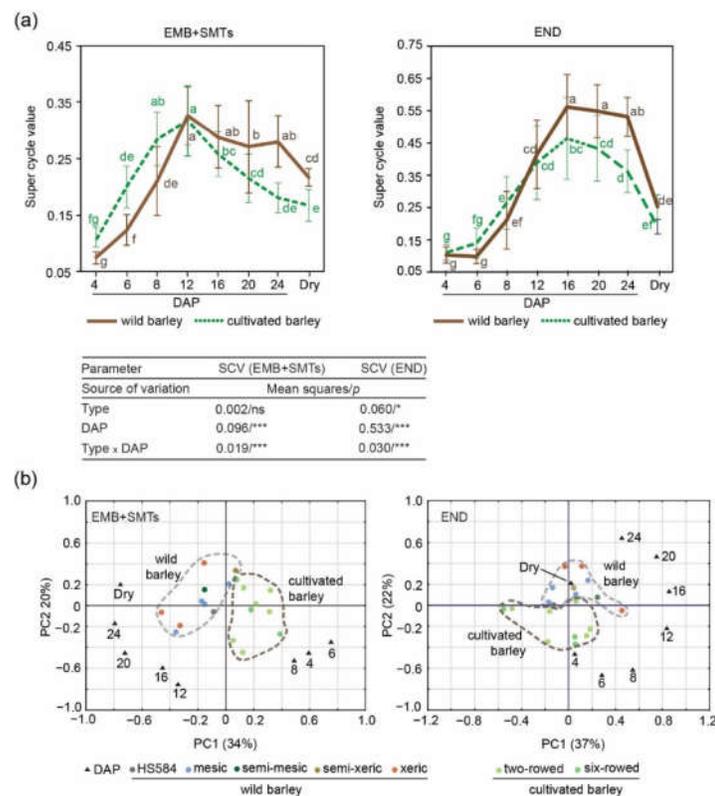


Figure 5. Comparison of the super cycle values (SCVs) for diploid seed tissues (EMB + SMTs) versus triploid endosperm (END) at given DAP between wild and cultivated barley. **(a)** Mean SCVs of EMB + SMTs and endosperm. Wild barley is represented by ten accessions (solid brown lines). Cultivated barley includes six two-rowed and three six-rowed cultivars (dashed green lines). Data for six-rowed cultivars are presented in Supplementary Figure S4. Data are the means (\pm SD) from three biological replicates, each with at least 5 individual measurements (Source data are presented in Supplementary Tables S5). Data marked with the same letter do not differ according to the Duncan test ($p \leq 0.05$). The sources of variance were as follows: two types of barley (wild and cultivated), eight-time point (DAP), and interaction between type and DAP. *, ***, Significant at $p \leq 0.05$, 0.001, respectively; ns—not significant. **(b)** Principal component (PC) analysis of SCVs in wild and cultivated barley. Numbers in the plots indicate DAP. The positions represent the contribution rates of the two main PCs to a given character. The dashed-line areas were added to highlight sample similarity.

3.4. Morphological and Cellular Changes during 48 Days of Wild Barley Seed Development

The need to explain the reasons for wild vs. cultivated barley endopolyploidy variation, inspired us to extend wild barley seed analysis in a broader experimental context and time. First, we monitored the dynamics of seed growth for HS584 from 4 to 48 DAP, and in dry seeds (Figure 6a; Supplementary Figure S5). At 4 DAP, SMTs constituted the dominant part of the seed. Both sagittal and transverse seed plans showed endosperm expansion, which accelerated seed growth from 6 DAP onwards (Supplementary Figure S5) and wild barley seeds reached the maximum growth (whole seed sagittal section areal $\sim 20 \text{ mm}^2$) at DAP 24. Interestingly, the endosperm changed color from green to gray from DAP 32, and the desiccation started to be visible from DAP 40. The most intense growth of the embryo occurred around DAP 16 (Supplementary Figure S5).

Next, we analyzed cell death which is a crucial cellular process during cereal seed development. To detect viable and non-viable cells, we performed Evans blue staining (Figure 6b). The stain penetrates the intracellular spaces of dead tissues and dyes them blue. Cell death followed a specific pattern in developing wild barley seeds. We detected regions of blue staining in the top (seed brush) and bottom parts of SMTs, but not in the

longitudinal elongation zone from 6 DAP onwards. In endosperm, very weak blue signals appeared in the central part at 12 DAP, and the area of staining and color intensity increased over time. AL was the only endosperm tissue free of staining at the end of seed development (Supplementary Figure S6). No staining was observed in the embryo at any stage of seed development.

Finally, we continued the measurement of nuclear C-values in seed tissues up to 48 DAP for the HS584, R_Pm, and BGsm accessions (Figure 6c; Supplementary Figure S7). For diploid seed tissues, the SCV profile differed between accessions. HS584 contained two peaks, at 12 and 32 DAP, reflecting probably the accumulation of endoreduplicated nuclei in SMTs and embryos, respectively. The R_Pm genotype had one symmetrical peak of endoreduplication from 6 to 28 DAP, which contrasted to very irregular peak of BGsm (Supplementary Figure S7). The SCV curve had a single broad peak profile for three studied accessions in endosperm tissues (Figure 6c). The differences concerned its width, reflecting the shifts between the start and end of the endopolyploidy period. Interestingly, the transition between low SCV at 48 DAP and its higher value in dry seed was very clear in both diploid and triploid seed tissues.

To summarize, these results demonstrate that (i) wild barley seed reaches the accumulation of growth at 20–24 DAP; (ii) SMTs and endosperm cells undergo cell death from 6 and 12 DAP, respectively; (iii) endoreduplication is more variable in a mixture of SMTs/embryo. All our observations suggest that wild barley seed ripening and desiccation continue after 48 DAP.

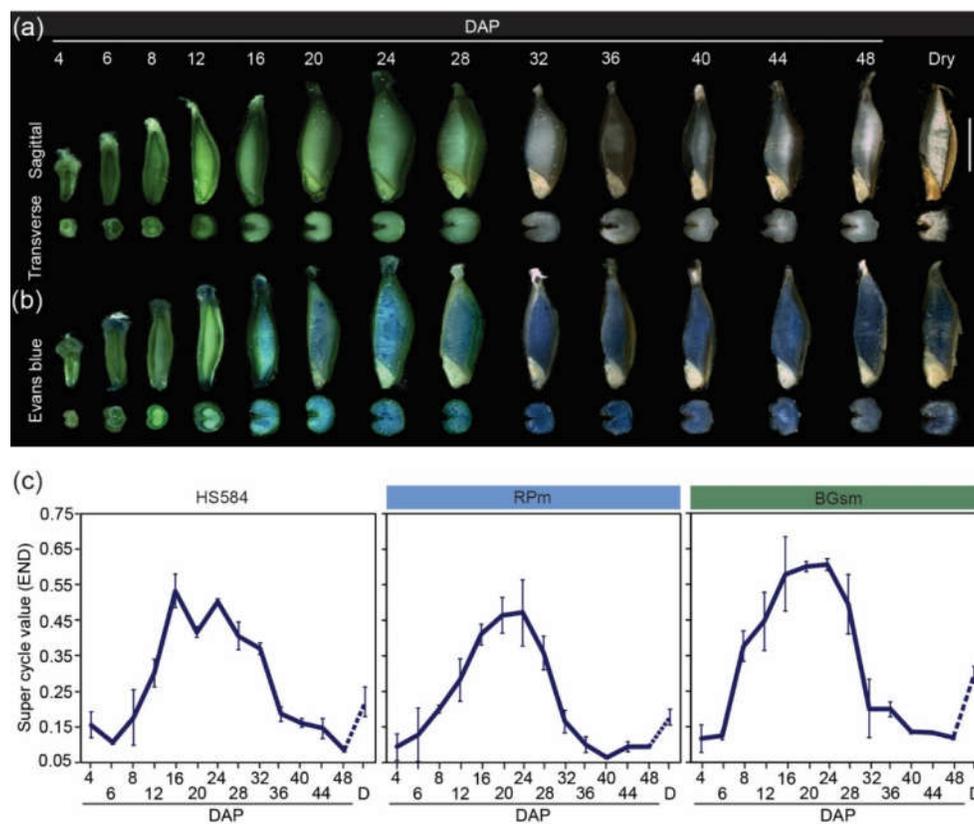


Figure 6. Time-course study of morphological and cellular events in developing wild barley seeds from 4 to 48 days after pollination (DAP). (a) and (b) Representative seed sections of HS584 (a) without and (b) with 0.1% Evans blue staining. The seeds shown are representative of at least 20 individuals not stained (a) and stained seed (b). Scale bar = 5 mm. (c) Endosperm super cycle values of three wild barley accessions: HS584, Rosh Pinna and Bar Giyyora. Complementary data for diploid seed tissues are presented in Supplementary Figure S7. Data are the means (\pm SD) from three biological replicates, each with at least 5 individual measurements (seeds). The dashed line between 48 DAP and dry seeds (D) samples represents the desiccation stage that was not analyzed in detail here.

4. Discussion

Wild barley, *H. vulgare* subsp. *spontaneum* is abundant in diverse ecogeographic regions in the Middle East and has been studied extensively from the phenotypic, genetic, and agronomic perspectives [37–40]. Wild barley plants growing in habitats with diverse environmental conditions, are exposed to numerous stressors, which directly influence their seed yield [39]. In wild barley, spike and seed traits were expressed so far only by spike length, grain number per spike or TKW [39]. Here, we focused on the size and biomass of dry wild barley kernels. We noted that the ranges of seed traits were much wider in subsp. *spontaneum* than in cultivated barley [11]. Interestingly, dry seeds of wild barleys were on average longer than in the cultivars. It seems that the wild barley seed length might be an interesting trait utilized by the breeders for seed yield improvement *per se*. One desert genotype—MHx deserves special attention. It is characterized by the lowest values of all measured seed features, which might be a result of a region-specific separation [40]. As expected, the seed biomass was higher in the cultivated barley because this is the main yield-related trait used during breeding [39]. The subsp. *spontaneum* accessions had on average longer awns than cultivated barleys, which is in agreement with previous observation [39]. Interestingly all accessions originating from the xeric environments possessed shorter awns, which is an example of an adaptation mechanism adjusting plants to the environment [40].

Until now, advanced methods detecting morphological and cellular changes during seed development have not been used, either from a domestication or stress adaptation point of view, in wild barley. Therefore, we investigated the dynamics of endoreduplication in the diploid and triploid seed tissues from the time shortly after pollination until dry seeds. In parallel, we monitored the morphological and PCD changes accompanying endoreduplication to understand better the complexity of wild barley seed formation (Figure 7).

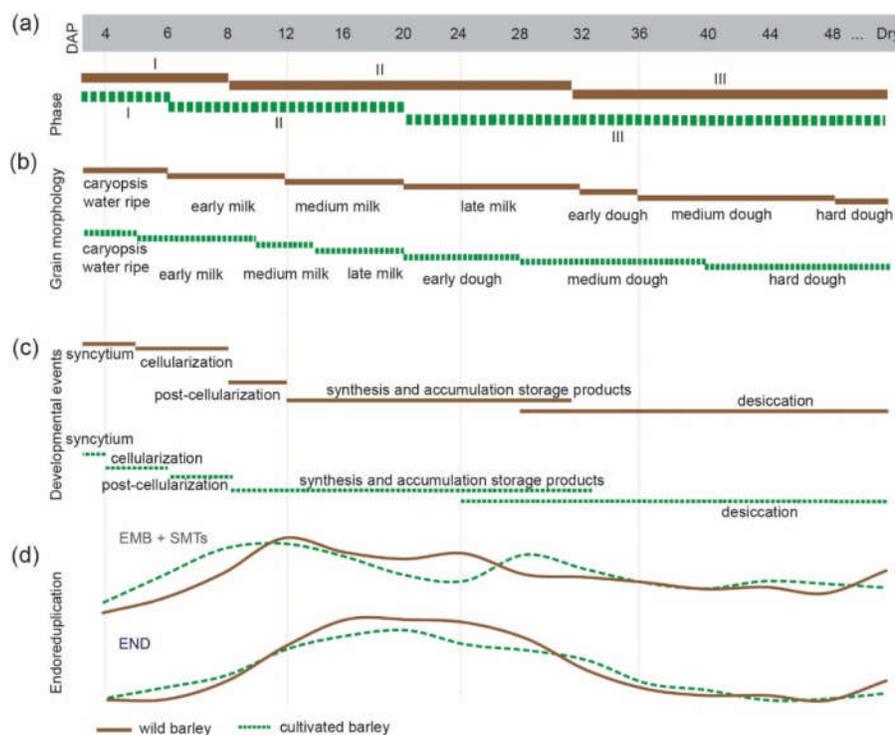


Figure 7. The model of phase transitions during wild (brown, solid lines) and cultivated barley [11] (green, dashed lines) seed development. Overview of (a) phase of barley grain development (based on [20,41]), and (b) grain morphological changes (based on [12]), (c) developmental events (based on [20,41,42]) and (d) endoreduplication dynamics. EMB + SMTs = embryo and seed maternal tissues and END = endosperm.

There are several parameters for quantification endoreduplication level [43]. Commonly used indicator is cycle value (a.k.a. endoreduplication index) [44]. However, this formula considers 4C (and 6C endosperm) nuclei as already endoreduplicated. Although some 4C nuclei (6C) might already be programmed for endoreduplication, others will be regularly cycling G2 nuclei. FCM does not recognize which 4C nuclei will undergo the mitotic cell cycle and which endocycle. Therefore, we recently introduced a new conservative formula, which considers that 8C nuclei (and 12C nuclei in endosperm) as the first unambiguous level of endoreduplication [11].

Many dicots possess non-endospermic seeds, where the developing embryo consumes most of the endosperm before the seed maturation. For the non-endospermic seeds, endoreduplication intensity is a marker of seed quality and maturation [36,45]. In contrast, grasses (*Poaceae*) have endospermic seeds which means that the endosperm forms the major and embryo the minor tissue mass of the fully developed seed. Besides, the high nutritional value of endosperm makes cereals the main crops worldwide to produce energy for humans and livestock. Endoreduplication appears during endosperm development and is correlated with the rapid growth of the caryopsis, the synthesis and accumulation of storage compounds, mainly starch and proteins in cereals [20]. We found that during seed development, both wild and cultivated barley endosperm underwent two rounds of endoreduplication resulting in 12C and 24C nuclei, respectively [11]. Two endocycles also appeared during wheat [46] and rice [47] endosperm development. Four and up to seven rounds of endoreduplication were found in the endosperm of sorghum [48] and maize [49], respectively. This suggests that the upper level of endopolyploidization is genetically regulated in the cereal endosperm, including *H. vulgare*. Further genetic variation most likely exists in the kinetics of endoreduplication which is suggested by the different SCV profiles observed in our study for different genotypes when grown under identical cultivation conditions. In wild barley, the major endoreduplication activity started ~8 DAP, i.e., two days later compared to the cultivated barley [11]. In both taxa, the SCV decreased after 32 DAP. The study performed in cultivated barley has already shown that endosperm endoreduplication nuclei were progressively degraded during the accumulation of the storage materials and ripening. This degeneration was initiated in highly endopolyploid nuclei and accompanied by accumulation of DNA damage and cell death [11]. Interestingly, we detected high proportion of endoreduplicated nuclei in dry seeds for subsp. *spontaneum*. Admittedly, desiccating and dry seeds of cultivated barley also contained endoreduplicated nuclei [11], but not in such proportion as in wild barley. Microscopic observations confirmed that endoreduplicated nuclei originated from AL in dry barley seeds [50]. Endoreduplicated AL nuclei are not observed in other cereals except for barley [50].

Here, we also found that wild barley has shifted the major seed/endosperm morphological and developmental phases and needs more time to complete seed ripening comparing to cultivated barley (Figure 6a–c) [11]. Delayed desiccation period and entrance into dough phases were the most obvious differences between wild and cultivated strains [12]. Our findings complement previous observations of several days difference in the heading and anthesis in wild *versus* cultivated barley [39]. Furthermore, the gray color of maturing endosperm in subsp. *spontaneum*, comparing to the white-yellow color of endosperm in cultivated barley [11], may reflect distinct compositions of storage compounds. So far, transcriptomic and metabolomic profiles of seed storage compounds are available for cultivated but not for wild barley [51]. The darker color of wild barley endosperm may also indicate the presence of secondary metabolites, e.g., anthocyanins or other reactive oxygen species scavenging molecules [52,53]. Taken together, all results collected for wild and cultivated barleys raised the question whether there is a link between higher endoreduplication level and different color of endosperm in wild barley. However, solving this question will require further studies, for example examination of the secondary metabolites using high-performance liquid chromatography. Extended analysis may help to better understand the mechanisms of stress adaptation and cereal seed improvement.

Based on the studies in cultivated barley, we concluded that endoreduplication in SMTs is correlated with starch deposition, and in embryos with differentiation of the tissues [11]. We detected two populations of endopolyploid nuclei (8C and 16C reflecting one and two endocycles, respectively) in a mixture of diploid seed tissues of wild barley which is similar to cultivated barley [11]. This is additional evidence suggesting that the number of endocycles is genetically controlled and species-specific [54]. We noted that in the mixture of SMTs/embryo of subsp. *spontaneum* seeds, that endoreduplication peaked two times at 12 and 24 DAP. Comparing wild and cultivated barley, the level of endoreplication expressed by SCV was the same, however, the peaks were shifted. We assign the first endoreduplication peak to SMTs, which correlates with their intensive growth. We assume that the second endoreduplication peak should be attributed to the embryo, and correlate with its rapid growth and tissue differentiation. Similarly to the endosperm, wild barley dry seed tissues contained higher proportion of endoreduplicated nuclei as compared to 48 DAP sample. However, this observation was exclusive only for wild, not cultivated barley (Figure 6). Among ten wild barley accessions, we found variation in the dynamics of SMTs/embryo and endosperm endopolyploidization. On the one hand, some accessions had shifted in time the endoreduplication peak, on the other hand, there were differences with SCV throughout the entire seed development. This finding unravels a new level of variation between wild barley populations. However, it has to be noted that the accessions used in our report represent only a limited diversity and that the variation is probably much greater in wild barley. Intra-specific endopolyploidy variation is quite common in both cultivated [39,40] and wild [55] plants. Studies performed in *Arabidopsis* revealed that endoreduplication levels are controlled by the interaction of multiple cell cycle-related genes [55].

Importantly, we detected a link between the amount of endoreduplicated nuclei and the ecogeographical origin of the wild barley accessions. Namely, accessions originating from the xeric environments tended to have higher SCV for both SMTs/embryo and endosperm tissues. This is analogous with the previous findings for Israeli accessions of wild barley, in the context of genetic variability detected by molecular markers [56]. Many studies have found endoreduplication more abundantly among plants that grow under environmentally challenging conditions [21,22,27,57]. Increasing DNA content may be integrated into the damage-induced oxidative stress-response systems, like for instance pentose phosphate pathway [58]. In this system, endoreduplication may promote compensation to damages by upregulation of gene expression involved in the overproduction of metabolites [58]. On the other hand, endopolyploidy is thought to play significant roles in plant physiology [21]. Altered phytohormone balances, changed after exposition to environmental stressors, probably trigger organ-specific endopolyploidization [24]. This may suggest an adaptive mechanism to an environmental gradient that results in differential endopolyploidy [24]. With only ten accessions used in this study, the identification of an obvious adaptive response to harsh environmental conditions is not conclusive. Therefore, to identify a potential link between environmental gradient and seed endoreduplication variation, future studies involving a larger number of genetically defined samples and mapping causal genes are necessary.

5. Conclusions and Future Perspectives

Both diploid and triploid barley seed tissues undergo two endocycles. This study of endoreduplication in wild barley seeds revealed a new level of variation appearing within subsp. *spontaneum*. Wild barley had a higher endoreduplication level in endosperm tissues comparing with the with cultivated one and the amount of endoreduplicated nuclei tended to be higher in xeric accessions. We are currently aiming to better understand how spatiotemporal seed endoreduplication patterns change under various stresses and whether these stresses are linked to stress adaptation.

Supplementary Materials: The following are available online at www.mdpi.com/2073-4425/12/5/711/s1, Figure S1: Phenotypic analysis of dry seeds of wild barley accessions originating from Israel; Figure S2: Principal component (PC) analysis of TKW, seed length and width for peeled seeds, and awn length; Figure S3: An example showing the interpretation of histograms of nuclear DNA content; Figure S4: Comparison of super cycle values in seed tissues of three six-rowed barley cultivars; Figure S5: Time-course study of wild barley HS584 seed growth; Figure S6: Analysis the aleurone layer viability; Figure S7: Comparison of super cycle values in diploid seed tissues of three selected wild barley accessions at given DAP; Tables S1: Percentage of 2C, 4C, 8C and 16C nuclei at a given DAP and in dry seeds in diploid seed tissues; Tables S2: Super cycle values at a given DAP calculated for SMTs; Tables S3: Percentage of 3C, 6C, 12C and 24C nuclei at a given DAP and in dry seeds in endosperm; Tables S4: Super cycle values at a given DAP calculated for endosperm; Tables S5: Super cycle values at a given DAP calculated for embryo/seed maternal tissues and endosperm.

Author Contributions: Conceptualization, A.P., A.N. and T.K.; methodology, A.N.; software, A.N. and P.P.S.; validation, A.N. and P.P.S.; formal analysis, A.N. and A.P.; investigation, A.N., P.P.S., M.K., D.W. and B.T.; resources, T.K.; data curation, A.N. and P.P.S.; writing—original draft preparation, A.N. and P.P.S.; writing—review and editing, A.P.; visualization, A.N.; supervision, A.P.; project administration, A.P.; funding acquisition, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Czech Science Foundation, grants numbers 18-12197S (A.P.) and 21-02929S (A.P.) and 20-25845Y (P.P.S.), Purkyně Fellowship from the Czech Academy of Sciences to A.P., and the ERDF project “Plants as a tool for sustainable global development” (No. CZ.02.1.01/0.0/0.0/16_019/0000827). Also, we thank COST action 16212 Impact of Nuclear Domains On Gene Expression and Plant Traits (INDEPTH) for creating the ground for developing this project and supporting meetings of the authors during its realization.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Marion Röder (Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Germany) for providing seeds of *H. vulgare* subsp. *spontaneum* HS584, Petr Capal for the maintenance of flow cytometers, Sariel Hubner for the photo of xeric site barley, Eva Jahnová and Suad Chalifa for technical assistance, and Zdenka Bursová for plant care.

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.

References

1. Badr, A.; Müller, K.; Schäfer-Pregl, R.; El Rabey, H.; Effgen, S.; Ibrahim, H.H.; Pozzi, C.; Rohde, W.; Salamini, F. On the origin and domestication history of barley (*Hordeum vulgare*). *Mol. Biol. Evol.* **2000**, *17*, 499–510, doi:10.1093/oxfordjournals.molbev.a026330.
2. Mascher, M.; Schuenemann, V.J.; Davidovich, U.; Marom, N.; Himmelbach, A.; Hübner, S.; Korol, A.; David, M.; Reiter, E.; Riehl, S.; et al. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* **2016**, *48*, 1089–1093, doi:10.1038/ng.3611.
3. Jakob, S.S.; Rödder, D.; Engler, J.O.; Shaaf, S.; Özkan, H.; Blattner, F.R.; Kilian, B. Evolutionary history of wild barley (*Hordeum vulgare* subsp. *spontaneum*) analyzed using multilocus sequence data and paleodistribution modeling. *Genome Biol. Evol.* **2014**, *6*, 685–702, doi:10.1093/gbe/evu047.
4. Krugman, T.; Nevo, E.; Beharav, A.; Sela, H.; Fahima, T. The Institute of Evolution Wild Cereal Gene Bank at the University of Haifa. *Isr. J. Plant. Sci.* **2019**, *65*, 129–146, doi:10.1163/22238980-00001065.
5. Von Korff, M.; Wang, H.; Léon, J.; Pillen, K. Development of candidate introgression lines using an exotic barley accession (*Hordeum vulgare* ssp. *spontaneum*) as donor. *Theor. Appl. Genet.* **2004**, *109*, 1736–1745, doi:10.1007/s00122-004-1818-2.
6. Nevo, E. Origin, evolution, population genetics and resources for breeding of wild barley, *Hordeum spontaneum*, in the Fertile Crescent. In *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*; Shewry, P., Ed.; Centre for Agriculture and Bioscience International: Wallingford, UK, 1992; pp. 19–43. ISBN 0-85198-725-7.
7. Fetch, T.G.; Steffenson, B.J.; Nevo, E. Diversity and sources of multiple disease resistance in *Hordeum spontaneum*. *Plant. Dis.* **2003**, *87*, 1439–1448, doi:10.1094/PDIS.2003.87.12.1439.

8. Grossi, M.; Giorni, E.; Rizza, F.; Stanca, A.M.; Cattivelli, L. Wild and cultivated barleys show differences in the expression pattern of a cold-regulated gene family under different light and temperature conditions. *Plant. Mol. Biol.* **1998**, *38*, 1061–1069, doi:10.1023/A:1006079916917.
9. Cai, K.; Chen, X.; Han, Z.; Wu, X.; Zhang, S.; Li, Q.; Nazir, M.M.; Zhang, G.; Zeng, F. Screening of worldwide barley collection for drought tolerance: The assessment of various physiological measures as the selection criteria. *Front. Plant. Sci.* **2020**, *11*, 1159, doi:10.3389/fpls.2020.01159.
10. Nevo, E.; Krugman, T.; Beiles, A. Genetic resources for salt tolerance in the wild progenitors of wheat (*Triticum dicoccoides*) and barley (*Hordeum spontaneum*) in Israel. *Plant. Breed.* **1993**, *110*, 338–341, doi:10.1111/j.1439-0523.1993.tb00599.x.
11. Nowicka, A.; Kovacik, M.; Tokarz, B.; Vrána, J.; Zhang, Y.; Weigt, D.; Doležel, J.; Pecinka, A. Dynamics of endoreduplication in developing barley seeds. *J. Exp. Bot.* **2021**, *72*, 268–282, doi:10.1093/jxb/eraa453.
12. Tottman, D.R.; Makepeace, R.J.; Broad, H. An explanation of the decimal code for the growth stages of cereals, with illustrations. *Ann. Appl. Biol.* **1979**, *93*, 221–234, doi:10.1111/j.1744-7348.1979.tb06534.x.
13. Evers, T.; Millar, S. Cereal grain structure and development: Some implications for quality. *J. Cereal Sci.* **2002**, *36*, 261–284, doi:10.1006/jcrs.2002.0435.
14. Radchuk, V.; Borisjuk, L. Physical, metabolic and developmental functions of the seed coat. *Front. Plant. Sci.* **2014**, *5*, 510, doi:10.3389/fpls.2014.00510.
15. Evers, A.D.; Blakeney, A.; O'Brien, L. Cereal structure and composition. *Aust. J. Agric. Res.* **1999**, *50*, 629–650.
16. Olsen, O.-A. Endosperm development: Cellularization and cell fate specification. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **2001**, *52*, 233–267, doi:10.1146/annurev.arplant.52.1.233.
17. Rodríguez, M.V.; Barrero, J.M.; Corbineau, F.; Gubler, F.; Benech-Arnold, R.L. Dormancy in cereals (not too much, not so little): About the mechanisms behind this trait. *Seed Sci. Res.* **2015**, *25*, 99–119, doi:10.1017/S0960258515000021.
18. D'amato, F. Endopolyploidy as a factor in plant tissue development. *Caryologia* **1964**, *17*, 41–52, doi:10.1080/00087114.1964.10796115.
19. Nagl, W. DNA endoreduplication and polyteny understood as evolutionary strategies. *Nature* **1976**, *261*, 614–615, doi:10.1038/261614a0.
20. Sabelli, P.A.; Larkins, B.A. The development of endosperm in grasses. *Plant. Physiol.* **2009**, *149*, 14–26, doi:10.1104/pp.108.129437.
21. De Veylder, L.; Larkin, J.C.; Schnittger, A. Molecular control and function of endoreplication in development and physiology. *Trends Plant. Sci.* **2011**, *16*, 624–634, doi:10.1016/j.tplants.2011.07.001.
22. Scholes, D.R.; Paige, K.N. Plasticity in ploidy: A generalized response to stress. *Trends Plant. Sci.* **2015**, *20*, 165–175, doi:10.1016/j.tplants.2014.11.007.
23. Gendreau, E.; Traas, J.; Desnos, T.; Grandjean, O.; Caboche, M.; Hofte, H. Cellular basis of hypocotyl growth in *Arabidopsis thaliana*. *Plant. Physiol.* **1997**, *114*, 295–305, doi:10.1104/pp.114.1.295.
24. Bhosale, R.; Boudolf, V.; Cuevas, F.; Lu, R.; Eekhout, T.; Hu, Z.; Van Isterdael, G.; Lambert, G.M.; Xu, F.; Nowack, M.K.; et al. A spatiotemporal dna endoploidy map of the *Arabidopsis* root reveals roles for the endocycle in root development and stress adaptation. *Plant. Cell* **2018**, *30*, 2330–2351, doi:10.1105/tpc.17.00983.
25. Lingua, G.; Fusconi, A.; Berta, G. The nucleus of differentiated root plant cells: Modifications induced by arbuscular mycorrhizal fungi. *Eur. J. Histochem.* **2001**, *45*, 9–20, doi:10.4081/1609.
26. Chandran, D.; Inada, N.; Hather, G.; Kleindt, C.K.; Wildermuth, M.C. Laser microdissection of *Arabidopsis* cells at the powdery mildew infection site reveals site-specific processes and regulators. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 460–465, doi:10.1073/pnas.0912492107.
27. Cookson, S.J.; Radziejewski, A.; Granier, C. Cell and leaf size plasticity in *Arabidopsis*: What is the role of endoreduplication? *Plant. Cell Environ.* **2006**, *29*, 1273–1283, doi:10.1111/j.1365-3040.2006.01506.x.
28. Tran, V.; Weier, D.; Radchuk, R.; Thiel, J.; Radchuk, V. Caspase-like activities accompany programmed cell death events in developing barley grains. *PLoS ONE* **2014**, *9*, e109426, doi:10.1371/journal.pone.0109426.
29. Radchuk, V.; Tran, V.; Radchuk, R.; Diaz-Mendoza, M.; Weier, D.; Fuchs, J.; Riewe, D.; Hensel, G.; Kumlehn, J.; Munz, E.; et al. Vacuolar processing enzyme 4 contributes to maternal control of grain size in barley by executing programmed cell death in the pericarp. *New Phytol.* **2018**, *218*, 1127–1142, doi:10.1111/nph.14729.
30. Young, T.E.; Gallie, D.R. Programmed cell death during endosperm development. *Plant. Mol. Biol.* **2000**, *44*, 283–301, doi:10.1023/A:1026588408152.
31. Wu, X.; Liu, J.; Li, D.; Liu, C.M. Rice caryopsis development II: Dynamic changes in the endosperm. *J. Integr. Plant. Biol.* **2016**, *58*, 786–798, doi:10.1111/jipb.12488.
32. Nevo, E. Evolution of wild barley at “Evolution Canyon”: Adaptation, speciation, pre-agricultural collection, and barley improvement. *Isr. J. Plant. Sci.* **2015**, *62*, 22–32, doi:10.1080/07929978.2014.940783.
33. Waddington, S.R.; Cartwright, P.M.; Wall, P.C. A quantitative scale of spike initial and pistil development in barley and wheat. *Ann. Bot.* **1983**, *51*, 119–130, doi:10.1093/oxfordjournals.aob.a086434.
34. Kovacik, M.; Nowicka, A.; Pecinka, A. Isolation of high purity tissues from developing barley seeds. *J. Vis. Exp.* **2020**, 1–2, doi:10.3791/61681.
35. Nevo, E.; Zohary, D.; Brown, A.H.D.; Haber, M. Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. *Evolution* **1979**, *33*, 815–833, doi:10.1038/hdy.1988.88.

36. Rewers, M.; Sliwinska, E. Endoreduplication intensity as a marker of seed developmental stage in the Fabaceae. *Cytom. Part A* **2012**, *81A*, 1067–1075, doi:10.1002/cyto.a.22202.
37. Nevo, E.; Beiles, A.; Kaplan, D.; Storch, N.; Zohary, D. Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum* (Poaceae), in Iran. *Plant. Syst. Evol.* **1986**, *153*, 141–164.
38. Hosseini, M.; Ghorbani, R.; Rashed Mohassel, M.H.; Yassaie, M. Correlation of environmental factors and phenotypic diversity of Iranian wild barley (*Hordeum Spontaneum* Koch) populations. *Acta Oecologica* **2019**, *95*, 19–25, doi:10.1016/j.actao.2018.12.004.
39. Shakhatareh, Y.; Haddad, N.; Alrababah, M.; Grandi, S.; Ceccarelli, S. Phenotypic diversity in wild barley (*Hordeum vulgare* L. ssp. *spontaneum* (C. Koch) Thell.) accessions collected in Jordan. *Genet. Resour. Crop. Evol.* **2010**, *57*, 131–146, doi:10.1007/s10722-009-9457-8.
40. Volis, S.; Mendlinger, S.; Ward, D. Adaptive traits of wild barley plants of Mediterranean and desert origin. *Oecologia* **2002**, *133*, 131–138, doi:10.1007/s00442-002-0999-0.
41. Dante, R.A.; Larkins, B.A.; Sabelli, P.A. Cell cycle control and seed development. *Front. Plant. Sci.* **2014**, *5*, 493, doi:10.3389/fpls.2014.00493.
42. Sreenivasulu, N.; Borisjuk, L.; Junker, B.H.; Mock, H.-P.; Rolletschek, H.; Seiffert, U.; Weschke, W.; Wobus, U. Barley grain development: Toward an integrative view. In *International Review of Cell and Molecular Biology*; Jeon, K., Ed.; Elsevier: Amsterdam, The Netherlands, 2010; Volume 281, pp. 49–89.
43. Rewers, M.; Sadowski, J.; Sliwinska, E. Endoreduplication in cucumber (*Cucumis sativus*) seeds during development, after processing and storage, and during germination. *Ann. Appl. Biol.* **2009**, *155*, 431–438, doi:10.1111/j.1744-7348.2009.00362.x.
44. Barow, M.; Meister, A. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. *Plant. Cell Environ.* **2003**, *26*, 571–584, doi:10.1046/j.1365-3040.2003.00988.x.
45. Sliwinska, E.; Lukaszewska, E. Polysomaty in growing in vitro sugar-beet (*Beta vulgaris* L.) seedlings of different ploidy level. *Plant. Sci.* **2005**, *168*, 1067–1074, doi:10.1016/j.plantsci.2004.12.003.
46. Chojecki, A.; Gale, M.; Bayliss, M. The number and sizes of starch granules in the wheat endosperm, and their association with grain weight. *Ann. Bot.* **1986**, *58*, 819–831, doi:10.1093/oxfordjournals.aob.a087264.
47. Kobayashi, H. Variations of endoreduplication and its potential contribution to endosperm development in rice (*Oryza sativa* L.). *Plant. Prod. Sci.* **2019**, *22*, 227–241, doi:10.1080/1343943X.2019.1570281.
48. Kladnik, A.; Chourey, P.S.; Pring, D.R.; Dermastia, M. Development of the endosperm of Sorghum bicolor during the endoreduplication-associated growth phase. *J. Cereal Sci.* **2006**, *43*, 209–215, doi:10.1016/j.jcs.2005.09.004.
49. Sabelli, P.A.; Larkins, B.A. Regulation and function of retinoblastoma-related plant genes. *Plant. Sci.* **2009**, *177*, 540–548, doi:10.1016/j.plantsci.2009.09.012.
50. Keown, A.C.; Taiz, L.; Jones, R.L. The nuclear DNA content of developing barley aleurone cells. *Am. J. Bot.* **1977**, *64*, 1248–1253.
51. Sreenivasulu, N.; Usadel, B.; Winter, A.; Radchuk, V.; Scholz, U.; Stein, N.; Weschke, W.; Strickert, M.; Close, T.J.; Stitt, M.; et al. Barley grain maturation and germination: Metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant. Physiol.* **2008**, *146*, 1738–1758, doi:10.1104/pp.107.111781.
52. Burešová, V.; Kopecký, D.; Bartoš, J.; Martinek, P.; Watanabe, N.; Vyhnánek, T.; Doležel, J. Variation in genome composition of blue-aleurone wheat. *Theor. Appl. Genet.* **2015**, *128*, 273–282, doi:10.1007/s00122-014-2427-3.
53. Jeewani, D.C.; Hua, W.Z. Grain color development and segregation of blue wheat. *Res. J. Biotechnol.* **2017**, *12*, 40–45.
54. Chevalier, C.; Nafati, M.; Mathieu-Rivet, E.; Bourdon, M.; Frangne, N.; Cheniclet, C.; Renaudin, J.P.; Gvaudant, F.; Hernould, M. Elucidating the functional role of endoreduplication in tomato fruit development. *Ann. Bot.* **2011**, *107*, 1159–1169, doi:10.1093/aob/mcq257.
55. Sterken, R.; Kiekens, R.; Boruc, J.; Zhang, F.; Vercauteren, A.; Vercauteren, I.; De Smet, L.; Dhondt, S.; Inzé, D.; De Veylder, L.; et al. Combined linkage and association mapping reveals CYCD5;1 as a quantitative trait gene for endoreduplication in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4678–4683, doi:10.1073/pnas.1120811109.
56. Hübner, S.; Höffken, M.; Oren, E.; Haseneyer, G.; Stein, N.; Graner, A.; Schmid, K.; Fridman, E. Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Mol. Ecol.* **2009**, *18*, 1523–1536, doi:10.1111/j.1365-294X.2009.04106.x.
57. Barow, M. Endopolyploidy in seed plants. *BioEssays* **2006**, *28*, 271–281, doi:10.1002/bies.20371.
58. Siddappaji, M.H.; Scholes, D.R.; Bohn, M.; Paige, K.N. Overcompensation in response to herbivory in Arabidopsis thaliana: The role of glucose-6-phosphate dehydrogenase and the oxidative pentose-phosphate pathway. *Genetics* **2013**, *195*, 589–598, doi:10.1534/genetics.113.154351

Appendix V

Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues

Nowicka, A., Ferková, L., Said, M., **Kovacik, M.**, Zwyrtková, J., Baroux, C.,
and Pecinka, A.

Journal of Experimental Botany

DOI: 10.1093/jxb/erad036

IF (2023): 6.9

RESEARCH PAPER

Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues

Anna Nowicka^{1,2,*}, Ľuboslava Ferková¹, Mahmoud Said^{1,3}, Martin Kovacik¹, Jana Zwyrková¹,
Célia Baroux⁴ and Ales Pecinka^{1,*}

¹ Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany of the Czech Academy of Sciences, Šlechtitelů 31, 779 00 Olomouc, Czech Republic

² The Polish Academy of Sciences, The Franciszek Górski Institute of Plant Physiology, Niezapominajek 21, 30-239 Krakow, Poland

³ Field Crops Research Institute, Agricultural Research Centre, 9 Gamma Street, Giza, Cairo, 12619, Egypt

⁴ Department of Plant and Microbial Biology, Zürich-Basel Plant Science Center, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

* Correspondence: pecinka@ueb.cas.cz or nowicka@ueb.cas.cz

Received 26 October 2022; Editorial decision 16 January 2023; Accepted 25 January 2023

Editor: Zoe Wilson, University of Nottingham, UK

Abstract

Rabl organization is a type of interphase chromosome arrangement with centromeres and telomeres clustering at opposite nuclear poles. Here, we analyzed nuclear morphology and chromosome organization in cycling and endoreduplicated nuclei isolated from embryo and endosperm tissues of developing barley seeds. We show that endoreduplicated nuclei have an irregular shape, less sister chromatid cohesion at 5S rDNA loci, and a reduced amount of centromeric histone CENH3. While the chromosomes of the embryo and endosperm nuclei are initially organized in Rabl configuration, the centromeres and telomeres are intermingled within the nuclear space in the endoreduplicated nuclei with an increasing endoreduplication level. Such a loss of chromosome organization suggests that Rabl configuration is introduced and further reinforced by mitotic divisions in barley cell nuclei in a tissue- and seed age-dependent manner.

Keywords: Barley, CENH3, CEREBA repeat, embryo, endoreduplication, endosperm, *Hordeum vulgare*, nuclear organization, Rabl configuration.

Introduction

Chromosome structure and organization play important roles in replication, transcription, and genome repair (Misteli, 2020). Their organization includes the formation of nucleosomes, as the basic unit of chromatin, and their assembly into higher order domains. These domains represent different chromatin states characterized by specific histone and/or DNA modifications, and vary in their transcription, replica-

tion, or DNA repair patterns (Roudier *et al.*, 2011; Sequeira-Mendes *et al.*, 2014). Individual interphase chromosomes occupy a specific nuclear space known as chromosome territories (CTs) (reviewed in, for example, Schubert and Shaw, 2011; Grob, 2020).

In *Arabidopsis thaliana* (Arabidopsis) with a small and repeat-poor genome, the peri(centromeric) regions form

Abbreviations: CENH3, centromere-specific histone H3; CEREBA, CENTROMERIC RETROELEMENT OF BARLEY; CT, chromosome territory; DAP, days after pollination; FISH, fluorescence *in situ* hybridization; FITC, fluorescein isothiocyanate; mip, maximum intensity projection; msv, middle slide view; NA, nucleus area; NCI, nucleus circularity index; ND, nucleus diameter; NP, nucleus perimeter; ROI, region of interest; SC, sister chromatid.

heterochromatic chromocenters from which euchromatic chromosome arms emanate and the telomeric regions often surround the nucleolus (Fransz *et al.*, 2002, 2003). The CTs of 45S *rDNA*-bearing chromosomes cluster preferentially around the nucleolus, but the CTs of the remaining chromosomes are positioned randomly in roots and leaves (Pecinka *et al.*, 2004) or show pairwise CT associations in the seed endosperm (Baroux *et al.*, 2017). In contrast, plants with large and repeat-rich genomes often harbor Rabl chromosome organization, with the centromeres and telomeres clustering at the opposite nuclear poles (Santos and Shaw, 2004), first described in 1885 by the cytologist and anatomist Carl Rabl based on nuclei of *Caudata* (Rabl, 1885). Because Rabl organization is widespread in cereals with large genomes such as bread wheat (17 Gbp/1C) or barley (5.1 Gbp/1C), and diminishes with a decreasing genome size in, for example, maize (2.4 Gbp/1C) or rice (0.43 Gbp/1C) (Fujimoto *et al.*, 2005a), it has been hypothesized that it is determined by the nuclear genome size (reviewed in, for example, Santos and Shaw, 2004). Although this correlation holds true over distantly related phylogenetic groups such as *Poaceae* and *Brassicaceae* (Němečková *et al.*, 2020; Shan *et al.*, 2021), it is not universal. For example, the majority of root nuclei (but not leaf nuclei) show Rabl in the small genome grass *Brachypodium distachyon* (0.35 Gbp/1C), while some *Liliaceae* species with giant genomes (~35–50 Gbp/1C) lack Rabl organization (Fujimoto *et al.*, 2005a; Idziak *et al.*, 2015). Therefore, other hypotheses suggest that Rabl might be a preserved organization of mitotic chromosomes. However, why this configuration is maintained in some but not other species remains unclear.

To investigate the relationships between occurrence of the Rabl configuration, endoreduplication, and tissue age, we made use of the endosperm and to a smaller extent embryo tissues of barley grains (Nowicka *et al.*, 2021a). In cereals, the endosperm progresses through several stages of development (reviewed in, for example, Olsen, 2001; Sabelli and Larkins, 2009), connected with major changes in transcriptional regulation (Sreenivasulu *et al.*, 2004; Pfeifer *et al.*, 2014). The endosperm forms most of the seed mass and serves as the main energy storage tissue for the embryo during germination. In contrast to vegetative tissues, cereal endosperm shows endoreduplication, a process during which the genome duplicates without mitosis (Sabelli and Larkins, 2009).

Endoreduplication in cereal grains does not contribute to the seed size (Nowicka *et al.*, 2021a), but a study in *Arabidopsis* suggests that it might be linked with high metabolic activity of certain cells (Baroux *et al.*, 2004). Endoreduplicated nuclei allow the study of whether chromosome organization patterns are correlated with ploidy. In *Arabidopsis*, endoreduplicated nuclei from leaves showed loss of positional sister chromatid (SC) cohesion and had generally reduced heterochromatin compaction (Schubert *et al.*, 2006). In rice, endoreduplication forced Rabl chromosome organization in xylem vessel cells (Prieto *et al.*, 2004). Finally, conserved Rabl organization was

observed in endoreduplicated nuclei of bread wheat embryo and endosperm tissues (Wegel and Shaw, 2005).

Here, we studied chromosome organization in mitotically cycling and endoreduplicated nuclei isolated from embryo and endosperm tissues of developing barley grains. Our data suggest that while the chromosomes of embryo nuclei are organized mostly according to a Rabl pattern, the nuclei of endosperm tissues adopt a non-Rabl organization with increasing C-value and number of days after pollination. Collectively, our study provides comprehensive characteristics of embryo and endosperm nuclear processes during the key stages of barley grain development.

Materials and methods

Plant materials and growth conditions

The spring cultivar (cv.) Compana (PI 539111, NSGC of the USDA-ARS, USA) of cultivated barley (*Hordeum vulgare* subsp. *vulgare*) was used in this study. For germination, grains were evenly spread in a Petri dish on filter paper soaked with distilled water, stratified at 4 °C in the dark for 48 h, and germinated in the dark at 25 °C for 3 d. Sprouting seedlings were planted into 5 × 5 × 5 cm peat pots with a mixture of soil and sand (2:1, v/v) and grown in a phytochamber under a long-day regime (16 h daylight with an intensity of 200 μmol m⁻² s⁻¹ and temperature 20 °C; 8 h night with 16 °C; 60% humidity). After 2 weeks, seedlings were transferred into 15 × 15 × 15 cm pots and grown under the same conditions until flowering. The day of pollination was monitored using the morphology of the stigma and anthers according to the Waddington scale (W10) (Waddington *et al.*, 1983) as we established previously (Kovacic *et al.*, 2020; Nowicka *et al.*, 2021b). Seeds were collected from the center of the spike at 4, 8, 16, 24, and 32 days after pollination (DAP).

Isolation of nuclei and flow sorting

Nuclei were isolated from root apical meristem (RAM), embryo, and endosperm tissues. For isolation of root nuclei: 70 roots of seedlings at 2 d after germination were cut ~1 cm from the apex and collected in a drop of distilled water. Next, roots were drained on a cellulose tissue paper, rinsed in 10 mM Tris buffer pH 7.0, fixed with 2% (v/v) formaldehyde/Tris buffer for 20 min on ice, and washed three times for 5 min each with Tris buffer also on ice. Root apices were excised ~1 mm from the tip and homogenized in 500 μl of LB01 buffer (15 mM Tris-HCl pH 7.5, 2 mM NaEDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl, and 0.1% Triton X-100) for 13 s at 15 000 rpm using a Polytron PT1300D homogenizer (Kinematica AG).

For isolation of embryo nuclei, ~100 embryos were manually dissected from seeds at 8 DAP, using an SZX16 binocular microscope (Olympus). First, seeds were peeled (manual removal of hulls) and placed on a Petri dish in a drop of 1× phosphate-buffered saline (PBS) (Kovacic *et al.*, 2020). Dissected embryos were collected into a 1.5 ml Eppendorf tube containing 200 μl of 1× PBS and kept on ice until the sampling was finished. Tubes were low-speed centrifuged (1 min, 2000 g), PBS was removed, embryos were rinsed in Tris buffer, fixed with 2% (v/v) formaldehyde/Tris buffer for 20 min on ice, and washed three times for 5 min with Tris buffer also on ice. Embryos were homogenized with a pellet pestle in the same Eppendorf tubes with 500 μl of LB01 buffer. For 16 DAP and older seeds, at least 50 embryos were manually dissected using a binocular microscope. Embryos were collected into a small beaker containing PBS and kept on ice. Subsequently, the material was rinsed in Tris buffer, pre-fixed with 4% (v/v) formaldehyde/Tris buffer for 30 min by vacuum infiltration on

ice, followed by fixation with the same solution for 30–40 min without a vacuum also on ice, and washed in Tris buffer. Samples were chopped in 2 ml of LB01 buffer with a razor blade on a Petri dish.

For isolation of endosperm nuclei, ~80 of the whole peeled seeds from 4 and 8 DAP endosperm samples were gathered into a small beaker kept on ice. Seeds were rinsed in Tris buffer, pre-fixed with 4% (v/v) formaldehyde/Tris buffer for 20 min on ice, fixed without vacuum for 40 min, and washed. Samples were chopped with a razor blade in 2–3 ml of LB01 buffer on a Petri dish. For endosperm samples at ≥16 DAP, ~60 whole peeled seeds with embryos removed were collected into a small beaker, cut with the razor blade into 1 mm thick transversal slides, pre-fixed for 30 min, then fixed for 40 min. Samples were chopped with a razor blade in 3–4 ml of LB01 buffer on a Petri dish.

For flow sorting of nuclei, the crude homogenates of all samples were double-filtered first through a 50 µm and then a 20 µm pore size mesh. Nuclei suspensions were stained with 2 µg ml⁻¹ DAPI. Approximately 500 nuclei for each C-value were flow-sorted into a 2 µl drop of sorting buffer (100 mM Tris-HCl pH 7.5, 50 mM KCl, 2 mM MgCl₂, 0.05% Tween-20, and 5% sucrose) on poly-lysine (Menzel Gläser, J2800AMNZ) or superfrost Plus (Menzel Gläser, J1810AMNZ) microscopic slides using a FACSaria II SORP flow cytometer and sorter (BD Biosciences, Santa Clara, CA, USA). Slides were air-dried for 1 h at room temperature and stored at -20 °C until use.

Mitotic chromosome preparations

Metaphase chromosomes were prepared from synchronized root tips (Lysák *et al.*, 1999). Briefly, germinated seedlings were transferred into Hoagland's nutrient solution containing 2 mM hydroxyurea for 18 h. Then the roots were washed in distilled water and cultured in hydroxyurea-free Hoagland's solution for 5.5 h. To accumulate cells at metaphase, the roots were treated for 2 h with Hoagland's solution containing 2.5 µM amiprofos-methyl (A0185, Duchefa Biochemie). Subsequently, the root tips were fixed in ice-cold 90% acetic acid for 10 min followed by three washes in 70% ethanol and stored in 70% ethanol at -20 °C. Chromosomes were prepared using the drop technique (Danilova *et al.*, 2012). In brief, maceration of root tips was performed at 37 °C for 57 min using an enzyme mixture consisting of 4% (w/v) cellulase Onozuka R-10 (Yakult Pharmaceutical Industry, 150422-01) and 1% (w/v) pectolyase Y23 (Duchefa, 9033-35-6) in KCl buffer (75 mM KCl, 7.5 mM EDTA, pH 4). The quality of chromosome spreads was evaluated using a phase-contrast microscope (Primo Star, Zeiss), and the slides with at least five metaphases were used for fluorescence *in situ* hybridization (FISH). Before a hybridization step, the slides were pre-treated with pepsin (10 µg ml⁻¹ in 10 mM HCl) at 37 °C for 10 min, then rinsed in 2× SSC followed by RNase A treatment (described below).

Fluorescence *in situ* hybridization

Three combinations of the following probes were used in the double- or triple-color FISH experiments. For detecting barley centromeres, a synthetic 28-mer oligonucleotide (5'-AGGGAGA-3')₄ probe labeled at the 5' end with Cy3 or Cy5 (Eurofins) was used. The probe targets a centromeric retroelement-like element *CEREBA* conserved among cereal centromeres (Hudakova *et al.*, 2001). For detecting the telomeres, we used a synthetic 28-mer oligonucleotide probe (5'-CCCTAAA-3')₄ corresponding to the Arabidopsis-type telomeric repeat and labeled at the 5' end with Cy3 or Cy5 (Eurofins). The 5S *rDNA* probe was amplified from cv. Compana genomic DNA with the primers 5'-GGATGCGATCATACCAGCAC-3' and 5'-GACATGCAACTATCTATTTGT-3' using biotin-dUTP or digoxigenin-11-dUTP (both Roche, 11093070910 and 11093088910) during PCR. The 45S *rDNA* probe was labeled with biotin-dUTP or digoxigenin-11-dUTP from the pTa71 plasmid containing a 9.1 kb fragment of

rDNA sequence from bread wheat (Gerlach and Bedbrook, 1979) using nick translation kits (both Roche, 11745824910 and 11745816910) according to the manufacturer's instructions.

FISH was performed as described (Karafiátová *et al.*, 2013; Nowicka *et al.*, 2020), with the following modifications. Preparations were air-dried at room temperature, rinsed in 2× SSC, treated with RNase A (50 µg ml⁻¹ in 2× SSC; Thermo Fisher, EN0601) for 30 min at 37 °C, and washed with 2× SSC and 1× PBS. Subsequently, slides were post-fixed with 4% formaldehyde/1×PBS for 15 min and washed with 1× PBS. A hybridization mixture contained a cocktail of two or three probes, each with a final concentration of 400 ng µl⁻¹, and 1 µg of sheared salmon sperm DNA (Invitrogen, AM9680), 50% (v/v) deionized formamide, 10% (v/v) dextran sulfate, and 2× SSC. For biotin-dUTP- and digoxigenin-11-dUTP-labeled probes, the hybridization mixture was heated for 4 min at 95 °C, cooled on ice, and denatured again together with target DNA on slides for 4 min at 80 °C. For oligo-probes, the step of hybridization mixture pre-denaturation was skipped. Biotin-dUTP was detected either by (i) streptavidin-Cy3 (1:500, Molecular Probes, SA1010) or (ii) goat anti-avidin conjugated with biotin (1:100, Vector Laboratories, NC9256157) followed by avidin conjugated with Texas Red (1:1000, Vector Laboratories, NC9172942). Digoxigenin-dUTP was detected either with (i) an anti-digoxigenin antibody conjugated with fluorescein isothiocyanate (FITC, 1:200 Roche, 11207741910) or (ii) a mouse anti-digoxigenin antibody (1:250, Roche, 11333062910) followed by application of a goat anti-mouse antibody conjugated with Alexa Fluor 488 (1:200, Molecular Probes, A32723). The preparations were counterstained with DAPI in Vectashield (Vector Laboratories, H-1200-10).

ImmunoFISH

Slides were air-dried at room temperature, post-fixed with 4% formaldehyde/1× PBS for 15 min, and washed with 1× PBS. Immunostaining was carried out as described (Jasencakova *et al.*, 2000) with minor modifications. In brief, preparations were incubated with the rabbit anti-barley-αCENH3-specific primary antibody (1:200; Sanei *et al.*, 2011) at 4 °C overnight and the secondary goat anti-rabbit-Alexa Fluor 488 (1:300, Molecular Probes, A11008) at 37 °C for 90 min. Before FISH, slides were fixed in 3:1 ethanol/acetic acid for 10 min, followed by 10 min fixation with 3.7% formaldehyde/1× PBS. Slides were washed with 1× PBS. FISH steps were performed as described above, excluding pepsin and RNase A treatment.

Microscopy

The images were taken with an AxioImager Z2 upright microscope (Zeiss, Oberkochen, Germany) equipped with a pE-4000 LED illuminator light source (CoolLED, Andover, UK), motorized four-position excitation filter wheel, laser-free confocal spinning disk device (DSD2, Andor, Belfast, UK), and a ×100/1.4 NA Oil M27 Plan-Apochromat (Zeiss) objective. Image stacks of 40–80 slides depending on the C-value of the nucleus, on average, with a 0.2 µm z-step were acquired separately for each fluorochrome using the appropriate excitation [DAPI λ=390/40 nm, green fluorescent protein (GFP) λ=482/18 nm, red fluorescent protein (RFP) λ=561/14 nm, Cy5=640/14 nm] and emission (DAPI λ=452/45 nm, GFP λ=525/45 nm, RFP λ=609/54 nm, Cy5=676/29 nm) filters. For fluorescence detection, the 4.2 megapixel sCMOS camera (Zyla 4.2) was used and the iQ 3.6.1 acquisition software (both Andor) was used to drive the microscope.

Image analysis

The images were converted into .ims format with Imaris File Converter 9.2.1 (Bitplane, Zurich, Switzerland) and exported as maximum intensity projection (mip) tif files with Imaris 9.7 software (Bitplane). For

visualization of the surface and shape of the nuclei, the Imaris 9.7 function 'Surface' was used for rendering the DAPI-stained nucleus surface and to obtain 3D nucleus images. Then, functions 'Slide viewer' and 'Section view' were applied to visualize inside the nucleus. For determination of the nucleus area, perimeter, and circularity, the nucleus area (NA) and perimeter (NP) of the X–Y middle slide view tif images were measured in FIJI (ImageJ2; <https://imagej.net/software/fiji/>) calibrated with an internal size control. The nucleus circularity index (NCI) was calculated according to the formula: $NCI = 4\pi \times NA / (NP)^2$ (Ankam et al., 2018).

To construct the karyotype, chromosomes were paired based on the chromosomal position of *rDNAs* and *CEREBA*. The karyotype was prepared in Adobe Photoshop 6.0 (Adobe Systems Corporation, San Jose, CA, USA). Individual chromosomes were classified according to Fukui et al. (1994) and Kapusi et al. (2012). For FISH signal scoring, the number of FISH signals per nucleus was quantified in FIJI with the 'Multipoint' tool using mip tif images. Quantitative analysis of CENH3 co-localization with *CEREBA* was performed in FIJI calibrated with an internal size control using fluorescence intensity 'Plot Profile' for both correlated signals.

Total fluorescence intensity measurements of all CENH3 and *CEREBA* per nucleus were done in FIJI using mip tif images. For each nucleus, 2–10 regions of interest (ROIs) defined manually with a constant size of $3.5 \times 3.5 \mu\text{m}$ were evaluated. For green and red channels, the same ROIs were analyzed, and for each of them the fluorescence intensity ratio of CENH3/*CEREBA* was calculated. For the DAPI channel ($3.5 \times 3.5 \mu\text{m}$), ROIs located in the middle of the nucleus were evaluated.

For Rab1 configuration analysis, Imaris applications 'Surface' and 'Spot detection' were used for rendering the nucleus surface and modeling the centromere/telomere arrangements, respectively. The space in the nucleus occupied by centromeres and telomeres was measured using Imaris 'Measurement point' in polygon mode. The detailed Imaris-based image analysis workflow is described separately (Randall et al., 2022). The Imaris statistic output files reporting on the distance between centromeres and telomeres were exported for each nucleus separately.

Image data normalization and statistical analysis

Scored numbers of FISH signals were normalized to the number of signals per nucleus at the G_1 phase (Supplementary Fig. S1). Hence, for the number of *5S major rDNA* loci, scoring values were divided by two or by three for embryo and endosperm data, respectively. In the case of *5S minor* and *45S rDNA* loci, raw data were divided by four and six for embryo and endosperm, respectively. For *CEREBA*, data were normalized to 14 and 21 for embryo and endosperm, respectively. Distances between centromeres and telomeres were expressed as a ratio to the nucleus diameter (ND).

All scoring and measurement, raw and normalized data were tested for Gaussian distribution. To return to Gaussian distribution, data expressed as percentages were first arcsine transformed. Next, relevant comparisons were carried out by two-way ANOVA (factor 1=C-value, factor 2=DAP) and post-hoc Duncan's multiple ranges ($P \leq 0.05$) test. To evaluate the statistical differences between embryo and RAM samples, one-way ANOVA (factor 1 tissue) was applied. Pearson's correlation coefficient analysis was used to visualize relationships among the measured and evaluated parameters (NCI, NA, NP, and co-localization between CENH3/*CEREBA*). Statistical analyses were performed in Statistica v. 12 (Statsoft Inc.) or Minitab v. 18 (Minitab). Boxplots were drawn using the ggplot GUI online tool (<https://shiny.gmw.rug.nl/ggplotgui/>).

Results

Endoreduplication affects the morphology of barley embryo and endosperm nuclei

We isolated G_1 (2C/3C), G_2 (4C/6C), and endoreduplicated (8C/ ≥ 12 C) nuclei from barley embryo and endosperm tissues

(embryo/endosperm C- values, respectively) at 4, 8, 16, 24, and 32 DAP as described (Nowicka et al., 2021a), and analyzed their morphology. Nuclei from highly dividing RAM tissues were used as somatic control. Based on our initial assessment, we noted differences in nuclear shape and therefore calculated the nucleus circularity index (NCI) (Ankam et al., 2018) using nucleus area (NA) and nucleus perimeter (NP) values (Materials and methods; Fig. 1A, B; Supplementary Figs S2, S3).

The 2C and 4C embryo nuclei had a nearly ideal circular shape ($NCI \geq 0.91$) during the whole of seed development, but the NCI of endoreduplicated (8C) nuclei from 24 and 32 DAP was significantly reduced to ~ 0.75 (two-way ANOVA; Fig. 1B; Supplementary Fig. S2D). Hence, the circularity of the embryo nuclei depended on the degree of endoreduplication but not the number of DAP. In contrast, NCI of endosperm nuclei was influenced not only by the C-value but also by DAP. The NCI of 3C and 6C nuclei was ~ 0.89 from 4 to 16 DAP, then reached its maximum of ~ 0.95 at 24 DAP and decreased to 0.82–0.88 at 32 DAP. Endoreduplicated endosperm nuclei exhibited an ellipsoid shape with NCIs between 0.75 (12C at 8 DAP) and 0.67 (24C at 24 DAP), and the ellipticity increased during seed maturation and desiccation (Fig. 1A, B; Supplementary Fig. S2D). Interestingly, rendering of the surface of endoreduplicated endosperm nuclei revealed grooves of variable dimensions that were not observed in the nuclei with lower C-values (Fig. 1C; Supplementary Fig. S4).

Loss of sister chromatid cohesion at 5S rDNA loci of endoreduplicated nuclei

The altered morphology of barley endoreduplicated seed nuclei stimulated us to also explore the chromosome organization. As detection of single-copy sequences in barley interphase nuclei is not possible using current cytology tools, we focused on the arrangement of major tandem repetitive regions *5S rDNA*, *45S rDNA*, and *CEREBA* centromeric repeats using FISH (Fig. 2A, B).

The genome of cv. Compana contains a cytologically distinguishable single large *5S rDNA* cluster at the bottom arm of chromosome 2H (hereafter *5S major*) and two smaller clusters at the bottom arms of chromosomes 3H and 7H (hereafter *5S minor*; Fig. 2A–C). In G_1 (2C) and G_2 (4C) nuclei of the embryo samples up to 16 DAP and root samples, we observed two separate *5S major rDNA* signals, which suggests SC cohesion during the G_2 phase and separation of the two homologous chromosomes in both G_1 and G_2 . However, in embryo nuclei from older (24 and 32 DAP) seeds, 4C nuclei contained mostly three to four *5S major* FISH signals and 8C nuclei contained five to seven such signals, indicating a reduced SC cohesion at *5S major* (Fig. 2D; Supplementary Fig. S5A). G_1 (3C) endosperm nuclei showed mostly three *5S rDNA major* FISH signals, but G_2 (6C) and once-endoreduplicated (12C) nuclei displayed SC separation progressing in a DAP-dependent manner. Finally, the majority of twice-endoreduplicated

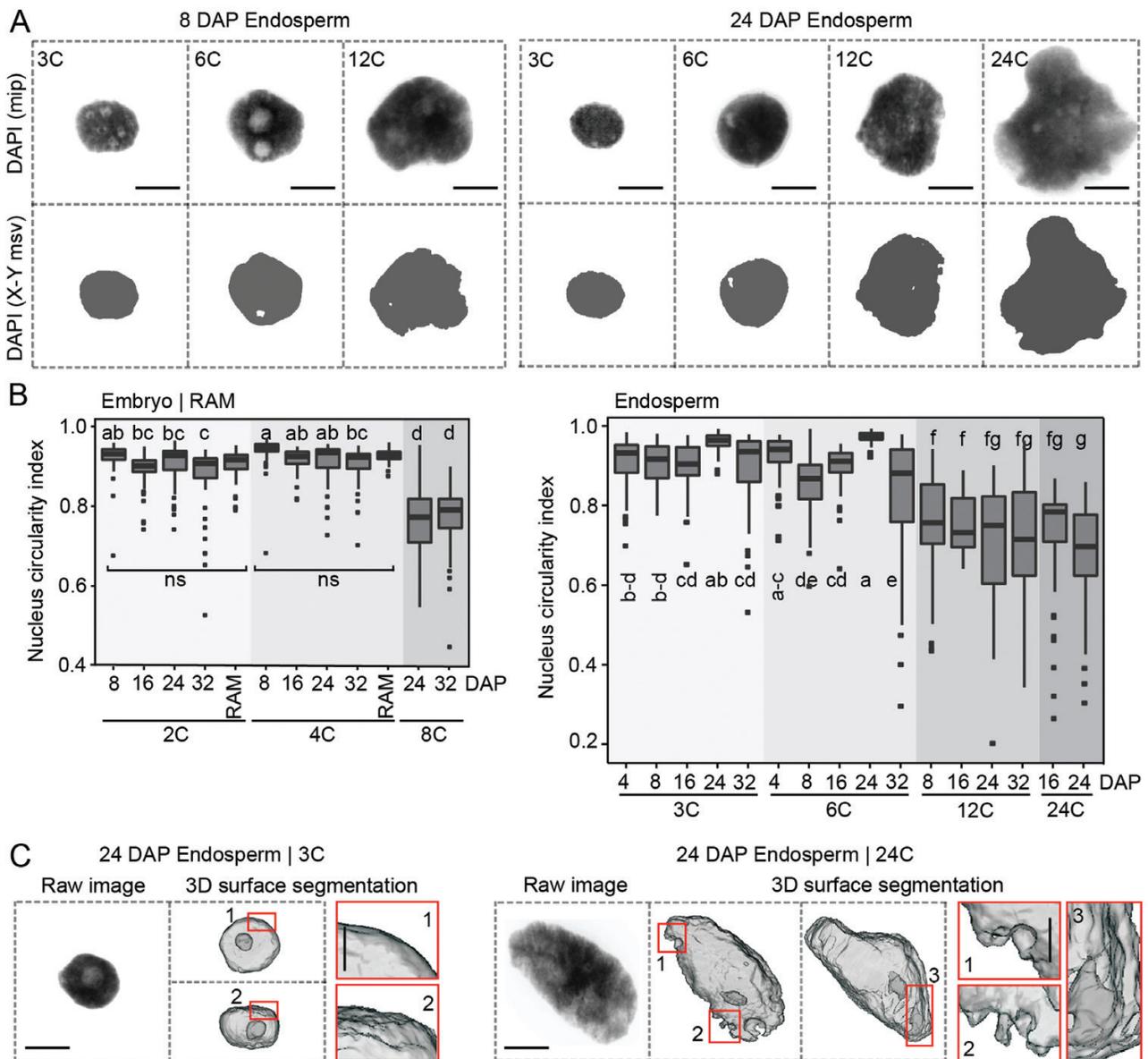


Fig. 1. Endoreduplicated barley grain nuclei have altered shape. (A) Representative DAPI-stained endosperm nuclei of different C-values collected at 8 and 24 days after pollination (DAP). The upper panels show 3D maximum intensity projections (mip) and the lower panels their 2D X-Y middle slide view (msv). Scale bars=10 µm. (B) Boxplots showing the nucleus circularity index (NCI) for nuclei of different tissues, C-values, and DAP. Root apical meristem (RAM) nuclei were used as the vegetative tissue control. NCI was calculated using the following formula $NCI=4\pi \times NA/(NP)^2$ (Ankam et al., 2018). Original data for NA and NP are depicted in [Supplementary Fig. S2](#). The lower and upper hinges of the boxplots correspond to the first and third quartiles of the data, and the black lines within the boxes mark the median. Whiskers mark 10% and 90% intervals. A total of 75 nuclei were measured in three microscopic slides. Black squares represent outliers. Different letters indicate significant differences ($P \leq 0.05$, two-way ANOVA, factor 1=C-value, factor 2=DAP, followed by Duncan post-hoc test). The summary of ANOVA is presented in [Supplementary Fig. S2D](#). Statistical significance between embryo and RAM samples was evaluated with one-way ANOVA, ns, not significant ($P_{2C}=0.453$, $P_{4C}=0.101$). (C) Example images of 24 DAP 3C and 24C DAPI-stained endosperm nuclei presented in 3D mip (left panel) and their surface reconstructions using Imaris software (right panels). Insets display in more detail the absence (3C) and presence (24C) of nuclear grooves. Scale bars=10 µm (main) and 2 µm (insets). Additional images are presented in [Supplementary Fig. S4](#).

endosperm (24C) nuclei had 21 FISH signals at 24 DAP, corresponding to full SC separation at *5S rDNA major* ([Fig. 2C, D](#); [Supplementary Fig. S5](#)). Statistical analysis by ANOVA showed that the *5S rDNA major* SC separation in embryo and en-

dosperm significantly increased with rising C-values, developmental progression (DAP), and as a result of an interaction between both factors ([Supplementary Fig. S5](#)). The same trend was observed for *5S rDNA minor loci* ([Supplementary Fig. S6](#)).

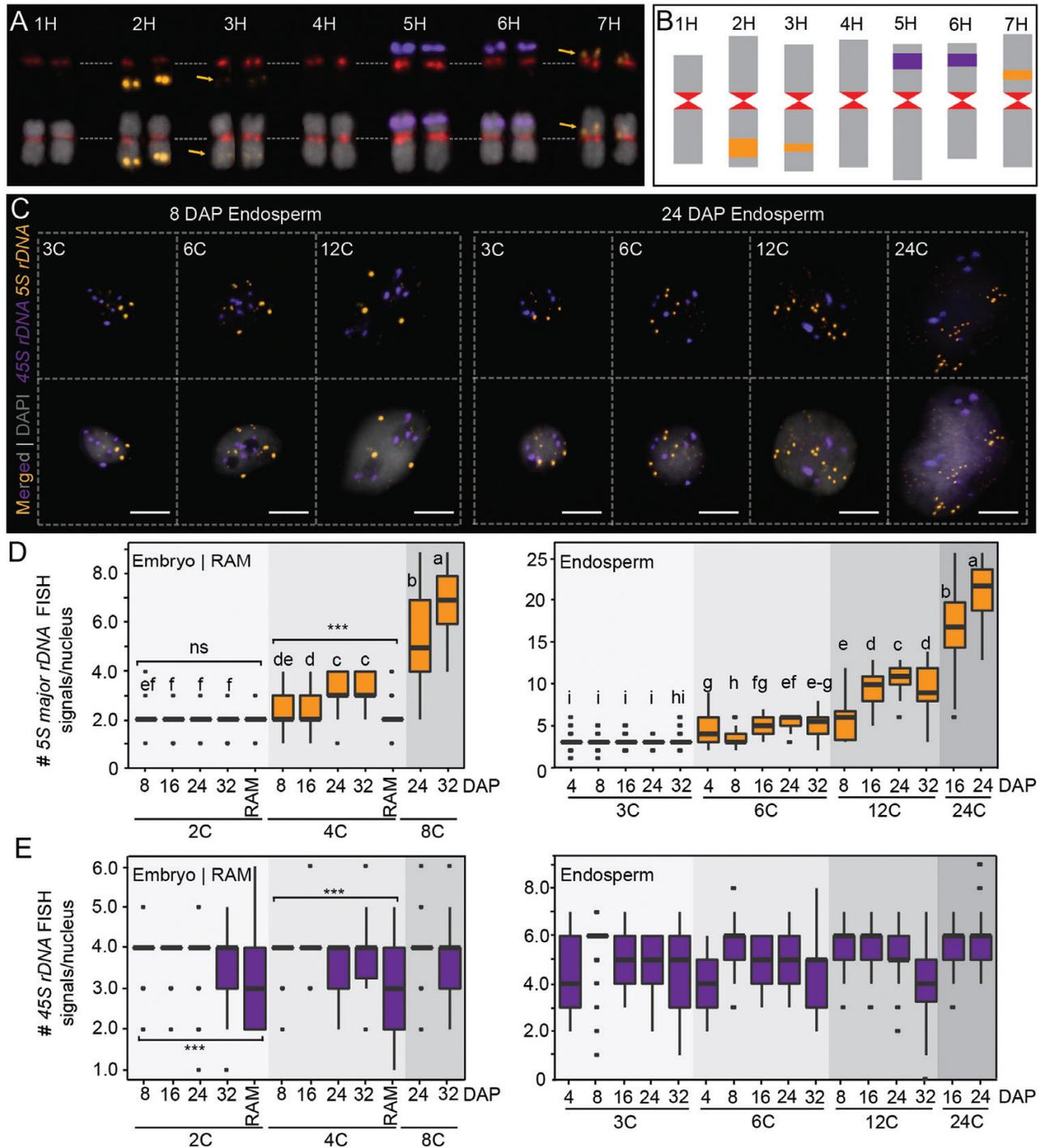


Fig. 2. Sister chromatid cohesion at *5S* and *45S* *rDNA* loci in barley seed nuclei. (A) Karyotype of cv. *Compana* showing localization of *5S rDNA*, *45S rDNA*, and *CEREBAs* centromeric repeats on metaphase chromosomes by FISH. DNA was counterstained with DAPI. Arrows indicate *5S minor rDNA* loci. Scale bar=10 μ m. (B) Ideogram of cv. *Compana* (based on A). (C) Representative endosperm nuclei with different C-values collected at 8 and 24 DAP after FISH with *5S* (orange) and *45S* (violet) *rDNA* probes. The larger and brighter *5S rDNA* signals correspond to the *5S major* loci. DNA was stained with DAPI (gray). Scale bars=10 μ m. (D, E) Boxplots showing the number of (D) *5S major* and (E) *45S rDNA* FISH signals per nucleus for different tissues, C-values, and DAP. Root apical meristem (RAM) nuclei were used as the vegetative tissue control. The lower and upper hinges of the boxplots correspond to the first and third quartiles of the data, and the black lines within the boxes mark the median. Whiskers mark 10% and 90% intervals. At least 75 nuclei from three microscopic slides were scored. Black squares represent outliers beyond the whiskers. Different letters indicate significant differences ($P \leq 0.05$, two-way ANOVA, factor 1=C-value, factor 2=DAP, followed by Duncan test). No significant differences were found for the *45S rDNA* ($P > 0.05$). The summary of ANOVA is presented in [Supplementary Figs S5B and S7A](#). Statistical significance between the embryo and RAM samples was evaluated with one-way ANOVA, *** significant at $P \leq 0.001$, ns, not significant ($5S P_{2C} = 0.771$). Normalized data ([Supplementary Fig. S1](#)) for *5S major* and *45S rDNA* signals are provided in [Supplementary Figs S5C and S7B](#), respectively.

Analogous analysis of the *45S rDNA*, located on barley chromosomes 5H and 6H (Kapusi *et al.*, 2012), revealed a lower than expected number of FISH signals. This suggests a persistent SC cohesion and tendency toward *45S rDNA* clustering (Fig. 2C, E; Supplementary Figs S5A, S7). Neither C-value nor DAP affected this pattern (two-way ANOVA, $P > 0.05$, no significant differences for single factors and their interaction, Supplementary Fig. S7). Hence, the organization of *45S rDNA* loci remained relatively intact, suggesting a locus-specific control of SC alignment in endoreduplicated barley seed nuclei.

Decondensation of *CEREBA* and reduction of *CENH3* in endoreduplicated nuclei

After FISH with *CEREBA* repeats, we observed on average 12 signals at 8 DAP and 11 signals at 32 DAP in 2C and 4C embryo nuclei (Fig. 3; Supplementary Fig. S8A). The average number of signals increased significantly to ~14–17 after endoreduplication (8C). In 3C and 6C endosperm nuclei, the average number of *CEREBA* foci ranged from nine to 17 at different DAPs. In 12C and 24C nuclei, the *CEREBA* signals appeared less compact, often splitting into several smaller foci,

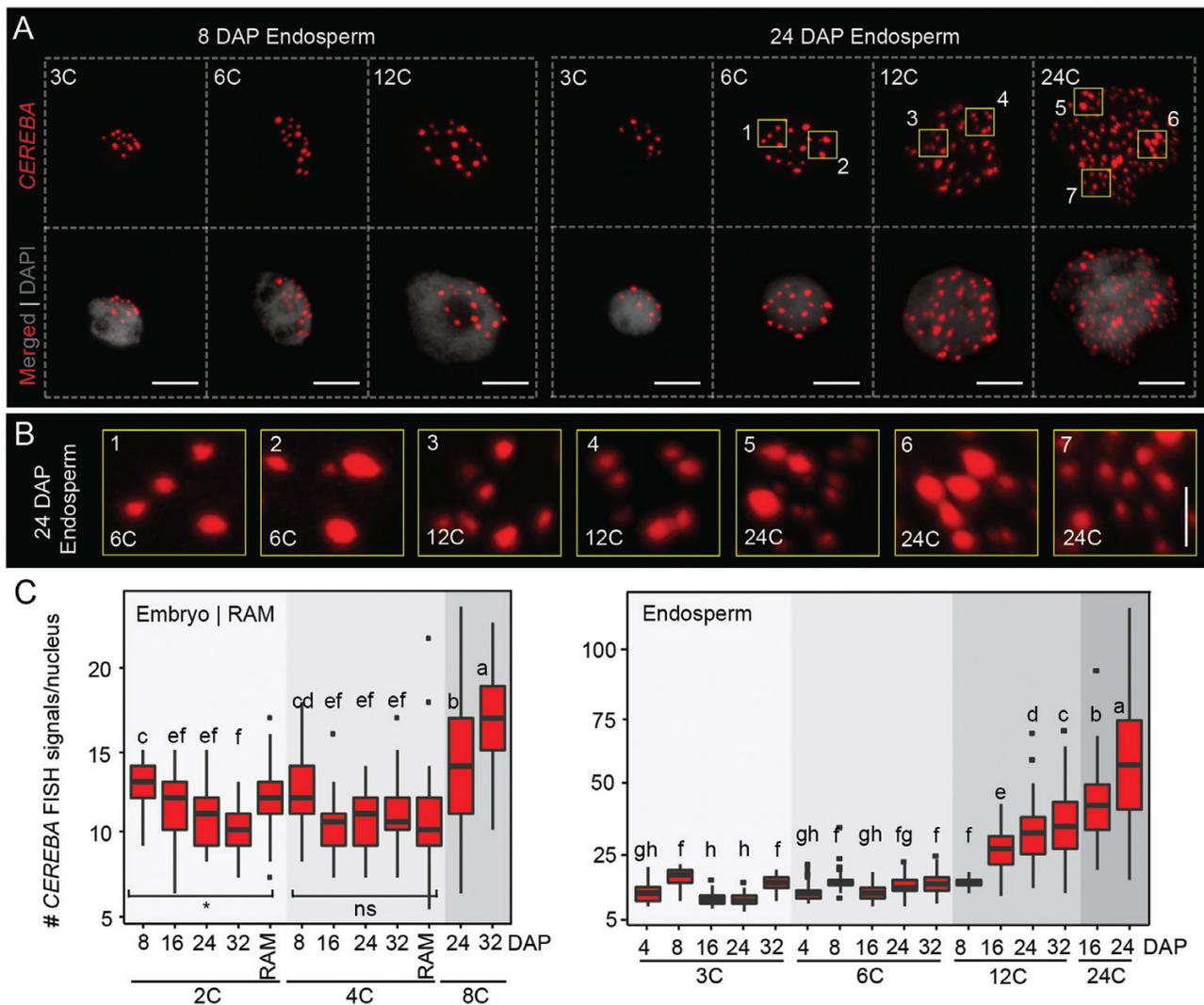


Fig. 3. *CEREBA* organization in barley seed nuclei. (A) Representative endosperm nuclei of different C-values collected at 8 and 24 DAP after FISH with *CEREBA* centromeric repeat (red). DNA was stained with DAPI (gray). Insets are enlarged in (B). Scale bars=10 μm . (B) Insets (1–7) marked in (A) show the variable size of *CEREBA* FISH signals in the 6C–24C endosperm nuclei. Scale bar=2 μm . (C) Boxplots showing the number of *CEREBA*-FISH signals per nucleus for different tissues, C-values, and DAP. Root apical meristem (RAM) was used as the vegetative tissue control. The lower and upper hinges of the boxplots correspond to the first and third quartiles of the data, and the black lines within the boxes mark the median. Whiskers mark 10% and 90% intervals. At least 75 nuclei from three microscopic slides were evaluated. Black squares represent outliers beyond the whiskers. Different letters indicate significant differences ($P \leq 0.05$, two-way ANOVA, factor 1=C-value, factor 2=DAP, followed by Duncan test). The summary of ANOVA is presented in Supplementary Fig. S8B. Statistical significance between embryo and RAM samples was evaluated with one-way ANOVA, * significant at $P \leq 0.05$, ns, not significant ($P_{4c} = 0.147$). Normalized data (Supplementary Fig. S1) are provided in Supplementary Fig. S8C.

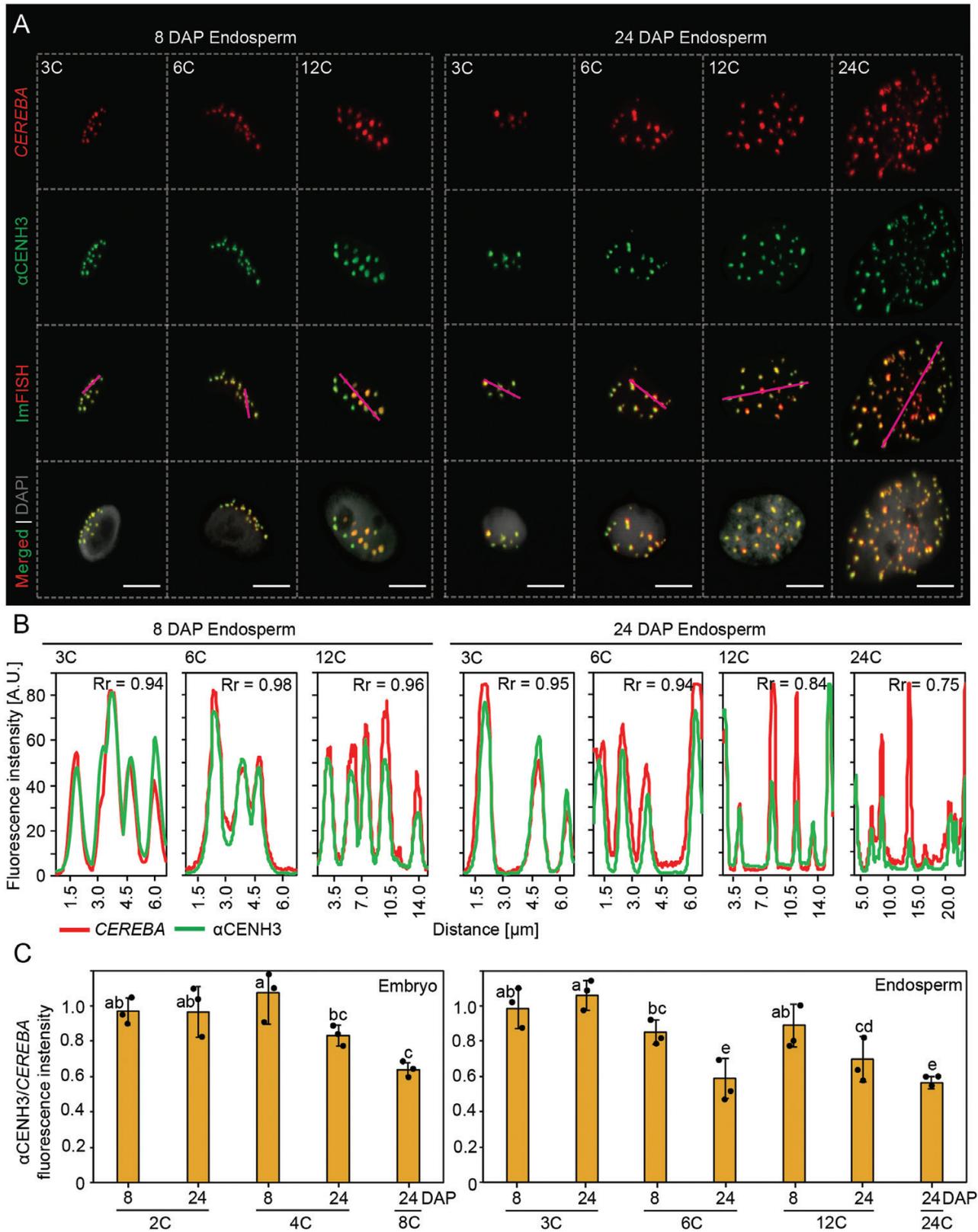


Fig. 4. Loss of HvCENH3 signals in endoreduplicated nuclei of seed tissues. (A) Representative endosperm nuclei of different C-values collected at 8 and 24 DAP after immunostaining followed by FISH (ImFISH) for barley α CENH3 (green) and CEREBAs (red). DNA was stained with DAPI (gray). Scale bars=10 μ m. (B) Fluorescence intensity plot profiles (y-axis; arbitrary units, A.U.) showing the quantified co-localization of HvCENH3 and CEREBAs

signals. Intersects used for quantification are highlighted by a pink line in the ImFISH images in (A). R_r displays Pearson's co-localization coefficient. Data for embryos are presented in [Supplementary Fig. S9](#). (C) α CENH3/*CEREBA* fluorescence intensity signal ratio (based on [Supplementary Fig. S10A](#)) measured for the same-size squared regions of interest (ROIs). Values are means (\pm SD) from three biological replicates (microscopic slides) marked as black spots, each with at least 100 measured ROIs. Different letters indicate significant differences ($P \leq 0.05$, two-way ANOVA, factor 1=C-value, factor 2=DAP, followed by Duncan test). The summary of ANOVA is shown in [Supplementary Fig. S10B](#). DAPI fluorescence measurements are presented in [Supplementary Fig. S10C](#).

and this trend was more pronounced with increasing DAP ([Fig. 3B, C](#)). Two-factor ANOVA revealed an additive effect of DAP in combination with C-value on the number of *CEREBA* foci in both embryo and endosperm nuclei ([Supplementary Fig. S8B, C](#)). The high number of *CEREBA* signals most probably indicates relaxation of centromeric repeats or a larger distance between individual sister chromatids at the centromeric region.

To understand whether a reduced compaction of the centromeric regions would prevent centromere maturation, we set out to measure the levels of CENH3 in different types of nuclei. For this, we performed *CEREBA* and barley α CENH3 immunofISH in 8 and 24 DAP embryo and endosperm nuclei, measured the fluorescence signal intensities over an intersecting line, and calculated their Pearson correlation ([Fig. 4A, B](#); [Supplementary Fig. S9](#)). This indicated that all *CEREBA* foci also contain a CENH3 signal, but the latter appeared to be weaker with increasing C-value in 24 DAP endosperm nuclei (Pearson correlation 0.95 in 3C and 0.75 in 24C). To quantify this interesting observation, we measured signal intensities within ROIs of fixed size and calculated the α CENH3/*CEREBA* ratio ([Fig. 4C](#); [Supplementary Fig. S10](#)). Since it was previously shown that the α CENH3 immunosignal reflects the amount of CENH3 ([Lermontova et al., 2006](#)), we measured both FISH and immunosignals in ROIs and calculated the α CENH3/*CEREBA* ratio. In both embryo and endosperm tissues, there was a significant reduction of α CENH3 relative to *CEREBA* repeats in endoreduplicated nuclei at 24 DAP ([Fig. 4C](#)). Interestingly, this was an effect not only of the C-value but also of DAP as, for example, 6C and 12C nuclei had significantly more α CENH3 at 8 DAP than at 24 DAP ([Fig. 4C](#); [Supplementary Fig. S10C](#)). Thus, the decondensation of centromeres occurring in endosperm nuclei correlates with a lesser loading of α CENH3.

Loss of Rabl chromosome configuration in seed endoreduplicated nuclei

In the light of the massive changes in chromosome organization, we asked whether barley seed nuclei retain a Rabl configuration. For this, we imaged the 3D distribution of FISH signals targeting the *CEREBA* and telomeric repeats in 8 and 24 DAP embryo and endosperm nuclei ([Figs 5, 6](#)). Besides nuclei with a typical Rabl configuration, we observed several types of nuclei with dispersed and non-polar centromeric and telomeric signals. To quantify the degree of signal dispersion versus clustering, we measured the shortest distance of each centromere signal to the next telomere signal and expressed

it relative to the diameter of the nucleus ([Fig 5A, B](#); image processing workflow as described in [Randall et al., 2022](#)). In the 2C and 4C embryo nuclei with a typical Rabl organization, the average, relative distance between centromere and telomere clusters was \sim 30–40% of ND. In 8C nuclei with more dispersed signals, the relative distance was only \sim 12% of ND. Assessing the distance distribution among all samples, ANOVA revealed that the C-value (but not DAP) affected the relative positioning of centromeres and telomeres in embryo nuclei ([Supplementary Fig. S11A](#)). Similar patterns were found in endosperm nuclei, but with an effect of C-value, DAP, and a combination thereof ([Fig. 5B](#); [Supplementary Fig. S11A](#)).

In addition, we noted that centromeric and telomeric signals are occasionally located away from the nuclear periphery, which might be another indication of altered chromosome organization. To quantify this reorganization, we measured the shortest distance of centromeric and telomeric FISH signals to the nuclear periphery, defined by the boundary of DAPI staining ([Randall et al., 2022](#)). We confirmed that centromeres and telomeres became more dispersed towards the interior of the nucleus in endoreduplicated embryo and endosperm nuclei, with a gradual relocation depending on the C-value and DAP ([Fig. 5A, C](#); [Supplementary Fig. S11B](#)). Furthermore, to quantify the dispersion of telomeres and centromeres, we calculated the volume occupied by connected centromere signals and the same for telomeres. We expressed these values relative to the volume of the nucleus to provide an estimate of the spatial dispersion ([Fig. 5A, D](#)). In the 2C and 4C embryo nuclei, these domains occupied 18–26% of the nuclear volume but this increased to 44–45% in 8C nuclei. Statistical analysis showed that the C-value (for centromeres and telomeres) and DAP (only for telomeres) influenced the expansion of the signals ([Supplementary Fig. S11C](#)). In endosperm, centromeres and telomeres covered 27–35% of the nuclear volume of 3C and 6C nuclei at 8 DAP. With increasing C-value and DAP, they dispersed over the nuclear volume even more and reached 68% and 70% of nuclear space, respectively, with 27% overlap in 24C endosperm nuclei at 24 DAP. In endosperm nuclei, the dispersion increased with the C-value, DAP, and their combination ([Supplementary Fig. S11C](#)).

Based on the above observations, we defined three arbitrary categories of nuclear organization: (i) Rabl; (ii) intermediate; and (iii) non-Rabl, with a median shortest centromere to telomere distance of 42, 23, and 7%, respectively ([Fig. 6A, B](#)). We quantified frequencies of these categories over the experimental points ([Fig. 6C](#)). The Rabl configuration was present in \geq 85% ($n=21$ of 25) of 2C and 4C embryo nuclei and in 77.3%

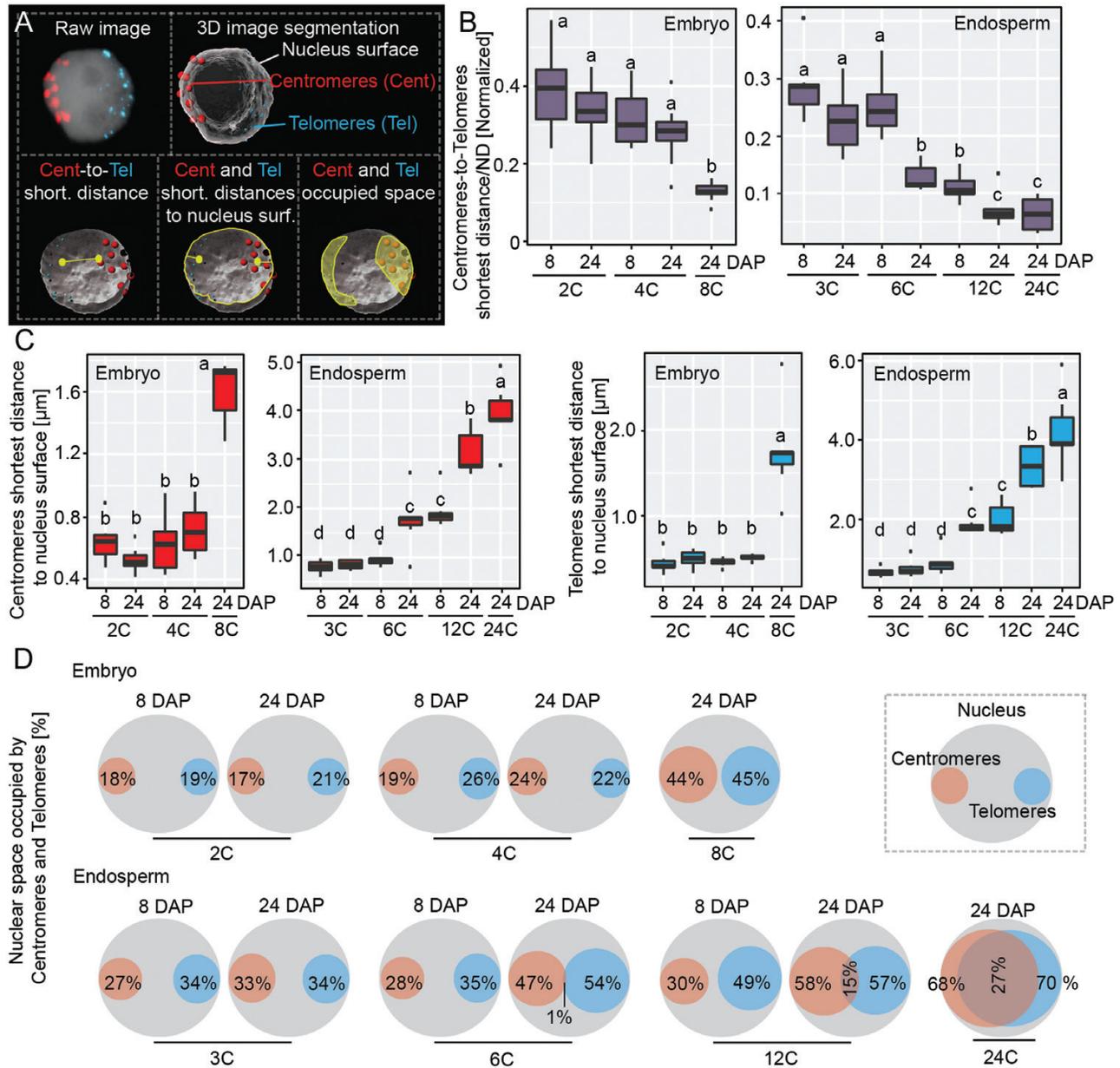


Fig. 5. Endoreduplication disrupts the Rab1 chromosome organization in barley nuclei. (A) Schematic overview of the processing of the raw images and quantified parameters applied for centromere and telomere positioning in the interphase nucleus. (B) Boxplots showing the shortest distance between centromeres and telomeres normalized to nucleus diameter (ND). The lower and upper hinges of the boxplots correspond to the first and third quartiles of the data, and the black lines within the boxes mark the median. Whiskers mark 10% and 90% intervals. Black squares represent outliers beyond the whiskers. The measurements were performed in Imaris software after FISH signal segmentation and nucleus surface rendering. Ten randomly selected nuclei of each C-value/time point were used for the analysis. Different letters indicate significant differences ($P \leq 0.05$, two-way ANOVA, factor 1=C-value, factor 2=DAP, followed by Duncan test). The summary of ANOVA is presented in [Supplementary Fig. S11A](#). (C) Boxplots showing the shortest distance of centromeres and telomeres to the nucleus surface. Data acquisition, plot organization, and statistics were performed as described in (B). The summary of ANOVA is presented in [Supplementary Fig. S11B](#). (D) Venn diagrams show the percentage of nuclear space occupied by centromeres and telomeres. The measurements were performed in Imaris software after FISH signal segmentation, nucleus surface rendering, and manual measuring of the nucleus territories occupied by centromeres and telomeres. Ten randomly selected nuclei of each C-value/time point were used for the analysis. The summary of ANOVA is presented in [Supplementary Fig. S11C](#).

($n=19$ of 25) of 8C nuclei. The Rab1-type nuclei were substituted by the intermediate type and the proportion of the non-Rab1 type remained very low (20%; $n=5$ of 25). In endosperm,

the Rab1 configuration appeared in the majority (64%; $n=16$ of 25) of 3C and 6C nuclei at 8 DAP. With increasing C-value and DAP, the proportion of nuclei with an intermediate and

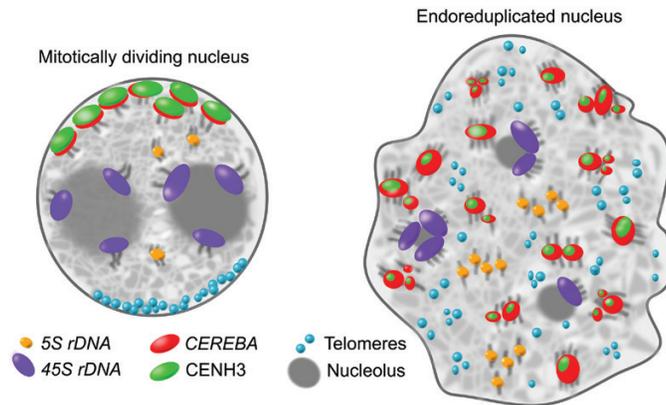


Fig. 7. Graphical summary of the major findings. Dividing nuclei have a round shape; adherent sister chromatids keep an equal amount of CENH3 histone and organize chromosomes according to a Rabl pattern. Endoreduplication alters the nuclear shape and causes positional loss of sister chromatid cohesion and loss of histone CENH3. In addition, they show a non-Rabl chromosome organization.

from developing barley embryo and endosperm tissues (Fig. 7). Our study shows that the tissue type, level of endoreduplication, and the age after pollination are the major determinants of the observed differences.

Chromosome organization has been explored in actively dividing meristematic and somatic tissues of barley, which contain mostly spherical nuclei with a smooth surface (Němečková et al., 2020). These nuclei are from cells that are either mitotically cycling or resting in the G_0/G_1 phase (Jasenčakova et al., 2000). Surprisingly, endoreduplicated endosperm nuclei adopted a very irregular shape, with channels lacking DNA staining, suggesting invagination of the nuclear membrane (Fig. 1). Similar shapes have been reported for endoreduplicated nuclei of several distantly related plants including *Allium cepa*, *Narcissus*, *Pisum sativum*, or *Solanum lycopersicum* (Collings et al., 2000, and references therein; Bourdon et al., 2011, 2012). This suggests that the effect of endoreduplication on the shape of the nucleus and particularly the regularity of its boundary is potentially widespread, consistent with the proposal that complex surface structures may be typical for nuclei of high ploidies (Pirrello et al., 2014). It is assumed that the grooves and invaginations may keep the necessary nucleus to cytoplasm surface ratio (Bourdon et al., 2012). Our work shows that invaginations are not a pure effect of endoreduplication in barley and that the seed developmental stage, tightly linked to its physiological state, plays a role. Some nuclear surface irregularities could be a result of metabolic activity in embryo and endosperm cells, possibly due to filling of cells with active and/or storage compounds or due to the altered cytoskeleton impacting the integrity of the nuclear envelope. This phenotype could also be linked to programmed cell death that is typical for the endosperm of most cereals (Young et al., 2000).

The hallmark of the Rabl configuration is the centromere and telomere clustering at opposite nuclear poles. However,

>130 years after its discovery (Rabl, 1885), the principles of this organization remain a matter of debate (Santos and Shaw, 2004). In some plants, the Rabl configuration was long thought to be the only type of genome organization. However, an increasing number of studies suggest tissue-specific variation in chromosomal organization (Fujimoto et al., 2005a; Idziak et al., 2015; Němečková et al., 2020; Shan et al., 2021). Some cells lose the Rabl pattern soon after entering the interphase, whereas others retain the organization throughout the interphase and until the next mitosis (Cowan et al., 2001). Nuclear genome size and chromosome length were postulated as two possible factors conferring the Rabl configuration (Santos and Shaw, 2004). While this holds for many species, some plants with giant genomes lack Rabl organization (Fujimoto et al., 2005a). Also, there are striking differences in genome organization between some closely related species. A well-described example are *Brachypodium* species, where *B. distachyon* shows a Rabl configuration in root nuclei while its relative *B. stacei*, with a similar genome size but twice as many chromosomes, does not (Idziak et al., 2015). This suggests that genome size alone cannot serve as a universal rule defining the Rabl organization. Recently, it was proposed that the Condensin II complex plays a major role in 3D interphase genome organization, and that an incomplete set of its subunits favors a Rabl-like pattern across the tree of life (Hoencamp et al., 2021). Applicability of this classification for the organization of plant chromosomes still requires investigations because all plants sequenced to date contain a full set of Condensin II subunits, in spite of their diverse chromosome organization (Fujimoto et al., 2005b; Schubert, 2009).

So far, barley has been considered as a species with a strict Rabl chromosome organization. We showed that there is variability in the chromosome configuration in barley seed tissues that is affected by the tissue type, seed developmental stage (days after pollination), and strongly by endoreduplication (C-value) and combination of the latter two factors. In contrast, Rabl organization was maintained in the embryo and endosperm nuclei of bread wheat (Wegel and Shaw, 2005), which could be due to analysis of younger tissues that contained only a small portion of endoreduplicated nuclei or, less likely, due to genuine species-specific differences. Importantly, correlation between loss of Rabl organization and degree of endoreduplication favors models suggesting that Rabl configuration is established and reinforced during mitotic cell divisions (Santos and Shaw, 2004). Our data show that the amount of nuclei with Rabl decreases not only with the number of DNA replications (that are not followed by mitosis), but also with the time since the last replication (Fig. 6). However, which molecular factors ensure relatively stable clustering of centromeres and telomeres in between divisions remains currently unknown.

The other observed changes in chromosome organization add to the little-explored organization of endoreduplicated nuclei of cereal seeds (Wegel and Shaw, 2005; Wegel et al., 2005; Bauer and Birchler, 2006). Here, repositioning of centromeres and telomeres from the nuclear envelope more into the nuclear space

may contribute to loss of Rabl configuration (Fig. 5) (Santos and Shaw, 2004). Another observed alteration in chromosome organization was related to absence of SC cohesion at *5S rDNA* loci (Fig. 2). Although we cannot draw any conclusions about the organization of singly-copy sequences, our data suggest that SCs are absent at least in some parts of the endoreduplicated barley chromosomes. This is reminiscent of the loss of SC cohesion along chromosome arms in Arabidopsis nuclei with a C-value of 4C or more (Schubert *et al.*, 2006). At centromeric regions, we found an increasing number of *CEREBA* foci in endoreduplicated nuclei (Fig. 3), which is similar to Arabidopsis (Schubert *et al.*, 2006; Baroux *et al.*, 2017). This suggests a relaxed control of heterochromatin compaction at centromeres upon endoreduplication in barley, which is in contrast to the situation in maize (Bauer *et al.*, 2006). We also observed reduction in centromeric histone CENH3 in endoreduplicated nuclei (Fig. 4). This is to be expected because CENH3 loading occurs in G₂ phase, that is skipped in the endoreduplication cycle (Lermontova *et al.*, 2007). Furthermore, data from Arabidopsis show that CENH3 is not produced during endoreduplicative S-phase (Lermontova *et al.*, 2011). Interestingly, we found a significant replication-independent loss of CENH3 in nuclei of the same C-value in later versus earlier seed developmental stages.

What the significance of the manifold changes in endoreduplicated barley nuclei is remains currently unknown. Speculatively, it could be linked with transcriptional reprogramming and a boost in synthesis of specific storage compounds in endosperm (Sabelli and Larkins, 2009). Furthermore, it could be related to a loss of mitotic activity and onset of the cellular trajectory towards programmed cell death that occurs in large parts of cereal endosperm (Nowicka *et al.*, 2021a).

In conclusion, our study highlights previously underappreciated dynamics in chromosome organization of barley embryo and endosperm nuclei upon endoreduplication. The most notable change is the progressive loss of polar chromosome organization and the disruption of centromere and telomere clusters. This shows that the Rabl chromosome arrangement is not a general rule for barley, and that mitosis may function as a mechanism reinforcing this organization. In general, these data help in understanding the principles and dynamics of genome organization during the course of plant development.

Supplementary data

The following supplementary data are available at [JXB online](#).

Fig. S1. Schematic drawing of G₁ diploid and triploid cells.

Fig. S2. Geometrical characteristics of flow-sorted embryo and endosperm nuclei.

Fig. S3. Pearson correlation coefficient analysis between nuclear geometry parameters.

Fig. S4. The complex shape of endoreduplicated nuclei.

Fig. S5. Sister chromatid organization for the *5S major rDNA* locus in barley seed nuclei.

Fig. S6. Sister chromatid organization for the *5S minor rDNA* loci in barley seed nuclei.

Fig. S7. Sister chromatid organization for the *45S rDNA* loci in barley seed nuclei.

Fig. S8. *CEREBA* organization in barley seed nuclei of different C-values and DAP.

Fig. S9. Co-localization of CENH3 and *CEREBA* signals in barley embryo nuclei.

Fig. S10. Fluorescence measurements of α CENH3, *CEREBA*, and DAPI signals.

Fig. S11. Statistical analysis supporting the Rabl chromosome configuration study.

Acknowledgements

We thank, Dr A. Doležalová for sharing the immunoFISH protocol and anti- α CENH3 antibody, E. Jahnová for technical assistance, and Z. Bursová for plant care. The authors thank COST Action no. CA 16212 'INDEPTH' for training in Imaris software.

Author contributions

AP and AN: design; AN, LF, MK, and JZ: performing experiments; MS: flow sorting of the nuclei; AN: data analysis and figure preparation; CB: providing expertise in microscopic image processing analysis by Imaris; AP and AN: writing, with contributions from all authors. All authors approved the final version of this article.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this work.

Funding

This research was funded by the Czech Science Foundation grants 18-12197S and 21-02929S (to AP), Purkyně Fellowship from the Czech Academy of Sciences to AP, and the European Regional Development Fund project 'Plants as a tool for sustainable global development' (no. CZ. 02.1.01/0.0/0.0/16_019/0000827).

Data availability

All data supporting the findings of this study are available within the paper and within its supplementary data published online.

References

- Ankam S, Teo BKK, Pohan G, Ho SWL, Lim CK, Yim EKF. 2018. Temporal changes in nucleus morphology, Lamin A/C and histone methylation during nanotopography-induced neuronal differentiation of stem cells. *Frontiers in Bioengineering and Biotechnology* **6**, 69.
- Baroux C, Fransz P, Grossniklaus U. 2004. Nuclear fusions contribute to polyploidization of the gigantic nuclei in the chalazal endosperm of Arabidopsis. *Planta* **220**, 38–46.

- Baroux C, Pecinka A, Fuchs J, Kreth G, Schubert I, Grossniklaus U.** 2017. Non-random chromosome arrangement in triploid endosperm nuclei. *Chromosoma* **126**, 115–124.
- Bauer MJ, Birchler JA.** 2006. Organization of endoreduplicated chromosomes in the endosperm of *Zea mays* L. *Chromosoma* **115**, 383–394.
- Bourdon M, Coriton O, Pirrello J, Cheniclet C, Brown SC, Poujol C, Chevalier C, Renaudin JP, Frangne N.** 2011. In planta quantification of endoreduplication using fluorescent *in situ* hybridization (FISH). *The Plant Journal* **66**, 1089–1099.
- Bourdon M, Pirrello J, Cheniclet C, et al.** 2012. Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. *Development* **139**, 3817–3826.
- Collings DA, Carter CN, Rink JC, Scott AC, Wyatt SE, Strömberg Allen N.** 2000. Plant nuclei can contain extensive grooves and invaginations. *The Plant Cell* **12**, 2425–2439.
- Cowan CR, Carlton PM, Cande WZ.** 2001. The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiology* **125**, 532–538.
- Danilova TV, Friebe B, Gill BS.** 2012. Single-copy gene fluorescence *in situ* hybridization and genome analysis: Acc-2 loci mark evolutionary chromosomal rearrangements in wheat. *Chromosoma* **121**, 597–611.
- Franz P, de Jong JH, Lysak M, Castiglione MR, Schubert I.** 2002. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proceedings of the National Academy of Sciences, USA* **99**, 14584–14589.
- Franz P, Soppe W, Schubert I.** 2003. Heterochromatin in interphase nuclei of *Arabidopsis thaliana*. *Chromosome Research* **11**, 227–240.
- Fujimoto S, Ito M, Matsunaga S, Fukui K.** 2005a. An upper limit of the ratio of DNA volume to nuclear volume exists in plants. *Genes and Genetic Systems* **80**, 345–350.
- Fujimoto S, Yonemura M, Matsunaga S, Nakagawa T, Uchiyama S, Fukui K.** 2005b. Characterization and dynamic analysis of *Arabidopsis* condensin subunits, AtCAP-H and AtCAP-H2. *Planta* **222**, 293–300.
- Fukui K, Kamisugi Y, Sakai F.** 1994. Physical mapping of 5S rDNA loci by direct-cloned biotinylated probes in barley chromosomes. *Genome* **37**, 105–111.
- Gerlach WL, Bedbrook JR.** 1979. Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Research* **7**, 1869–1885.
- Grob S.** 2020. Three-dimensional chromosome organization in flowering plants. *Briefings in Functional Genomics* **19**, 83–91.
- Hoencamp C, Dudchenko O, Elbatsh AMO, et al.** 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**, 984–989.
- Hudakova S, Michalek W, Presting GG, Hoopen R, Santos K, Jasencakova Z, Schubert I.** 2001. Sequence organization of barley centromeres. *Nucleic Acids Research* **29**, 5029–5035.
- Idziak D, Robaszkiewicz E, Hasterok R.** 2015. Spatial distribution of centromeres and telomeres at interphase varies among *Brachypodium* species. *Journal of Experimental Botany* **66**, 6623–6634.
- Jasencakova Z, Meister A, Walter J, Turner BM, Schubert I.** 2000. Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *The Plant Cell* **12**, 2087–2100.
- Kapusi E, Ma L, Teo CH, Hensel G, Himmelbach A, Schubert I, Mette MF, Kumlehn J, Houben A.** 2012. Telomere-mediated truncation of barley chromosomes. *Chromosoma* **121**, 181–190.
- Karafiátová M, Bartoš J, Kopecký D, Ma L, Sato K, Houben A, Stein N, Doležel J.** 2013. Mapping nonrecombining regions in barley using multicolor FISH. *Chromosome Research* **21**, 739–751.
- Kovacic M, Nowicka A, Pecinka A.** 2020. Isolation of high purity tissues from developing barley seeds. *Journal of Visualized Experiments* **164**, e61681.
- Lermontova I, Schubert V, Fuchs J, Klatte S, Macas J, Schubert I.** 2006. Loading of *Arabidopsis* centromeric histone CENH3 occurs mainly during G₂ and requires the presence of the histone fold domain. *The Plant Cell* **18**, 2443–2451.
- Lermontova I, Fuchs J, Schubert V, Schubert I.** 2007. Loading time of the centromeric histone H3 variant differs between plants and animals. *Chromosoma* **116**, 507–510.
- Lermontova I, Koroleva O, Rutten T, Fuchs J, Schubert V, Moraes I, Koszegi D, Schubert I.** 2011. Knockdown of CENH3 in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *The Plant Journal* **68**, 40–50.
- Lysák MA, Čhalíková J, Kubaláková M, Šimková H, Künzel G, Doležel J.** 1999. Flow karyotyping and sorting of mitotic chromosomes of barley (*Hordeum vulgare* L.). *Chromosome Research* **7**, 431–444.
- Misteli T.** 2020. The self-organizing genome: principles of genome architecture and function. *Cell* **183**, 28–45.
- Němečková A, Koláčková V, Vrána J, Doležel J, Hřibová E.** 2020. DNA replication and chromosome positioning throughout the interphase in three-dimensional space of plant nuclei. *Journal of Experimental Botany* **71**, 6262–6272.
- Nowicka A, Tokarz B, Zwyrtková J, et al.** 2020. Comparative analysis of epigenetic inhibitors reveals different degrees of interference with transcriptional gene silencing and induction of DNA damage. *The Plant Journal* **102**, 68–84.
- Nowicka A, Kovacic M, Tokarz B, Vrána J, Zhang Y, Weigt D, Doležel J, Pecinka A.** 2021a. Dynamics of endoreduplication in developing barley seeds. *Journal of Experimental Botany* **72**, 268–282.
- Nowicka A, Sahu PP, Kovacic M, Weigt D, Tokarz B, Krugman T, Pecinka A.** 2021b. Endopolyploidy variation in wild barley seeds across environmental gradients in Israel. *Genes* **12**, 711.
- Olsen O-A.** 2001. Endosperm development: cellularization and cell fate specification. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**, 233–267.
- Pecinka A, Schubert V, Meister A, Kreth G, Klatte M, Lysak MA, Fuchs J, Schubert I.** 2004. Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. *Chromosoma* **113**, 258–269.
- Pfeifer M, Kugler KG, Sandve SR, et al.** 2014. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091.
- Pirrello J, Bourdon M, Cheniclet C, Bourge M, Brown SC, Renaudin JP, Frangne N, Chevalier C.** 2014. How fruit developmental biology makes use of flow cytometry approaches. *Cytometry* **85A**, 115–125.
- Prieto P, Santos AP, Moore G, Shaw P.** 2004. Chromosomes associate premeiotically and in xylem vessel cells via their telomeres and centromeres in diploid rice (*Oryza sativa*). *Chromosoma* **112**, 300–307.
- Rabl C.** 1885. Über Zelltheilung. *Morphologisches Jahrbuch* **10**, 214–330.
- Randall R, Jourdain C, Nowicka A, et al.** 2022. Image analysis workflows to reveal the spatial organization of chromatin and chromosomes. *Nucleus* **13**, 277–299.
- Roudier F, Ahmed I, Berard C, et al.** 2011. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *The EMBO Journal* **30**, 1928–1938.
- Sabelli PA, Larkins BA.** 2009. The development of endosperm in grasses. *Plant Physiology* **149**, 14–26.
- Sanei M, Pickering R, Kumke K, Nasuda S, Houben A.** 2011. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proceedings of the National Academy of Sciences, USA* **108**, E498–E505.
- Santos AP, Shaw P.** 2004. Interphase chromosomes and the Rabl configuration: does genome size matter? *Journal of Microscopy* **214**, 201–206.
- Schubert I, Shaw P.** 2011. Organization and dynamics of plant interphase chromosomes. *Trends in Plant Science* **16**, 273–281.
- Schubert V.** 2009. SMC proteins and their multiple functions in higher plants. *Cytogenetic and Genome Research* **124**, 202–214.

- Schubert V, Klatter M, Pecinka A, Meister A, Jasencakova Z, Schubert I.** 2006. Sister chromatids are often incompletely aligned in meristematic and endopolyploid interphase nuclei of *Arabidopsis thaliana*. *Genetics* **172**, 467–475.
- Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C.** 2014. The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *The Plant Cell* **26**, 2351–2366.
- Shan W, Kubová M, Mandáková T, Lysak MA.** 2021. Nuclear organization in crucifer genomes: nucleolus-associated telomere clustering is not a universal interphase configuration in Brassicaceae. *The Plant Journal* **108**, 528–540.
- Sreenivasulu N, Altschmied L, Radchuk V, Gubatz S, Wobus U, Weschke W.** 2004. Transcript profiles and deduced changes of metabolic pathways in maternal and filial tissues of developing barley grains. *The Plant Journal* **37**, 539–553.
- Waddington SR, Cartwright PM, Wall PC.** 1983. A quantitative scale of spike initial and pistil development in barley and wheat. *Annals of Botany* **51**, 119–130.
- Wegel E, Shaw PJ.** 2005. Chromosome organization in wheat endosperm and embryo. *Cytogenetic and Genome Research* **109**, 175–180.
- Wegel E, Vallejos RH, Christou P, Stöger E, Shaw P.** 2005. Large-scale chromatin decondensation induced in a developmentally activated transgene locus. *Journal of Cell Science* **118**, 1021–1031.
- Young TE, Gallie DR.** 2000. Programmed cell death during endosperm development. *Plant Molecular Biology* **44**, 283–301.

Appendix VI

Core promoter of barley embryo

Pavlu, S., Nikumbh, S., **Kovacic, M.**, An, T., Lenhard, B., Simkova, H.
and Navratilova, P.

Computational and Structural Biotechnology Journal

DOI: 10.1016/j.csbj.2023.12.003

IF₂₀₂₃: 6.155



Research article

Core promoterome of barley embryo

Simon Pavlu^{a,b,1}, Sarvesh Nikumbh^{c,2}, Martin Kovacik^{a,b,3}, Tadaichi An^{d,4}, Boris Lenhard^{e,f,5},
Hana Simkova^{a,*,6}, Pavla Navratilova^{a,*,7}

^a Institute of Experimental Botany of the Czech Academy of Sciences, Slechtitelu 31, 77900 Olomouc, Czech Republic

^b Department of Cell Biology and Genetics, Faculty of Science, Palacky University, Slechtitelu 27, 78371 Olomouc, Czech Republic

^c Merck Sharp & Dohme (UK) Limited, 120 Moorgate, London EC2M 6UR, UK

^d DNAFORM Precision Gene Technologies, 230-0046 Yokohama, Kanagawa, Japan

^e Computational Regulatory Genomics, MRC London Institute of Medical Sciences, London, UK

^f Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, Hammersmith Hospital Campus, London, UK



ARTICLE INFO

Keywords:

Core promoter
Cap Analysis of Gene Expression
Morex
Initiator
TOR-signaling
Hordeum vulgare
Transcription regulation

ABSTRACT

Precise localization and dissection of gene promoters are key to understanding transcriptional gene regulation and to successful bioengineering applications. The core RNA polymerase II initiation machinery is highly conserved among eukaryotes, leading to a general expectation of equivalent underlying mechanisms. Still, less is known about promoters in the plant kingdom. In this study, we employed cap analysis of gene expression (CAGE) at three embryonic developmental stages in barley to accurately map, annotate, and quantify transcription initiation events. Unsupervised discovery of de novo sequence clusters grouped promoters based on characteristic initiator and position-specific core-promoter motifs. This grouping was complemented by the annotation of transcription factor binding site (TFBS) motifs. Integration with genome-wide epigenomic data sets and gene ontology (GO) enrichment analysis further delineated the chromatin environments and functional roles of genes associated with distinct promoter categories. The TATA-box presence governs all features explored, supporting the general model of two separate genomic regulatory environments. We describe the extent and implications of alternative transcription initiation events, including those that are specific to developmental stages, which can affect the protein sequence or the presence of regions that regulate translation. The generated promoterome dataset provides a valuable genomic resource for enhancing the functional annotation of the barley genome. It also offers insights into the transcriptional regulation of individual genes and presents opportunities for the informed manipulation of promoter architecture, with the aim of enhancing traits of agronomic importance.

1. Introduction

Biotechnological research aims to enhance plant utility and achieve greater resilience by exerting complete control over gene expression at two fundamental levels. Firstly, control of transcription is achieved by regulating the quantity of mRNA produced from a specific gene. The second level involves post-transcriptional events that govern the translation of mRNA into proteins. A comprehensive understanding of these

processes is gained by elucidating the function of each nucleotide within the specific genomic sequences involved in controlling expression. The core promoter serves as the ultimate platform for integrating signals of transcriptional regulation, and the complete set of these core promoters in a given species is referred to as the 'promoterome'. The core promoter of eukaryotic RNA polymerase II is defined as the minimal sequence at the 5' end of a gene, essential for initiating transcription. This sequence is recognized and bound by general transcription factors through the

* Corresponding authors.

E-mail addresses: simkovah@ueb.cas.cz (H. Simkova), navratilova@ueb.cas.cz (P. Navratilova).

¹ 0009-0009-2917-6337

² 0000-0003-3163-4447

³ 0000-0002-7470-1585

⁴ 0009-0007-2293-6239

⁵ 0000-0002-1114-1509

⁶ 0000-0003-4159-7619

⁷ 0000-0003-3719-2897

TFIID complex, the largest multiprotein entity composed of the TATA-binding protein (TBP) and approximately 12–13 TBP-associated factors (TAFs) [1]. Such binding is a prerequisite for the assembly of the RNA polymerase pre-initiation complex (PIC), leading to the subsequent events that trigger transcription initiation. Therefore, core promoter regions encompass sequences at the 5' termini of all mRNAs transcribed by RNA polymerase II, starting from the very first nucleotide of the full-length transcript, which are often imprecisely defined in genome annotations. Crucially, promoter diversity stems from varying cofactor binding and activity – a pivotal aspect for modulating gene responsiveness to transcriptional enhancers, the mechanisms of which are not yet fully elucidated in plants [2–4].

Promoter mechanics and sequence composition have been extensively studied in metazoans, with the aim of categorizing these elements to provide a roadmap for their understanding. This topic has been reviewed comprehensively in [5]). A number of well-defined core promoter sequence elements have been characterized, including the initiator element (Inr), TATA box, TFIIB recognition element (BRE), core promoter motif ten element (MTE), downstream promoter element (DPE), among others, which often co-occur in various combinations (reviewed in [6]). Transcription initiation in humans and *Drosophila* is known to commence from a well-characterized, extended Initiator sequence (YYANWYY and TCAKTY, respectively), which may even be sufficient for transcription initiation on its own [7]. Despite their significance and evolutionary conservation, new and lineage-specific promoter elements, located both upstream and downstream of the transcription start site (TSS), continue to be identified in plants (e.g., the Y patch [8,9] or the TC-motif [10]), protozoa and animals [11–13]. The interplay between these elements and variants of the initiator influences the pre-initiation complex (PIC) composition, TSS selection, level of polymerase engagement, transcriptional burst size, cis-regulatory responsiveness, and polymerase pausing [14,15]. These factors collectively have direct implications for gene expression regulation. The common initiator sequence, located at positions – 1 and + 1 relative to the TSS, is a pyrimidine-purine (PyPu) dinucleotide, a pattern confirmed in *Arabidopsis*, maize and rice promoters [8,9]. The sequence downstream of the initiator, within the 5' untranslated region (5'UTR), determines the binding affinity for specific TATA-box associated factors (TAFs) and may include the 5' Terminal OligoPyrimidine (TOP) motif, a sequence responsive to target of rapamycin (TOR) signalling. TOR is a protein kinase conserved across eukaryotes that orchestrates metabolic regulation by promoting anabolic processes in the presence of nutrients. In *Arabidopsis*, TOR is essential for early embryogenesis and for survival under stressful conditions [16].

Despite the variety of scenarios, each species has a limited number of synergistic motif/transcription factor (TF) combinations, with two primary configurations being highlighted: TATA-Inr-MTE and BREu-Inr-MTE, as supported by a PIC structure study [17]. This classification is consistent with the two principal classes of promoters identified in mammals based on the number and distribution of transcription start sites: focused promoters, characterized by a single initiation site approximately 5 base pairs wide ('sharp' initiation), and dispersed promoters, which feature multiple initiation sites spanning up to 100 base pairs ('broad' initiation) [18]. These classes correlate closely with chromatin configurations, including nucleosome positioning, histone variants [19] and their posttranslational modifications [20], which determine the motif-independent 'PIC catchment area' [21]. Intriguingly, these core promoter configurations are also associated with distinct categories of genes, a pattern that appears to be consistent across many organisms, including plants [9,22].

The path to characterizing promoters can be taken from gene annotation or promoter prediction methods, however, the direct proof of promoter position and distribution of TSSs is provided by sequencing of long capped RNA species called Cap analysis of gene expression (CAGE) [23] or similar methods, e.g. GRO-seq, Smar2C2, TSS-seq, [9,24,25]. CAGE, in particular, has been adopted as a fundamental method by large

genome consortium projects, aiding in accurate gene annotation and providing insights into gene expression levels [26,27]. Compared to metazoans, plant promoter characterization has been the focus of only a few systematic studies, largely confined to *Arabidopsis* [28], soybean, rice, sorghum, wheat and maize [9,22]. These studies suggest that sequence features of promoters may vary among plant species. Detailed functional analysis of promoters, using monocot and dicot reporter systems across three species - *Arabidopsis*, maize, and sorghum - has confirmed species-specific, sequence-dependent variations in promoter activity and strength [29]. This indicates that there are differences in the transcriptional machinery both between and within these two groups. The authors have experimentally demonstrated the effects of promoter mutations, highlighting the significance of informed promoter design in transgenesis.

Considering the vital importance of Triticeae crops—wheat, barley, and rye—in human and animal diets, barley (*Hordeum vulgare* L.) presents itself as a suitable model for real-world studies of promoter architecture due to its status as a diploid species with a reference genome assembled to near telomere-to-telomere (T2T) completion [30,31] and the availability of established transgenesis protocols [32]. This makes barley a suitable candidate for research with potential applicability to related Triticeae species. In this study, we generated CAGE datasets from three developmental stages of barley embryos in order to identify promoters involved in various differentiation processes and to elucidate the dynamics of transcription start site usage throughout development. Our data uncovered discrepancies between annotated and detected TSSs, as well as the existence of alternative TSSs, which presumably influence both regulatory and protein-coding sequences. The application of a novel core-promoter sequence clustering approach, combined with the examination of gene functions and epigenetic characteristics, indicated a functional divergence among genes associated with different promoter categories, reflecting disparate cofactor dependencies and modes of transcriptional regulation. Overall, we have demonstrated the power of CAGE promoterome profiling across multiple plant developmental stages in providing nucleotide-level positions along with initiation metrics for all potential alternative TSSs of every active gene. The resultant promoterome dataset lays the groundwork for deciphering the logic of transcriptional regulation and offers a valuable genomic resource for research and agricultural biotechnology applications in barley.

2. Methods

2.1. CAGE

Barley cv. Morex was grown in growth chambers at 16/8 hrs light cycle, 16 °C day/12 °C night temperature. For the 4DAG seedlings, seeds were germinated on wet tissue paper at 20 °C for 4 days before harvesting and removing remnants of seed coat and endosperm. The 8- and 24DAP embryos were staged according to their time of fertilization, size and phenotype and dissected as described previously [33]. Total RNA was extracted by Monarch® Total RNA Miniprep Kit (NEB cat#T2010S) and its quality was checked by Bioanalyzer (Agilent) to ensure that the RIN (RNA integrity number) values were over 7.0. CAGE libraries were sequenced using single-end reads of 150 bp on a NovaSeq instrument (Illumina). CAGE library preparation, sequencing, and read mapping on MorexV3 annotation were performed by DNAFORM (Yokohama, Kanagawa, Japan).

2.2. CAGE data and motif analysis

Obtained reads (CAGE tags) were mapped to the MorexV3 genome using BWA (version 0.7.17). Unmapped reads were then mapped by HISAT2 (version 2.0.5) while rRNA reads were filtered. Mapping rates varied between 41% and 87% of total reads. Regions that had a 90% overlap between replicates were extracted by BEDtools (version 2.12.0).

Tag count data for each of the samples were clustered using the CAGER program, which merged neighbouring TSSs with mutual distance < 20 bp into a single tag cluster (TC). Clusters with tags per million (TPM) < 0.1 were discarded just as singletons that had a TPM signal < 5. MorexV3 annotation (both high and low confidence genes) was used to assign a genomic category to each promoter candidate by applying the ‘annotatePeak’ function from the ChipSeeker package [34]. The genomic features were assigned according to the following hierarchy: promoter > 5’UTR/3’UTR > exon > intron > proximal, where ‘promoter’ has been defined as ranging from – 500 to + 100 bp and ‘proximal’ from – 1000 to – 500 bp relative to the annotated TSS (aTSS) of the nearest gene (Fig. 1B). For each gene ID, the hierarchically highest feature was assigned. In case two candidates had the same hierarchical significance, the candidate located closer to the aTSS was taken as the primary promoter. The rest were set aside as the ‘secondary promoter dataset’. For further analysis, the ‘consensus’ clusters were determined using the ‘aggregateTagClusters’ CAGER function, which aggregates TSSs from the three samples, merging those with mutual distance < 100 bp. The set of consensus clusters created by this method was filtered and split between the primary and secondary TC datasets as described above.

The consensus candidates from the CAGER were further clustered according to their sequence similarity using a seqArchR program (v1.2.0, Nikumbh 2023). The seqArchR configurations were kept default with the exceptions of ‘bound’ being set to 10⁸– 8 and ‘chunk size’ to 5000 for primary promoters (500 for the secondary set) to better suit our dataset. The resulting 49 clusters for the primary dataset were further collated into nine final seqArchR clusters based on sequence logo similarity using the seqArchRplus utilities. The 27 secondary clusters were collated into the final seven. Scripts with more details can be found in the github repository.

The +/- 50 bp around all TC’s dominant TSSs were subjected to peak-motif position analysis by RSAT (Regulatory Sequence Analysis Tools) followed by the hierarchical matrix-clustering method to generate a summary radial tree, both under custom settings documented in the rsat_analysis.txt deposited in the GitHub repository. The TOP score was calculated according to [35] from the 4DAG CAGE BAM file. The calculation script is available at: https://github.com/carsonthoreen/tss_tools/blob/master/tss_analyzer.py.

2.3. Tissue specificity and GO-term annotation of promoter clusters

The tissue specificity values were calculated by applying the Tau algorithm written as R script (deposited at <https://rdrr.io/github/roonysgalbi/tispec/man/>) on tpm (transcripts per million) matrix from 18 distinct samples from the EoRNA (<https://ics.hutton.ac.uk/eorna/index.html>) datasets.

To increase the accuracy of barley GO term annotation we have utilized the Gene Ontology Meta Annotator for Plants (GOMAP)-singularity pipeline [36], which combines three different annotation techniques using the protein fasta file as an input. With the new barley GO-term annotation in hand, the GO-terms were assigned to each gene of the seqArchR clusters. Using the ‘Enricher’ function from the ‘clusterProfiler’ package with default settings we determined the enrichment of GO terms within the clusters, calculating the p-value based on the hypergeometric model. The enriched GO terms together with their p-values were then loaded into the REVIGO web interface [37] to reduce the redundancy of the enriched set and differentiate between the BP (biological process), CC (cellular compartment) and MF (molecular function) GO-term categories. The raw TSV REVIGO data for each of these categories were then used to produce dot plots, showcasing only

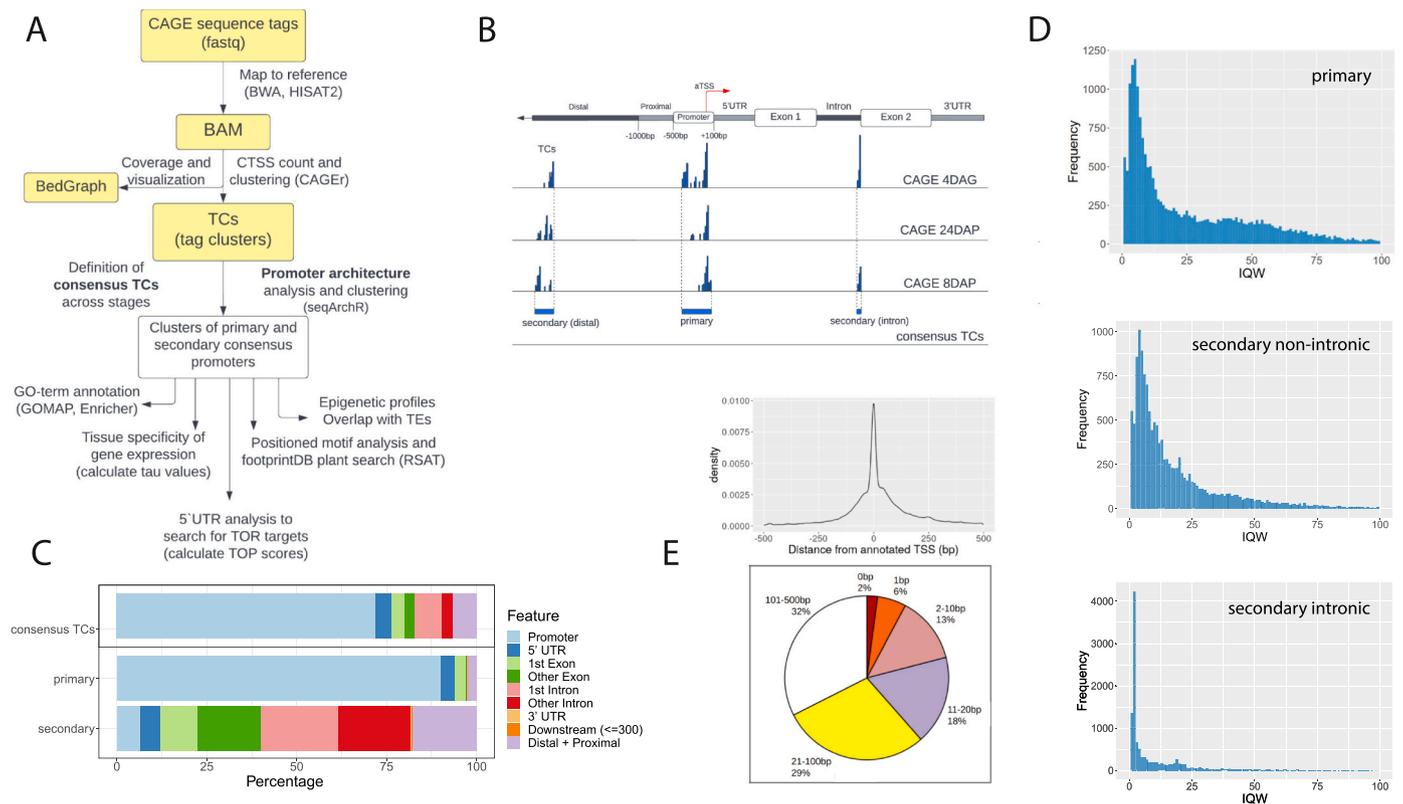


Fig. 1. Initial analysis of the barley CAGE dataset. A) CAGE data analysis workflow. B) Definition of the consensus tag clusters (TCs) and of the primary and secondary promoter set. The aTSS corresponds to the TSS as in MorexV3 annotation. C) Annotation of consensus promoters. The consensus TCs dataset (top bar) was split into the primary and secondary sets (second and third bars) using a filtering method described in the main text. D) Distribution of 24DAP promoter interquartile width (IQW) values for primary, secondary non-intronic and secondary intronic TCs. E) Distribution of distances between dominant CTSSs and aTSSs for barley genes expressed in 8DAP, 24DAP and 4DAG embryos (top), and proportions of expressed genes with a dominant CTSS located within the designated distance from the aTSS (bottom). The analyses were done for dominant CTSSs whose distance from the aTSS was not greater than 500 bp.

the top five most enriched GO terms for each of the seqArchR clusters.

2.4. Histone modification ChIP-seq and MNase-seq data analysis

Barley embryos grown and collected for the CAGE were used for nuclei isolation and MNase digestion followed by native ChIP-seq as described previously [38], with modifications detailed in [31]. These ChIP-grade antibodies against modified histones were used: H3K4me3 (Abcam ab8580) and H3K27me3 (Diagenode C15410195). Resulting sequencing libraries, including those generated from MNase-digested input, were sequenced on the NovaSeq6000 platform. Reads from the ChIP-seq pipeline went through qualitative trimming by Trim Galore (v. 0.6.2) and mapping to the MorexV3 reference genome was performed using the Bowtie 2 package (version 2.4.2). Duplicated reads were then removed using Picard tools (version 2.9.0) and MACS2 software was used to call peaks. For creating heatmaps, the deeptools functionalities computeMatrix and plotHeatmap were utilized, with the kmeans clustering set to 2, to identify any divergence within the promoter clusters. ChromHMM analysis was performed according to the ChromHMM protocol [39]. MNase-digested input WIG files were used to calculate nucleosome positions around dominant CAGE TSSs (CTSSs) using DANPOS software <https://github.com/sklasfeld/DANPOS3>. The resulting WIG files were plotted using deeptools' computeMatrix and plotProfile functions.

2.5. RNA-seq and CAGE correlation data analysis

The 4DAG [40] and 8/24DAP [41] RNA-seq datasets were analyzed using the RSEM software with STAR mapping pipeline [42,43]. The CAGEr datasets were constructed using CAGEr, merging replicas and normalizing tag counts to TPM. ChrUn records were filtered out of both CAGE and RNA-seq datasets and only records that had a TPM value higher than 0.1 were considered.

3. Results

3.1. Cap analysis of gene expression profiles transcription initiation events in barley embryos

To identify transcription initiation events genome-wide in barley cv. Morex, we applied CAGE to total RNA isolated from three embryonic developmental stages: eight days after pollination (8DAP), 24 days after pollination (24DAP) and four days after germination (4DAG), each in two highly correlated replicates each (BAM file Pearson correlation 0.99; Figure S1A). Sequencing of the CAGE samples yielded 52,313,604; 57,927,548; and 60,825,231 CAGE tags from the 8DAP, 24DAP, and 4DAG samples, respectively. Data analysis followed the workflow depicted in Fig. 1A. We employed the CAGEr software [44] to merge replicates, identify CTSSs, and aggregate them into 49,848; 51,903 and 54,276 TCs for 8DAP, 24DAP, and 4DAG, respectively, with custom parameter settings (Figure S1B). Each TC was characterized by its interquartile width (IQW), indicative of promoter breadth, and tag counts, expressed as tags per million, reflecting the expression level. Additionally, each TC was annotated with the position of a dominant CTSS and the associated gene identifier (ID). All TCs were annotated using the ChipSeeker software [34] according to genomic features and gene IDs, considering both high- and low-confidence genes annotated on the MorexV3 pseudomolecules [30] (Figure S1C). TCs located more than 1000 bp from the nearest annotated TSS - a total of 7662; 7296 and 7390 TCs for 8DAP, 24DAP, and 4DAG, respectively - were considered unassociated with the nearest gene and were analyzed as a secondary TC set. We applied a hierarchical approach to retain a single TC, representing the most likely promoter, for each gene (see the Methods section and Fig. 1B for details). This procedure resulted in a primary TC set comprising 19,289; 19,567 and 20,878 promoter candidates for 8DAP, 24DAP, and 4DAG, respectively. The remaining TCs were allocated to

the secondary TC set, which included alternative promoters as well as intragenic and intergenic TSSs, the latter likely encompassing promoters of presumed unannotated genes.

Our objective was to establish a generalized classification of promoter types, therefore we pooled the primary TCs from all three embryonic stages. Overlapping TCs were merged into a set of 34,897 consensus clusters. Following the same filtering and annotation protocols as described above, we derived 21,610 'primary consensus' and 13,287 'secondary consensus' clusters (Fig. 1B, C, and Table S1). These consensus clusters represented the central dataset for the majority of our analyses.

A comprehensive catalog of the TCs, which includes details such as the consensus-TC coordinates in the MorexV3 genome, the assigned gene IDs, positions of the dominant TSSs, feature annotations, classifications, IQWs, and TPMs of all TCs in the sense direction, is provided in Data S1. This dataset is also available in the Eukaryotic Promoter Database (EPD) at <https://epd.expasy.org/epd/>. Given the recognized importance of promoter width in distinguishing different functional classes of promoters, we compared the IQW distributions between the primary and secondary consensus cluster sets. We observed a tendency towards narrower TCs in the secondary set (Figs. 1D and 2). Within the secondary TC set, we identified a subgroup of extremely narrow (width 1–2 bp) TCs, which we termed 'singletons'. These singletons were frequently initiated at the first intron-exon boundary (CAG sequence) and were characterized by notably high TPM values.

3.2. CAGE data improve gene annotation and provide transcription initiation levels

Promoters are characterized by distinct clusters of CAGE peaks. Identifying the most frequently used dominant CTSSs allows for the positioning of the TSSs with single-base pair precision, which ideally corresponds to the gene annotation. To compare the position of these CTSSs with the MorexV3 annotated TSSs (aTSSs), we utilized dominant CTSSs from the primary set, ensuring that their distance from the aTSSs did not exceed 500 bp. This comparison revealed a notable degree of misalignment. Allowing a 20-bp deviation in either direction, 39% of expressed genes (detected by both CAGE and RNA-seq) occurred on the TSS position. Another 29% of aTSSs were located within 100 bp, while the remaining 31% were situated within 500 bp of the dominant CTSS (Fig. 1E). From this primary set, 2583 promoters were associated with the low-confidence gene category. Given these findings, the presence of a CAGE signal in close proximity to the aTSS could be a reason to reconsider the categorization for these specific genes. Supported by RNA-seq and epigenomic data, the CAGE signal enabled us to propose multiple putative unannotated genes, exemplified in Figure S1D.

A summary of tag counts within a CTSS serves as a measure of 5' transcription initiation level, providing an index of gene expression quantified as tags per million. We conducted both qualitative (i.e., which genes were detected by either or both methods) and quantitative comparisons between CAGE and RNA-seq datasets. The quantitative analysis involved calculating correlations of tag/transcript counts for active genes detected by both methods (Figure S2A, B and Data S2). The comparison indicated that CAGE is limited in detecting low levels of gene expression (Figure S2C). Conversely, genes whose transcripts were detected by CAGE but not RNA-seq were not predominantly of low expression (Figure S2D), suggesting that discrepancies may arise from other biological or technical factors.

3.3. Promoter architecture clustering and the initiator

Based on previous extensive research on promoters of other species (reviewed in [45] and elsewhere), we posited that regions 50 bp up- and downstream of the dominant TSS likely contained a PIC-binding sequence. We employed seqArchR [46], a recently developed software that uses unsupervised approach using non-negative matrix

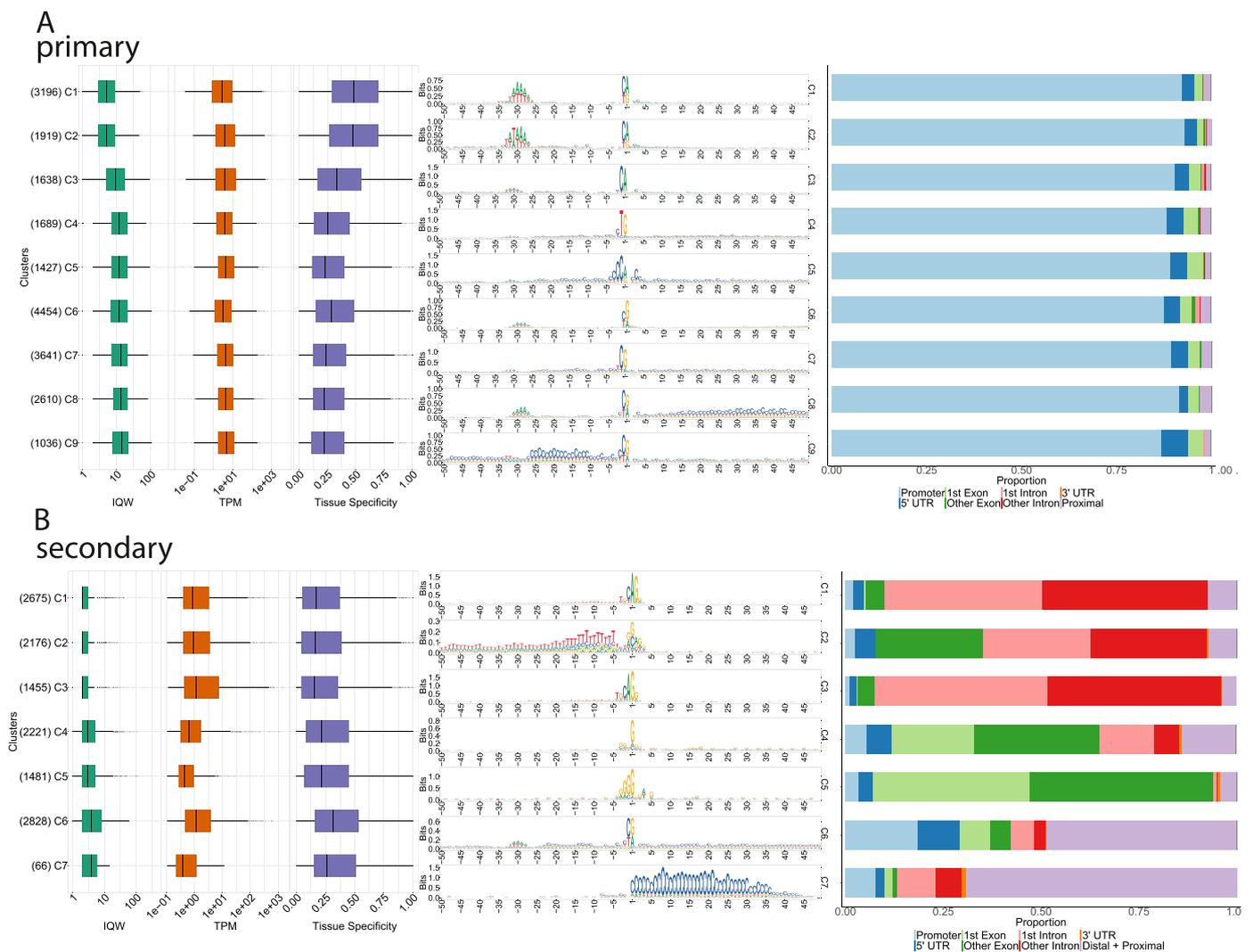


Fig. 2. : Clustering of CAGE promoter sequence architectures. Clusters of consensus promoters were generated by the seqArchR algorithm and ordered by median IQWs. The composed plots for **A)** primary and **B)** secondary promoter clusters include boxplots for IQW, gene expression level values (tags per million (TPM), log-transformed) and tissue specificity (tau values), followed by sequence logos and genomic feature annotation bar plots. The numbers of genes per cluster are given in parentheses.

factorization, to cluster promoter sequences based on their motifs at near-fixed distances from a reference point, such as TSS. These clusters are characterized by de novo-identified sequence elements, such as position-specific motifs and the nucleotide composition of the input sequences. Initially, we analyzed the primary TC sets from all three developmental stages, resulting in 15, 15, and 16 distinct sequence architecture clusters for 8DAP, 24DAP and 4DAG, respectively (Figure S3). In order to provide a generalized classification of promoter types across stages, we created a set of primary consensus promoters and subjected it to the seqArchR analysis. The resulting clusters, defined by their sequence architectures - including the initiator sequence, positioned TATA-box and other sequence motifs - and supplemented with the IQW and expression levels, were manually collated into fewer final clusters, ultimately settling to nine (Fig. 2A). The final number of clusters is the minimum number that still effectively differentiates the main TATA-box positions and other main architectures discovered in the default, non-collated seqArchR output. Further merging was tested but rejected because it would distort positionally restricted motifs, rendering them non-recognizable, and would merge distinct initiator configurations. The initiator, corresponding to the dominant CTSS and discernible as the sequence logo around positions $-1/+1$, emerged as a highly significant motif in all clusters. Notably, barley lacks an extended initiator sequence

motif, including the 'TCT' element, and this absence extends even to the promoters of ribosomal protein genes. The TATA-box containing promoters (clusters 1–3 of the primary set) are predominantly transcribed from a 'CA' initiator, which is associated mostly with genes exhibiting high expression levels. This underscores the pivotal role of the TATA box in dictating both the initiator sequence and promoter activity (i.e. transcriptional burst frequency). This feature is in stark contrast with the more varied pyrimidine-purine (PyPu) constellation ('CG' or 'TG'), prevalent in promoters lacking a distinct TATA box, which tend to exhibit lower expression levels (clusters 4–9, as shown in Figs. 2A and 3A).

The same clustering method was also applied to the secondary TCs (Fig. 2B). Seven distinct clusters best resolve all positioned motif configurations, which can be divided into three main groups. The TCs from clusters 1–3 originate from the first intron-exon boundary. These clusters are characterized by their pronounced sharpness and TPM values that are frequently higher than, and not synchronized with, the TPM of the primary TC of the hosting gene. These TCs are transcribed in the sense direction, and the base at the TSS corresponds mostly to one of the two bases forming the conserved eukaryotic splice acceptor site AG. An example is a gene coding for a 40 S ribosomal protein, which has been assigned to the secondary cluster 1 (Figure S4A). Conversely, secondary

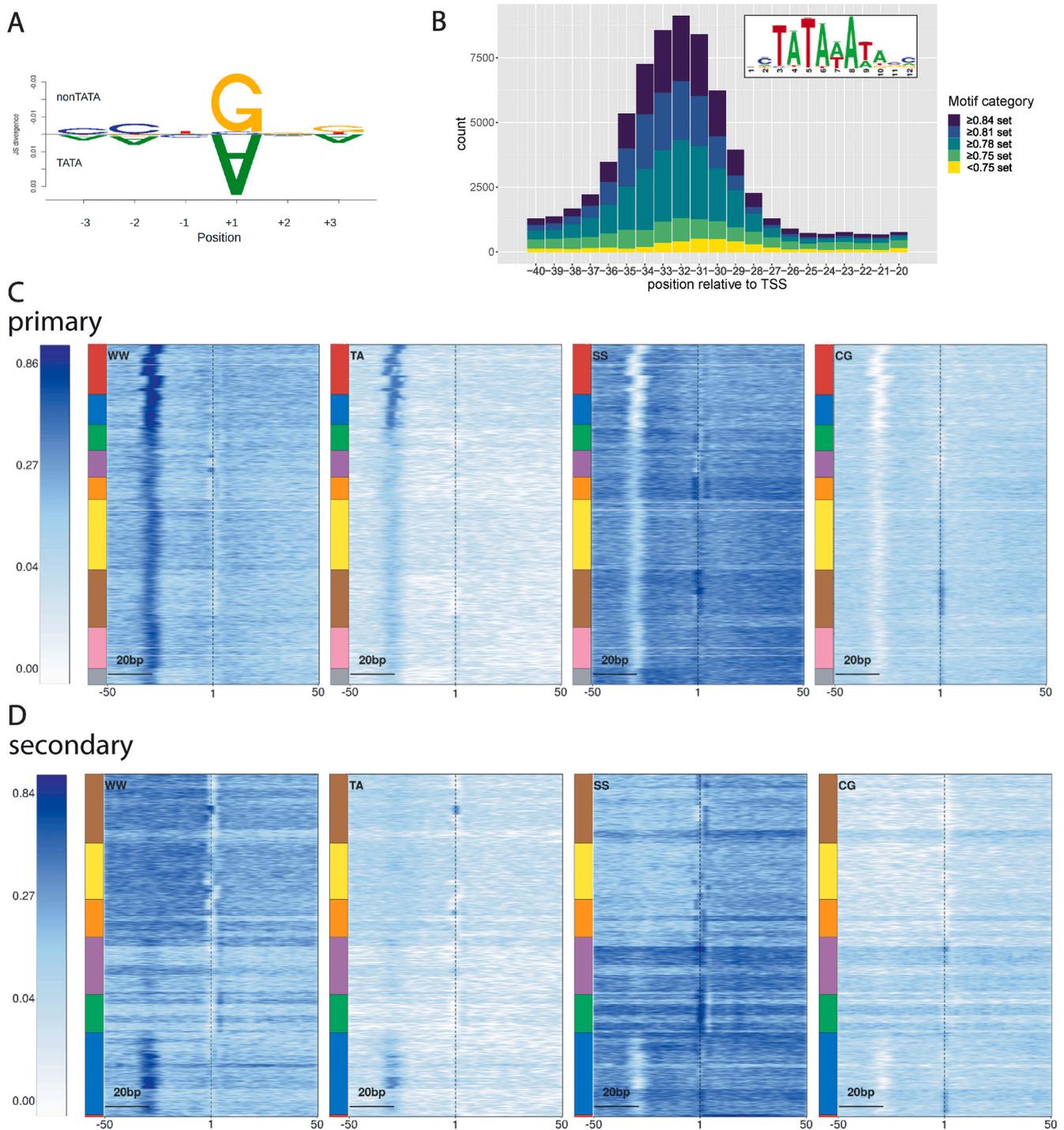


Fig. 3. Sequence analysis of the initiator, TATA-box and dinucleotide content in barley promoter clusters. **A)** TATA vs. non-TATA box promoters differ in the +1 position (A vs. G). **B)** Variability in the sequence and position of TATA-like motifs in the consensus promoter set. The motif categories are based on the degree of Pearson correlation with the canonical TATA-box position weight matrix (PWM). Pentamers included in the individual sets are listed in Table S3. **C, D)** Dinucleotide motif heat maps for the primary **C)** and the secondary **D)** promoter clusters. The heatmaps show enrichment of the given motif in the sequence with values 0–1. The multi-coloured bars left of the heat maps indicate boundaries between the clusters, ordered as in Figs. 2A and 2B, respectively. Note the presence of the W box in all primary promoters and the differences in CG distributions.

clusters 4 and 5 are characterized by both sense and antisense transcription initiated often at G-rich sequences originating at the exons, introns or intron-exon boundaries (Figure S4B). As there may be multiple secondary TCs per gene, gene ID redundancy and overlap between distinct clusters can occur. In the merged set of secondary clusters 1–3, out of a total of 6306 gene IDs, 4190 were unique. Similarly, in the merged set of secondary clusters 4–5, 2866 out of 3702 IDs were unique.

Notably, there is an overlap of 1084 gene IDs between these two sets of clusters, which is statistically significant according to Fisher's exact test (p-value of 0.000461). To determine whether this overlap might also indicate overlapping functions, we examined the GO enrichment in the 1084 genes with secondary TCs from both cluster sets 1–3 and 4–5. We isolated these genes from the rest of the unique gene IDs. The resulting Figure S8C demonstrates that the gene functions of these genes are

confined to the same categories of metabolic and housekeeping functions that have been assigned to the original clusters. Interestingly, the genes with secondary TCs in both cluster sets 1–3 and 4–5 exhibit the highest enrichment levels.

Clusters 6 and 7 likely represent true alternative promoters, which are generally TATA-less, and promoters of potentially unannotated genes, as shown in Figure S1D.

3.4. The TATA box and other previously defined motifs in barley promoters

Consensus primary clusters 1–3, characterized by a positioned TATA-box-like motif, delineate the TSSs within significantly narrower regions (sharp promoters), in contrast to the broader regions associated with non-TATA promoters. This pattern is consistent across all studied organisms, although the position of the TATA box and the frequency of its occurrence do vary. To quantitatively analyze the TATA-box sequence with reproducibility, we utilized the position weight matrix (PWM) of the canonical plant TATA box (Fig. 3B inset). This PWM was generated

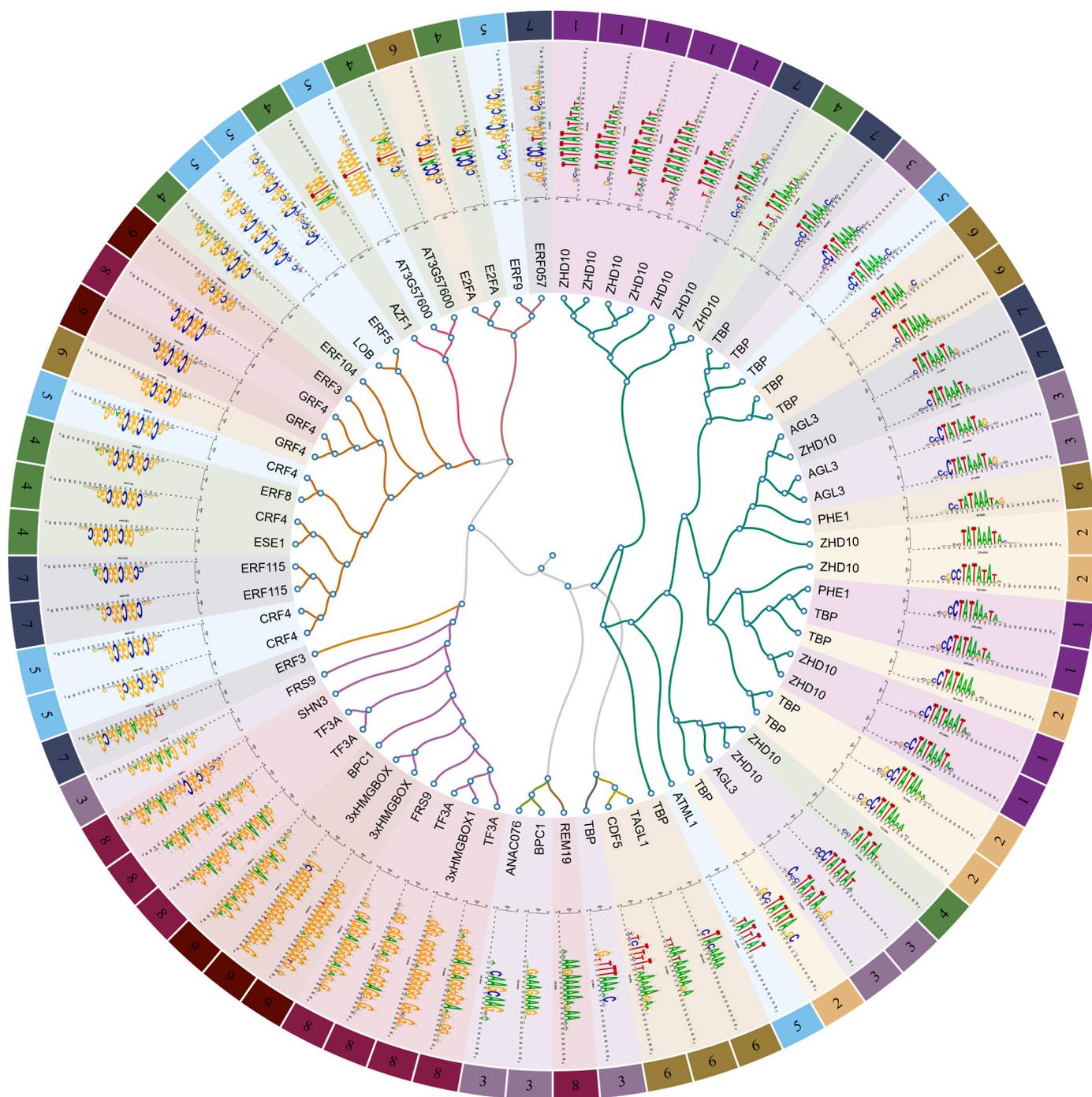


Fig. 4. Positioned sequence motifs in consensus core promoters. Region \pm 50 bp relative to the TSS was searched for TFBS motifs collected in footprintDB. plants database. Proteins at the tree branches represent the motifs with the highest probability of occurrence identified in the search. The colors and numbers in the outer circle correspond to distinct primary promoter clusters.

from 134 unrelated plant promoter sequences deposited in the Eukaryotic Promoter Database [47]. A strand-specific search for the TATA-box frequency using this PWM and the FIMO tool [48] with the default p-value threshold of $1e-4$ revealed that TATA boxes were present in 7.8%, 7.9%, and 9% of all promoters active at the 8DAP, 24DAP, and 4DAG stage, respectively. Employing a more lenient p-value of $1e-3$, which included motifs diverging significantly from the canonical TATAWAW, the proportion of active promoters increased to 20%, 20%, and 23%, respectively.

Positional analysis using the same canonical PWM and TA pentamers as in [49] revealed that the starts of the TATA-box and more variable W-box motifs are located at -29 to -36 upstream of the initiator site, peaking at -32 . Greater distances were observed for the more conserved TATA-box motifs (Fig. 3B). A certain proportion of TATA-box promoters are associated with distinct cytosine nucleotides immediately upstream of the TATA box, as indicated by subsequent motif discovery (Fig. 4). This could suggest the presence of the BRE-upstream sequence (BREu, SSRGCC), although neither the full BREu nor the BRE-downstream sequence (BREd, RTDKKKK) were detected (Figure S5). Apart from the TATA boxes and specific initiators, other previously recognized core promoter motifs were not observed (Figure S5). Nonetheless, dinucleotide heat maps produced by seqArchR provide a comprehensive view of the common PyPu dinucleotides at TSSs, which include the omnipresent W boxes ($W=A/T$) and a high content of SS dinucleotides ($S=G/C$) in the majority of barley promoters (Fig. 3C, D).

To detect less pronounced motifs within the core promoter regions of each cluster and to identify, which regulatory proteins might target them putatively, we supplemented the cluster analysis with an analysis using the RSAT toolkit [50]; <https://github.com/rsa-tools/rsat-code>. Specifically, we utilized a ‘peak-motif positioned’ function that detects oligonucleotides showing a positional bias, i.e. having a non-homogeneous distribution in the sequence set, followed by a TFBS search against footprintDB.plants [51]. For a comprehensive summary, the collection of motifs from all nine consensus clusters in the primary set underwent hierarchical clustering, which is depicted as a radial tree (Fig. 4, Data S5). This analysis confirmed that, in addition to the well-known canonical TATA box motif (TATAWAW), there is a significant occurrence (approximately 5.5% of cases, p-value = $1e-4$) of extended TA repeats, which are sequences of 6–15 TA dinucleotides that do not align with the TBP matrix in footprintDB. Instead, these sequences seem to be targeted by the zinc finger protein ZHD10, which plays a role in establishing leaf polarity during development via gibberellic acid signaling pathway. Furthermore, other putative W-box motifs are consistent with the binding matrices of AGL3/PHE1, both of which are MADS-box homeotic transcription factors.

Additionally, other motifs within barley core promoters have a low-complexity nature, such as the previously described pyrimidine-rich Y patch [8], which has been experimentally demonstrated to enhance gene expression in the maize reporter system [29]. This motif, characterized by the sequence CTCTTCCTC or its reverse complement GAGGAAGAAG, is present in over 14% of barley core promoters under strict search conditions (p-value $\leq 1e-5$) and in up to 70% when using relaxed criteria (p-value $\leq 1e-4$). The TFBS analysis conducted with RSAT attributes this motif to BPC1 (BASIC PENTACYSTEINE1), a GA-repeat-binding protein with an octodinucleotide binding preference, which has a homolog in barley known as BBR. This protein is thought to be involved in the Polycomb-mediated transcriptional regulation of developmental genes through its interaction with Polycomb Repressive Elements [52]. Other significant findings include 3xHMG-box proteins associated with cell proliferation and implicated in the organization of plant mitotic chromosomes [53], as well as hormone-responsive factors TF3A, FRS, and SHN3. Another group of low complexity motifs prevalent in barley core promoters is the GCC box, which is recognized by a GCC box-binding factor (GBF) and/or ethylene-responsive factors (ERFs); these are all hormone-responsive proteins implicated in stress

responses and developmental processes. The analysis of an extended sequence range (± 100 bp) did not reveal any novel motifs (Figure S6). Other TF bindings in core promoters include E2F, MYB, and NAC transcription factors.

3.5. CT-rich motifs downstream TSSs might act as nutrient-sensing TOP motifs

Although the search for TFBS within the promoter sequences yielded numerous matches, the region immediately downstream of the TSSs may function beyond transcriptional regulation. This region has been demonstrated to act as a translational signal, particularly via the TOP motif, which commences with a cytosine at the 5' cap, followed by a sequence of uracils and/or cytosines, with few or no adenines or guanines. To investigate the potential of certain barley promoter sequences to serve as a platform for TOR signaling, we calculated TOPscore values for our CAGE dataset for 4DAG using the algorithm described in reference [35]. This revealed a set of mRNAs that begin with a likely 5'TOP motif as potential subjects to the TOR-5'TOP nutritional signaling pathway. Overall, the distribution and means of TOPscores for each cluster were significantly skewed ($p < 0.001$, Welch's two-sample t-test) toward lower values in TATA-box clusters 1–3 and cluster 6, while the remaining clusters displayed significantly higher values (Fig. 5A). A total of 558 barley genes had TOPscore values greater than 3, considered to carry the bona fide TOP motif signature in their 5'UTRs. GO term analysis of these candidate genes revealed significant enrichment in ribosomes, the Golgi apparatus, and plastids, with functional involvement primarily in sugar metabolism, cell division, and plant growth (Fig. 5B). This suggests that they are likely direct targets of the nutrient-sensing pathway.

3.6. Tissue specificity and GO enrichment analysis

To evaluate the relationship between gene features driven by specific promoters and the categories of these promoters, we initially examined the expression breadth or specificity of these genes throughout plant development and in various tissues. We computed a tau value for each gene using a TPM matrix encompassing a wide array of tissues as presented in the developmental transcriptomic dataset [40], which is accessible in the EoRNA database [54]. The distribution of tau values for individual promoter clusters again distinctly separates the clusters containing TATA boxes, which exhibit significantly higher tissue specificity, from those without TATA boxes, which are generally more ubiquitously expressed (Fig. 2A, Figure S7). Remarkably, the genes that display a pronounced capped-mRNA signal at the intron-exon junction, namely the TCs in secondary clusters 1–3, are the most ubiquitous (Fig. 2B).

To conduct a GO enrichment analysis, we generated a detailed plant-centered GO annotation using the GOMAP toolkit [36]. Compared to the published MorexV3 GO-term annotation [30], generated by the Automated Assignment of Human Readable Descriptions (AHRD) pipeline [55], GOMAP generated additional 3491 terms, assigning one to each barley gene, including those with a missing GO term in the published annotation (37% of the analyzed gene set). Annotation of several well-defined gene categories (e.g., histones, auxin-responsive genes, MADS-box) confirmed a better definition of gene functions compared to the published version (Figure S8A, B). The new barley GO-term annotation is available in GAF format as Data S3.

The enrichment analysis of GO terms within individual promoter clusters revealed that genes regulated by the most active and narrowly defined TATA-box promoters were annotated as responsive to environmental stimuli, stress, and signals related to hormonal, developmental, and organ growth processes (see Fig. 6A). These included a response to karrikin - a plant growth regulator structurally similar to strigolactones - implicated in seed germination, nitrate response, peroxidase activity, and protein folding. In a targeted approach, a manually curated subset of

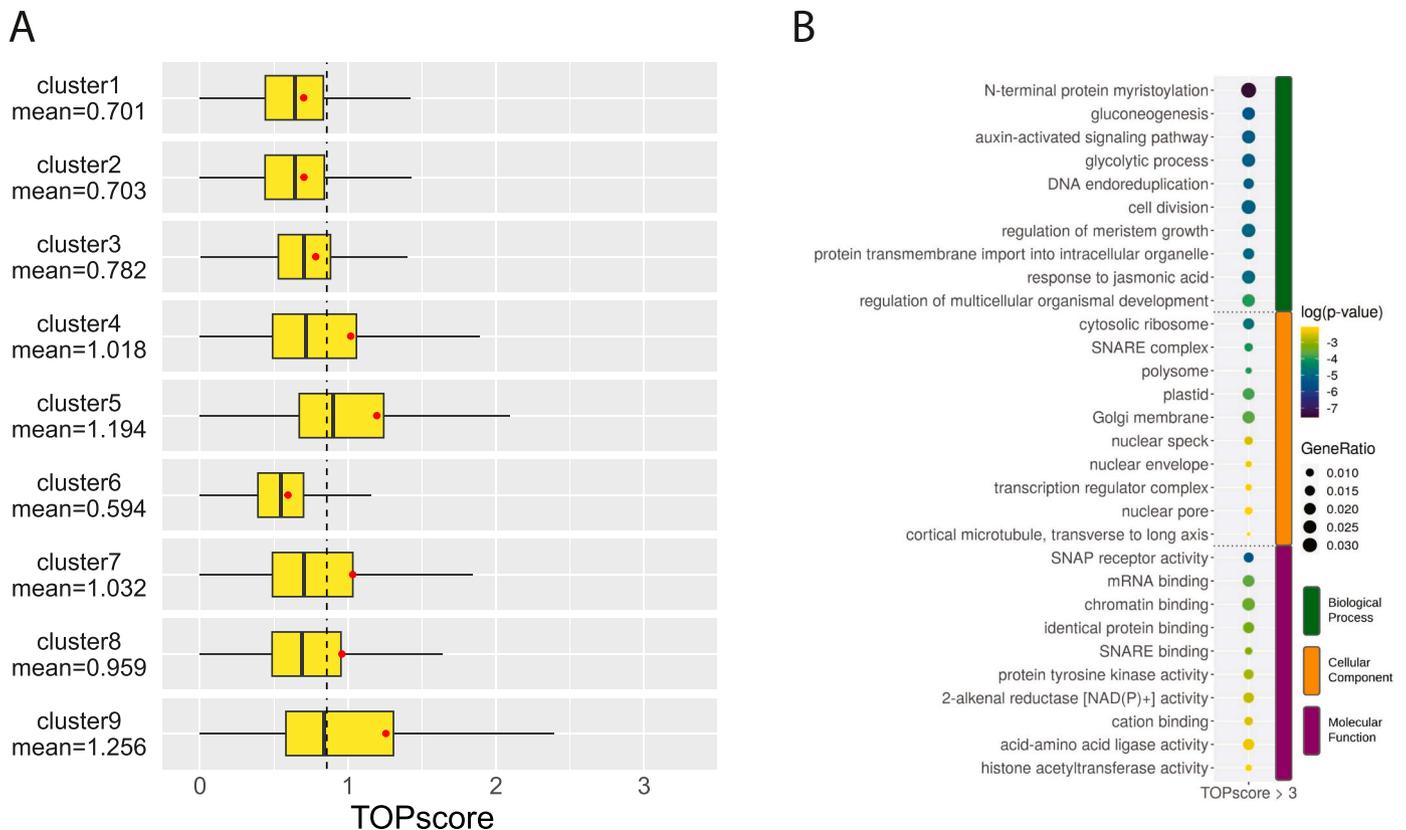


Fig. 5. Analysis of TOP Motifs Across Promoter Clusters. **A**) Distributions of TOPScores in the primary consensus clusters. Each cluster exhibited either a significantly lower or higher TOPScore mean (represented by red dots) compared to the dataset's overall mean (mean=0.86, indicated by a dashed line), with statistical significance (p -value < 0.001, Welch Two-Sample t -test). **B**) GO enrichment analysis of genes with promoters containing the candidate TOP motif (TOPScore > 3).

116 genes responsive to auxin showed a clear association with the first three promoter clusters (Fig. 6C), as did genes encoding histones. Clusters 4–9, lacking a TATA box, exhibited a substantial overlap in functions attributable to basic cellular processes and some metabolic activities, as well as to translational regulation (including RNA modification and Golgi apparatus functions), exemplified by genes encoding ribosomal proteins (Fig. 6C). Notably, GO terms pertaining to transcriptional regulation - such as transcription factor binding to cis-regulatory regions or chromatin binding, which are typically linked with developmental genes - were represented in both categories of promoters in the primary dataset. Unexpectedly, secondary clusters 1–5 displayed a significant degree of functional similarity (Fig. 6B). These functions could be characterized as predominantly metabolic, related to ribosome structure and function, and involved in glucose metabolism, photorespiration, as well as RNA methylation and binding.

3.7. Promoter developmental shifts across barley embryo stages

Altering promoter selection is one of the mechanisms for cell type differentiation. The use of an alternative promoter, which manifests as a distinct TC associated with the same gene, can coincide with a change in promoter architecture. This often involves switching between TATA-box and TATA-less promoter types. The putative alternative promoters identified in our study predominantly belong to secondary cluster 6. A subset of these alternative promoters is developmentally regulated, which we have termed "moving promoters". The average length of promoter shift between stages examined in our study was approximately 500 bp, typically encompassing the 5'UTR, the first exon, or intron, thereby influencing the length of the UTR or the coding sequence. In our CAGEr consensus dataset, we sought instances of these shifts, considering only those TSS pairs that involved at least one of the following

elements: promoter, 5'UTR, and promoter-proximal sequence (ranging from -500 to -1000 bp relative to the TSS). We then assessed the potential TSS shifts between embryonic stages based on the TPM values of each TC in the pair. Comparing pairs of stages with the highest stage-specific TPM values, we revealed 182, 154 and 160 genes with developmentally changed TSSs between the stage pairs 8DAPx4DAG, 8DAPx24DAP and 24DAPx4DAG, respectively (Data S4). Of these genes, 60, 34, and 51 involved the coding region, potentially altering the amino acid sequence, while 110, 115, and 103 affected the promoter/5'UTR region, potentially impacting transcriptional, translational, and transport signals. For instance, an alternative promoter in the first intron of an Argonaute protein gene is active in embryos at 4DAG and produces a transcript truncated at the 5' end compared to the 8DAP stage. The transcription initiation alternatives were also reflected in the RNA-seq and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) data (Fig. 7A). Fig. 7B presents an example of a 5'UTR/promoter shift for a kinase-like protein gene.

3.8. Epigenetic characteristics of barley promoters

The two primary modes of transcription regulation - TATA-dependent, characterized by sharp promoters (or TCs), and TATA-independent, associated with broad promoters (or TCs) - are known to correlate with distinct promoter-proximal nucleosome positioning in metazoans [56]. To investigate the relationship between nucleosome positioning and barley promoter architectures, we utilized data from MNase-digested DNA sequencing of 24DAP embryos. Resulting plots of nucleosome distributions show a dominant nucleosome immediately downstream of the TSS - a characteristic shared by all promoters - but also two distinct profiles for other adjacent nucleosomes (Fig. 7C, detailed in Figure S9). In the region upstream of the TSS, the narrower

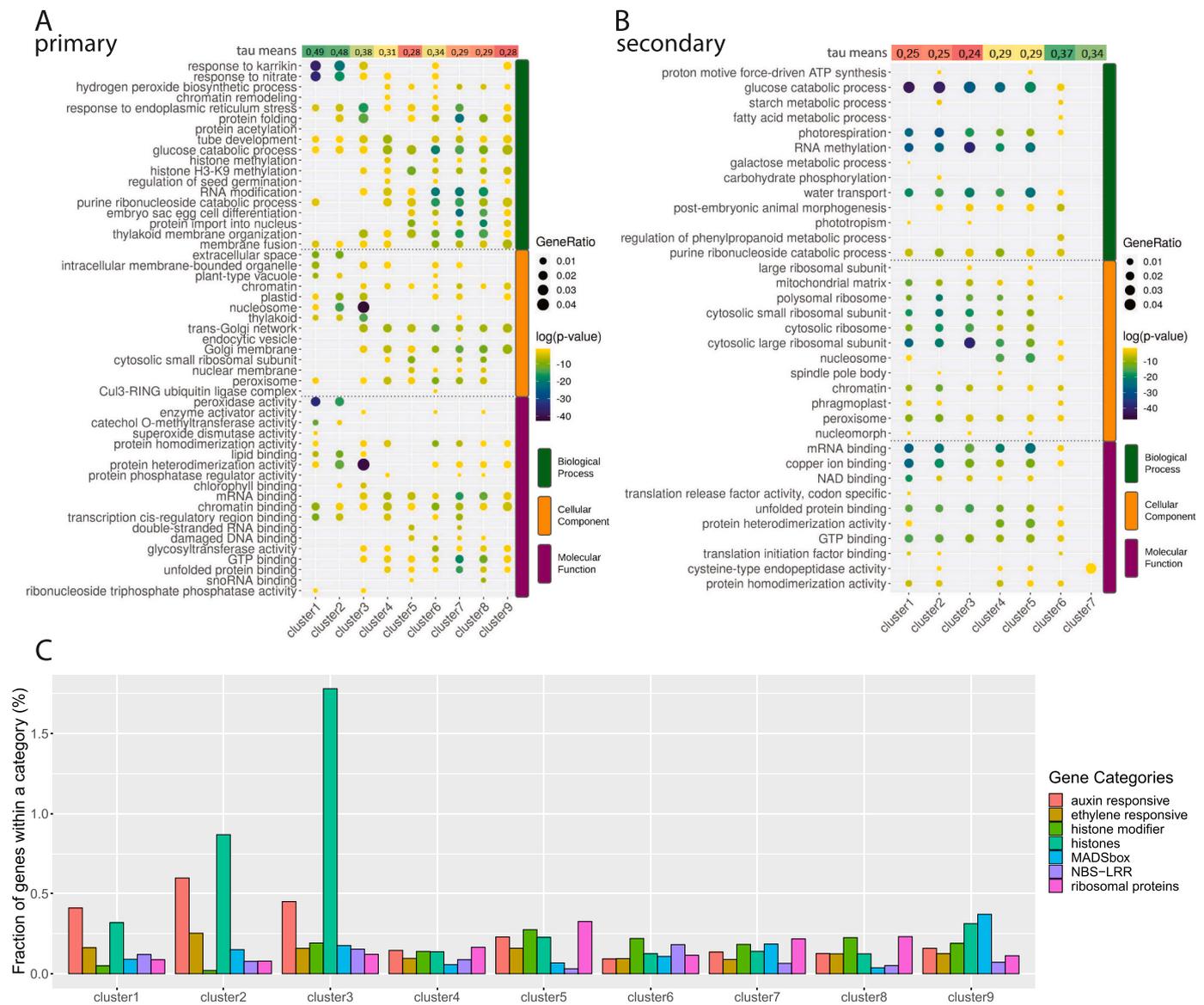


Fig. 6. GO enrichment analysis. The plot shows the top five GO terms associated with primary **A)** and secondary **B)** consensus clusters. The uppermost row, labeled as 'tau means', indicates the tissue specificity scores for genes in each cluster. In panel A, Clusters 1–3 are identified as TATA-box clusters. **C)** illustrates the distribution of selected gene categories within the clusters of the primary set.

TATA-box promoters exhibit a well-positioned nucleosome, as opposed to non-TATA promoters, which display a nucleosome-free region. Conversely, several nucleosomes downstream of the TSS appear to be very precisely positioned in the latter category of promoters but are less well-positioned in the TATA-box group.

The key promoter nucleosomes are typically marked by histone-3 post-translational modifications, specifically H3K4me3 and H3K9ac, in active genes. We conducted native chromatin immunoprecipitation (ChIP) using antibodies specific to histone modifications, including those two activating ones as well as the Polycomb- and facultative heterochromatin-related H3K27me3. K-means clustering of each cluster profile revealed that TATA clusters contained two distinct composite profiles: H3K4me3 plus H3K9ac and H3K4me3 plus H3K27me3, resembling the bivalent marks considered to poise the expression of developmental genes (Figure S10A, B). In contrast, the silencing H3K27me3 mark was almost absent from the non-TATA promoters, which aligns with their lower tissue specificity. Chromatin-state (CS)

analysis using ChromHMM [57] integrates genome-wide profiles of multiple histone modifications obtained by ChIP-seq. This method employs combinatorial and spatial patterns of marks to annotate each tissue sample. We utilized ChromHMM to model the presence or absence of each of the three histone modifications, thereby learning the CSs. These states were then applied to generate annotations for each tissue/stage, determining the most probable CS for each genomic segment. Finally, we performed enrichment analysis on these annotations for individual stage-specific clusters (Fig. 7D). This analysis confirmed that all TATA-box promoter clusters contained both activating marks (H3K4me3 and H3K9ac), as well as, albeit to a lesser extent, the silencing H3K27me3 mark. Notably, the proportion of promoters bearing the silencing mark was observed to increase concomitantly with cell differentiation.

An important feature of promoters, which is closely related to the epigenetic landscape, is their transposable element (TE) content. In the human genome, approximately 18% of start sites have been defined by

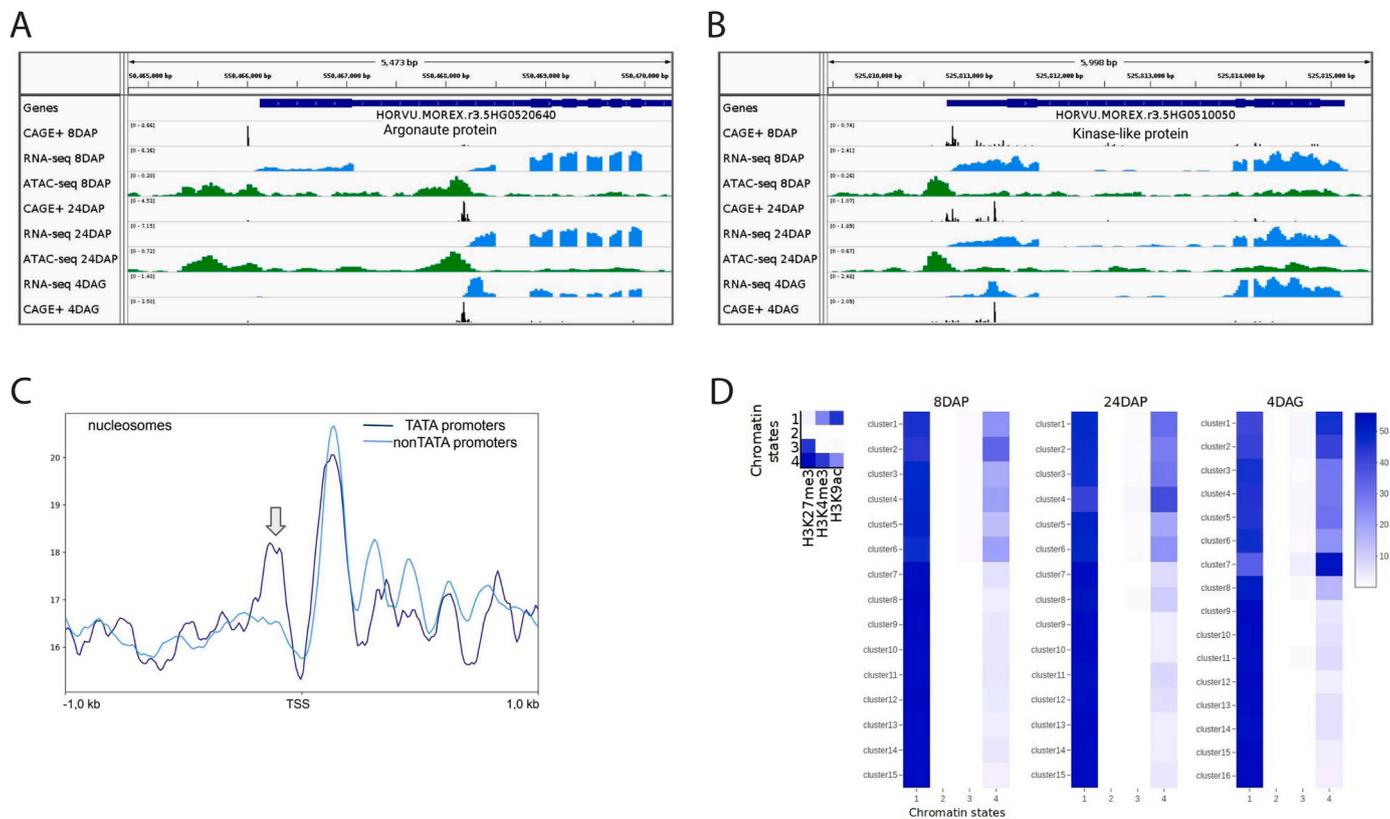


Fig. 7. Barley promoter shifts and epigenetic context. A) A developmentally regulated alternative first exon in an Argonaute protein gene, accompanied by the dynamics of the epigenetic environment, as shown by open-chromatin profiles (green peaks) across three embryonic developmental stages. B) An example of an alternative TSS that shifts during development altering the 5' UTR sequence. The shift is validated by differences in RNA-seq and open-chromatin profiles. C) The nucleosome positioning for the TATA (merged clusters 1–3 of the primary consensus set) and non-TATA promoters (clusters 4–9). A well-positioned upstream nucleosome (indicated by an arrow) is correlated with the presence of the TATA box, contrasting with the nucleosome-free region and well-positioned downstream nucleosomes in TATA box-lacking promoters. D) Four chromatin states, inferred using ChromHMM from ChIP-seq datasets at 8DAP, 24DAP, and 4DAG, are based on three key histone modifications (left). Stage-specific heat maps (right) display the enrichment of these states across promoter clusters at different stages. The color key represents the fold enrichment of each chromatin state within each promoter cluster. Notably, the silencing H3K27me₃ mark, predominantly found in chromatin states 3 and 4, shows higher enrichment in TATA-box clusters (i.e., clusters 1–6 for 8DAP and 24DAP, and clusters 1–7 for 4DAG), as further detailed in Fig. S3.

CAGE as overlapping with TEs [58]. Intriguingly, an inspection of the degree of overlap between TEs, as annotated in [40,59], and regions within ± 50 bp of TSSs revealed that only 5–6% of TATA box-containing promoters overlapped with TEs. In contrast, 10% (cluster 6) to 15% (cluster 5) TE-overlapping promoters in the non-TATA category did overlap, as shown in Figure S10C). Predominant TE families included DNA transposons *CACTA* and retrotransposons such as *Copia* and *Gypsy*, along with others in the LTR-RLX class that are unknown. The overlap of these elements with promoters raises new questions about their role in plant gene regulation, development, and evolution.

4. Discussion

We investigated barley core promoter sequences utilizing precisely delineated transcription start sites identified by CAGE and employing a novel, unbiased data analysis approach. This approach accounts for the positional restrictions of motifs within the promoters and categorizes them based on their overall sequence architecture. It is independent of pre-existing motif knowledge and circumvents the detection of non-conserved noise, which typically arises from the examination of inaccurate promoter sequences and non-positioned motifs. Moreover, our method does not rely on the previously emphasized distinction between sharp and broad promoters, eliminating the need to define a boundary between the two categories. Still, our findings align with those from studies on metazoan promoteromes [5], indicating that the presence of

the TATA box is crucial for governing TSS selection within a more constrained region associated with the CA initiator. Conversely, the presence of other W-box sequences, or their absence, appears to result in a more relaxed or flexible TSS selection that includes various PyPu initiator configurations.

Although the applied methods vary slightly, the proportion of promoters containing a relaxed form of the TATA-box-like motif was found to be approximately 20% in both the barley and the human promoterome [60]. A notably higher proportion (38%) was observed in maize genes active in both root and shoot tissues [22], which may reflect the fact that the TATA-box promoters are associated with tissue-specific expression primarily in adult tissues [61], as opposed to the early developmental stages analyzed in our study.

The TATA-box position is not strictly conserved across species, and its distance from the initiator element does not appear to be directly proportional to genome size. In barley, the TATA box is found within a wider range (from -29 to -36 nucleotides upstream of the TSS), which is slightly shifted in comparison to the positions commonly observed in metazoan genomes. In metazoans, the TATA box is typically located at a more constrained position, usually at the 30th or 31st nucleotide upstream of the TSS [60]. This variation may be associated with wider transcriptional core sequences (TCs) linked to the TATA box in plant species as opposed to those in animals. Similar spacing has been reported in *Arabidopsis* [29], approximately at -30 nucleotides upstream of the TSS, and in maize, sorghum, rice, and wheat at -34 nucleotides [9], indicating some degree of conservation among plant species.

In our observations, we identified a W-box variant that does not conform to the consensus sequence of the TBP-motif position weight matrix. This raises the question of whether factors other than TBP, evolutionarily related or not, can substitute for this essential protein in the pre-initiation complex, as predicted by motif analysis tools. Conversely, it is also worth investigating whether TBP can bind to motifs diverging from the TATA box, as has been reported in yeast [62]. In addition, the BREu, a C-rich element located upstream of the TATA box, was found to enhance promoter activity in maize when inserted into an artificial promoter in a transgenic assay. This effect contrasts with findings in tobacco and is particularly notable given that the motif is reportedly absent in maize [29]. We found only a partial match to this element (the ‘CC’ sequence), leaving its significance in barley to be functionally tested.

The two principal TSS selection modes are associated with nucleosome positioning, tissue specificity, and the epigenetic profile. These align with the concept of two distinct regulatory environments. However, the intricate interplay between plant hormones—which perform both developmental and housekeeping functions—and the inherent plasticity of plant cells, obscures the distinction between housekeeping and developmental gene promoters. This is probably reflected in our GO term analysis. In relation to this, the bivalent chromatin states identified in TATA-box-containing promoters facilitate timely activation while also enabling repression in the absence of differentiation signals [63]. Given the relatively complex nature of our samples, the presence of H3K27me3 could also be indicative of the repression of tissue-specific promoters in a subset of cell types. Therefore, the bivalent histone modification status should be confirmed or ruled out through a sequential ChIP experiment.

Our comparison of the dominant TSSs from CAGE datasets to the most recent (MorexV3) barley annotation resulted in relatively frequent misalignment: 61% CTSSs were located over 20 bp and 32% more than 100 bp distant from the aTSS. This discrepancy may be partly attributed to promoter shifts between tissues and developmental stages, given that the embryonal samples analyzed in our study were only marginally represented in the RNA-seq dataset [40] utilized for the MorexV3 annotation. This underscores the importance of creating a comprehensive promoterome atlas across multiple tissues and cell types for a given organism. Such an atlas would make it possible to focus on the relevant regulatory regions in gene cloning and editing projects. In this context, CAGE stands out as a cost-effective complementary technology capable of significantly enhancing even high-quality genome annotations that are primarily based on RNA-seq data.

To understand the differences between CAGE and RNA-seq, it is essential to recognize that these methodologies target disparate segments of RNA molecules: specifically, the capped 5′ ends and random RNA fragments, respectively [64]. The decreased sensitivity of CAGE is likely due to the TPM threshold applied to filter out widespread, yet low-level, intergenic signals, regardless of whether they are background noise or actual transcriptional events. Additionally, the depth of sequencing may have also contributed to this limitation. The requisite sequencing depth for barley was extrapolated from numerous studies conducted on human cells, which may have resulted in an underestimation. Such an underestimation can occur due to the broader promoter regions, along with distinct levels of intergenic transcription and background noise present in barley compared to those in human cells.

In comparison to other cereals, barley exhibited a smaller proportion, approximately 60.5%, of transcription start regions (TSRs) located in gene-proximal regions, while rice, maize, and sorghum had approximately 69–74% and wheat 49–54% of TSRs within a 1,000-bp distance from annotated genes [9]. These differences may be attributed to variations in detection techniques (CAGE versus Smar2C2) and data processing pipelines or could potentially signify differences in numbers of unannotated genes in the cereal annotation.

In addition to transcripts that overlap with annotated promoters, we identified over 2000 predominantly single-base-wide CTSSs emanating

from the first intron-exon junction at the splice acceptor site. Similar transcripts, transcribed in the same direction as the gene, have been previously observed in mammalian CAGE datasets and are characterized as products of post-transcriptional cleavage and recapping. It has been hypothesized that these transcripts give rise to truncated mRNA isoforms that could potentially be translated into proteins with truncated C- or N-termini. [65]. The authors referred to these as ‘intraexonic CAGE tags’ and observed that they tend to manifest as tissue-specific variants, indicating a highly regulated process that presumably augments mRNA abundance. Alternatively, these CTSSs may arise as artifacts of co-transcriptional splicing or due to a deceleration and subsequent re-capping of RNA polymerase or may be a subject to post-transcriptional cleavage by Argonaute and used as regulatory non-coding RNA [66]. Lastly, they might result from intron-dependent loop formation, which is associated with a mechanism of transcriptional enhancement, as reviewed in [65]. Further characterization of these transcripts through rapid amplification of cDNA ends (RACE), RNA-binding protein immunoprecipitation, or the detection of corresponding translation products at the protein level would be a valuable extension of this research.

Our results demonstrate that in barley, the core promoter motifs identifiable are limited to the PyPu initiator, the W box, and the Y patch - previously established elements - supplemented by stretches of low-complexity sequences that exhibit some dynamic behavior and flexibility during development. The promoter sequence architecture seems to evolve through the exaptation of novel regions, including transposable elements, or via alterations in ancient promoters due to degeneration and changes in binding preferences. We posit that ancient cellular functions are more commonly orchestrated by sequences containing the TATA box and simple TF-binding motifs, whereas more specialized functions tend to be associated with less conserved promoter sequences. These less conserved sequences may facilitate polymerase scanning [67], interaction with co-activators, and initiation in response to distal TFBS.

Artificial promoter design and regulatory sequence manipulation are common engineering methods to drive or influence transcription levels. The knowledge of PIC interactions have been exploited by humans to create a highly active core promoter, termed the ‘super core promoter’, that is capable of engaging in high-affinity interactions with TFIID through the presence of optimal versions of the TATA, Inr, MTE, and DPE motifs [17]. Similarly, a plant ‘super-promoter’ has been designed by [29] who also indicated that for optimal results, species-specific promoters might be preferable in transgenic designs. Although functional validation of each individual promoter type is beyond the scope of this article, such validation can be readily addressed in the future using *in vitro* assays with a reporter gene, including tests for hormone, stress, and nutrient effects on promoter activity.

Our study, conducted on developing and germinating barley embryos, yielded comprehensive information regarding the initiation of transcription in 21,610 genes active during these specific stages. To enhance this dataset and obtain a more complete understanding of transcriptional regulation in barley, we plan to generate and analyze additional CAGE datasets, particularly from floral tissues, which are expected to reveal generative-tissue-specific regulatory mechanisms. Furthermore, we hypothesize that species-specific promoter models, constructed from a limited number of tissue-specific CAGE datasets, could facilitate genome-wide promoter prediction without necessitating the production of an exhaustive array of new datasets [68].

Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used chatGPT in order to correct English grammar and refine scientific style without changing text content. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

This project has been supported by the Czech Science Foundation [grant number 21–18794S] and by the European Regional Development Fund project “Plants as a tool for sustainable global development” [No. CZ.02.1.01/0.0/0.0/16_019/0000827].

Computational resources were provided by the e-INFRA CZ project (ID:90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

CRediT authorship contribution statement

P.N. and H.S.: Project conceptualization; P.N., S.P.: Investigation; S. N., B.L. and S.P.: Software design; S.P., P.N., and T.A. Formal analysis; M. K.: Resources (RNA-seq data); P.N., S.P.: Writing - Original Draft; P.N., H.S. and S.P.: Writing - Review & Editing. H.S.: Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Katerina Holusova and Helena Tvardikova for sequencing assistance and Zdenka Bursova for plant maintenance.

Code and data availability

CAGE data were deposited into the Gene Expression Omnibus (GEO) database under accession number GSE227219.

The data generated in the study will be shared in the EPD database (<https://epd.expasy.org/epd/>). The scripts, high-resolution figures and other large data files are available at: <https://github.com/MorexV3CAGE>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.csbj.2023.12.003](https://doi.org/10.1016/j.csbj.2023.12.003).

References

- Burley SK, Roeder RG. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* 1996;65:769–99.
- Bergman DT, Jones TR, Liu V, Ray J, Jagoda E, Siraj L, et al. Compatibility rules of human enhancer and promoter sequences. *Nature* 2022;607:176–84.
- Martinez-Ara M, Comoglio F, van Arensbergen J, van Steensel B. Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Mol Cell* 2022;82:2519–2531.e6.
- Neumayr C, Haberle V, Serebreni L, Karner K, Hendy O, Bojja A, et al. Differential cofactor dependencies define distinct types of human enhancers. *Nature* 2022;606:406–13.
- Haberle V, Lenhard B. Promoter architectures and developmental gene regulation. *Semin Cell Dev Biol* 2016;57:11–23.
- Vo Ngoc L, Wang Y-L, Kassavetis GA, Kadonaga JT. The punctilious RNA polymerase II core promoter. *Genes Dev* 2017;31:1289–301.
- Smale ST, Baltimore D. The “initiator” as a transcription control element. *Cell* 1989;103–13. [https://doi.org/10.1016/0092-8674\(89\)90176-1](https://doi.org/10.1016/0092-8674(89)90176-1).
- Yamamoto YY, Ichida H, Matsui M, Obokata J, Sakurai T, Satou M, et al. Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genom* 2007;8:67.
- Murray A, Mendieta JP, Vollmers C, Schmitz RJ. Simple and accurate transcriptional start site identification using Smar2C2 and examination of conserved promoter features. *Plant J* 2022;112:583–96.
- Bernard V, Brunaud V, Lecharny A. TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genom* 2010;11:166.
- Cordon-Obras C, Gomez-Liñan C, Torres-Rusillo S, Vidal-Cobo I, Lopez-Farfan D, Barroso-Del Jesus A, et al. Identification of sequence-specific promoters driving polycistronic transcription initiation by RNA polymerase II in trypanosomes. *Cell Rep* 2022;38:110221.
- Marbach-Bar N, Bahat A, Ashkenazi S, Golan-Mashiach M, Haimov O, Wu S-Y, et al. DTIE, a novel core promoter element that directs start site selection in TATA-less genes. *Nucleic Acids Res* 2016;44:1080–94.
- Danks GB, Navratilova P, Lenhard B, Thompson EM. Distinct core promoter codes drive transcription initiation at key developmental transitions in a marine chordate. *BMC Genom* 2018;19:164.
- Shao W, Alcantara SG-M, Zeitlinger J. Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing. *eLife* 2019. <https://doi.org/10.7554/eLife.41461>.
- Vo Ngoc L, Cassidy CJ, Huang CY, Duttke SHC, Kadonaga JT. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev* 2017;31:6–11.
- Menand B, Desnos T, Nussaume L, Berger F, Bouchez D, Meyer C, et al. Expression and disruption of the Arabidopsis TOR (target of rapamycin) gene. *Proc Natl Acad Sci USA* 2002;99:6422–7.
- Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, Kadonaga JT, et al. Human TFIID Binds to Core Promoter DNA in a Reorganized Structural State. *Cell* 2013;152:120–31.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 2006;38:626–35.
- Deal RB, Henikoff S. Histone variants and modifications in plant gene regulation. *Curr Opin Plant Biol* 2011;116–22. <https://doi.org/10.1016/j.cpb.2010.11.005>.
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WWMP, van Schaik FMA, Varier RA, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 2007;131:58–69.
- Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, Nepal C, et al. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* 2014;507:381–5.
- Mejía-Guerra MK, Li W, Galeano NF, Vidal M, Gray J, Doseff AI, et al. Core Promoter Plasticity Between Maize Tissues and Genotypes Contrasts with Predominance of Sharp Transcription Initiation Sites. *Plant Cell* 2015;27:3309–20.
- Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA* 2003;100:15776–81.
- Yamamoto YY, Yoshitsugu T, Sakurai T, Seki M, Shinozaki K, Obokata J. Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J* 2009;60:350–62.
- Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;322:1845–8.
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science* 2005;309:1559–63.
- FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest ARR, Kawaji H, Rehli M, Baillie JK, de Hoon MJL, et al. A promoter-level mammalian expression atlas. *Nature* 2014;507:462–70.
- Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, Sandelin A. Characterization of Promoter Bidirectionality and Antisense RNAs by Inactivation of Nuclear RNA Decay Pathways. *Plant Cell* 2020;32:1845–67.
- Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, et al. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nat Plants* 2021;7:842–55.
- Mascher M, Wicker T, Jenkins J, Plott C, Lux T, Koh CS, et al. Long-read sequence assembly: a technical evaluation in barley. *Plant Cell* 2021;33:1888–906.
- Navrátilová P, Toegelová H, Tulpová Z, Kuo Y-T, Stein N, Doležel J, et al. Prospects of telomere-to-telomere assembly in barley: Analysis of sequence gaps in the MorexV3 reference genome. *Plant Biotechnol J* 2022;20:1373–86.
- Schreiber M, Mascher M, Wright J, Padmarasu S, Himmelbach A, Heavens D, et al. A Genome Assembly of the Barley “Transformation Reference” Cultivar Golden Promise. *G3* 2020;10:1823–7.
- Kovacik M, Nowicka A, Pecinka A. Isolation of High Purity Tissues from Developing Barley Seeds. *J Vis Exp* 2020. <https://doi.org/10.3791/61681>.
- Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3.
- Philippe L, van den Elzen AMG, Watson MJ, Thoreen CC. Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5' TOP motifs. *Proc Natl Acad Sci* 2020;5319–28. <https://doi.org/10.1073/pnas.1912864117>.
- Wimalanathan K, Lawrence-Dill C.J. Gene Ontology Meta Annotator for Plants (GOMAP). [doi:10.1101/809988](https://doi.org/10.1101/809988).
- Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 2011;6:e21800.
- Neumann P, Navrátilová A, Schroeder-Reiter E, Koblížková A, Steinbauerová V, Chochořová E, et al. Stretching the rules: monocentric chromosomes with multiple centromere domains. *PLoS Genet* 2012;8:e1002777.
- Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017;12:2478–92.
- Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, et al. A chromosome conformation capture ordered sequence of the barley genome. *Nature* 2017;544:427–33.
- Kovacik M, Nowicka A, Zwyrtková J, Střejčková B, Vardanega I, Esteban E, et al. The transcriptome landscape of developing barley seeds reveals H3K27me3 dynamics in endosperm tissues. 2023.07.26.550659 bioRxiv 2023. <https://doi.org/10.1101/2023.07.26.550659>.
- Applied Research Applied Research Press. RSEM: Accurate Transcript Quantification from RNA-Seq Data with Or Without a Reference Genome. 2015.

- [43] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma* 2011;12:323.
- [44] Haberle V, Forrest ARR, Hayashizaki Y, Carninci P, Lenhard B. CAGER: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res* 2015;43:e51.
- [45] Levine M, Tjian R. Transcription regulation and animal diversity. *Nature* 2003;424:147–51.
- [46] Nikumbh S, Lenhard B. Identifying promoter sequence architectures via a chunking-based algorithm using non-negative matrix factorisation. *PLoS Comput Biol* 2023. <https://doi.org/10.1101/2023.03.02.530868>.
- [47] Cavin P erier R, Junier T, Bucher P. The Eukaryotic Promoter Database EPD. *Nucleic Acids Res* 1998;26:353–7.
- [48] Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 2011;27:1017–8.
- [49] Wragg JW, Roos L, Vucetic N, Cveticic N, Lenhard B, M uller F. Embryonic tissue differentiation is characterized by transitions in cell cycle dynamic-associated core promoter regulation. *Nucleic Acids Res* 2020;48:8374–92.
- [50] Santana-Garcia W, Castro-Mondragon JA, Padilla-G alvez M, Nguyen NTT, Elizondo-Salas A, Ksouri N, et al. RSAT 2022: regulatory sequence analysis tools. *Nucleic Acids Res* 2022. <https://doi.org/10.1093/nar/gkac312>.
- [51] Sebastian A, Contreras-Moreira B. footprintDB: a database of transcription factors with annotated cis elements and binding interfaces. *Bioinformatics* 2014;30:258–65.
- [52] Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, et al. Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat Genet* 2017;49:1546–52.
- [53] Pedersen DS, Coppens F, Ma L, Antosch M, Marktl B, Merkle T, et al. The plant-specific family of DNA-binding proteins containing three HMG-box domains interacts with mitotic and meiotic chromosomes. *N Phytol* 2011;192:577–89.
- [54] Milne L, Bayer M, Rapazote-Flores P, Mayer C-D, Waugh R, Simpson CG. EORNA, a barley gene and transcript abundance database. *Sci Data* 2021;8:90.
- [55] Boecker F. AHRD: Automatically annotate proteins with human readable descriptions and Gene Ontology terms. Universit ats- und Landesbibliothek Bonn. 2021. Available: <https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/9344>.
- [56] Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, Zhu J, et al. Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet* 2011;7:e1001274.
- [57] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–6.
- [58] Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature* 2012;489:101–8.
- [59] Wicker T, Schulman AH, Tanskanen J, Spannagl M, Twardziok S, Mascher M, et al. The repetitive landscape of the 5100 Mbp barley genome. *Mob DNA* 2017;8:22.
- [60] Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, et al. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol* 2006;7:1–18.
- [61] Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* 2018;19:621–37.
- [62] Seizl M, Hartmann H, Hoeg F, Kurth F, Martin DE, S oding J, et al. A Conserved GA Element in TATA-Less RNA Polymerase II Promoters. *PLoS One* 2011;6. <https://doi.org/10.1371/journal.pone.0027595>.
- [63] Voigt P, Tee W-W, Reinberg D. A double take on bivalent promoters. *Genes Dev* 2013;27:1318–38.
- [64] Kawaji H, Lizio M, Itoh M, Kanamori-Katayama M, Kaiho A, Nishiyori-Sueki H, et al. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Res* 2014;24:708–17.
- [65] Mercer TR, Dinger ME, Bracken CP, Kolle G, Szubert JM, Korbie DJ, et al. Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. *Genome Res* 2010;20:1639–50.
- [66] Haberman N, Digby H, Faraway R, Cheung R, Jobbins AM, Parr C, et al. Abundant capped RNAs are derived from mRNA cleavage at 3'UTR G-Quadruplexes. 2023.04.27.538568 *bioRxiv* 2023. <https://doi.org/10.1101/2023.04.27.538568>.
- [67] Qiu C, Jin H, Vvedenskaya I, Llenas JA, Zhao T, Malik I, et al. Universal promoter scanning by Pol II during transcription initiation in *Saccharomyces cerevisiae*. *Genome Biol* 2020;21:1–31.
- [68] Wang Y, Peng Q, Mou X, Wang X, Li H, Han T, et al. A successful hybrid deep learning model aiming at promoter identification. *BMC Bioinforma* 2022;23:1–20.

Appendix VII

Analysis of transcriptome landscape in developing barley seeds

Kovačik, M., Nowicka, A., and Pecinka, A.

In: Book of abstracts of the Plant Biotechnology: Green for Good V.

Olomouc, Czech Republic, 2019



Analysis of transcriptome landscape in developing barley seeds

Martin Kovacik, Anna Nowicka & Ales Pecinka



The Czech Academy of Sciences, Institute of Experimental Botany (IEB), Centre of the Region Haná for Biotechnological and Agricultural Research (CRH), Šlechtitelů 31, 783 71 Olomouc, Czech Republic

Introduction

The nuclear endosperm of monocots represents most important renewable source of food, feed, and industrial raw materials. Seed development is started by double-fertilization giving rise to the diploid (2n, 2C) embryo and triploid (3n, 3C) endosperm, respectively. Developing embryo and endosperm is surrounded by a diploid seed coat of maternal origin. In order to understand molecular and cellular mechanisms taking place in different seed tissues and during developmental stages, we performed transcriptomic analysis. Total RNA was isolated and sequenced from embryo, endosperm and seed coat nuclei at 4, 8, 16 and 24 days after pollination (DAP) to cover main events of endosperm development.



Dissection

Seed coats



At 4 DAP seed coats are full of water and make up most of the caryopsis. During development, are getting dry and at 24 DAP are mostly dried and stuck with endosperm and embryo.

Seed coats dissected at 4, 8, 16 and 24 DAP.

Endosperm



Development of the endosperm has four main stages: syncytium, cellularization, differentiation and maturation.

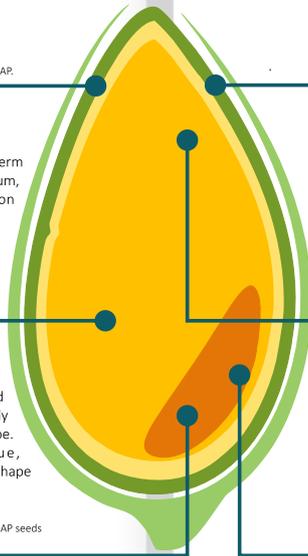
Endosperm dissected at 4, 8, 16 and 24 DAP seeds

Embryo



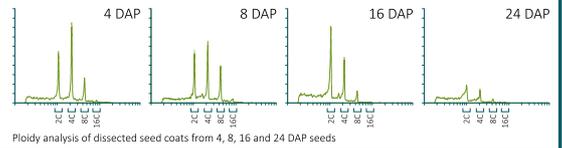
At early beginning of seed development, embryo is only 200 µm big and has round shape. As development continue, embryo is expanding and its shape become more oval.

Embryo dissected from 4, 8, 16 and 24 DAP seeds



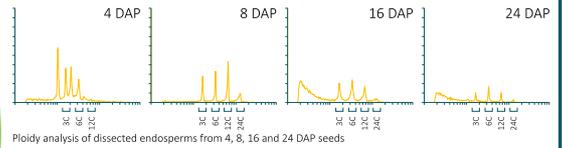
Quality control of dissected parts

Seed coats



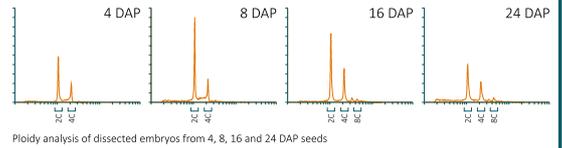
Ploidy analysis of dissected seed coats from 4, 8, 16 and 24 DAP seeds

Endosperm



Ploidy analysis of dissected endosperms from 4, 8, 16 and 24 DAP seeds

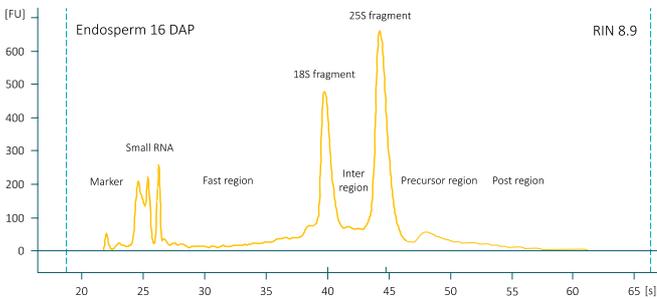
Embryo



Ploidy analysis of dissected embryos from 4, 8, 16 and 24 DAP seeds

Quality control of isolated RNA

RNA was isolated using RNeasy Plant Mini Kit (QIAGEN) and Spectrum Plant Total RNA Kit (Sigma).



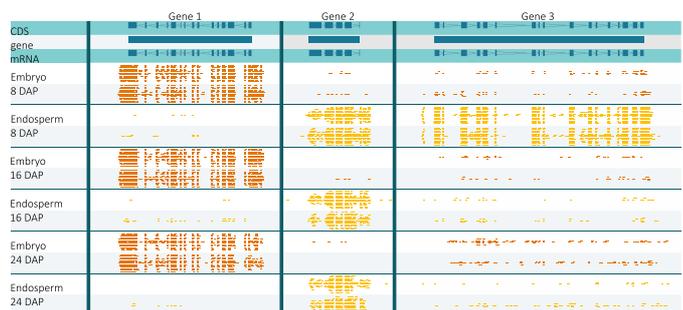
Agilent 2100 Bioanalyzer electropherogram summary

Mapping statistics



Average mapping statistics using HISAT2

Mapping visualisation



Examples of mapped reads on chromosome 4H

Current list of prepared samples

	4 DAP	8 DAP	16 DAP	24 DAP
Seed coats	Prepared	Prepared	Prepared	to be prepared
Endosperm	to be prepared	Prepared	Prepared	Prepared
Embryo	to be prepared	Prepared	Prepared	Prepared

Appendix VIII

Transcriptome landscape of endosperm in developing barley seeds

Kovačik, M., Nowicka, A., Vardanega, I., Provart, N., J., Simon, R., and Pecinka, A.

In: Book of abstracts of the 26th International Conference on Sexual Plant Reproduction

Prague, Czech Republic, 2022



Transcriptome landscape of endosperm in developing barley seeds

Martin Kovacik¹, Anna Nowicka¹, Isaiá Vardanega², Nicholas J. Provar², Rüdiger Simon² & Ales Pecinka¹

¹Institute of Experimental Botany, Czech Acad Sci, Centre of the Region Haná for Biotechnological and Agricultural Research, CZ-779 00 Olomouc, Czech Republic, ²Institute for Developmental Genetics, Heinrich-Heine-University Düsseldorf, Düsseldorf 40225, Germany, ³Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, 25 Wilcocks St., Toronto, ON M5S 3B2, Canada

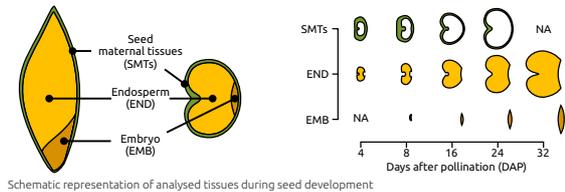


Introduction

Cereal grains are the major source of food and feed. The largest part of cereal seeds is occupied by triploid endosperm, representing a specialized tissue for embryo protection and nourishment. In order to understand molecular and cellular mechanisms governing cereal seed development, we used barley (*Hordeum vulgare*), which is a diploid temperate zone cereal crop. **To provide a spatiotemporal information about the seed developmental process, we performed transcriptomic study of endosperm, embryo and seed maternal tissues at early, middle and late stages of barley grain development.** Analysis of differential gene expression and co-expression networks pointed out to the major biological processes going on in different grain tissues at different times after fertilization, such as cellularization, differentiation and storage components synthesis. Furthermore, we defined a set of tissue-specific marker genes, which can be used to follow tissue origin and stage of seed development. This opens new avenues towards functional analysis of barley seed development.

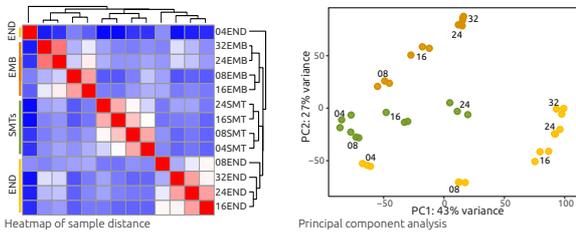
Analysed tissues and timepoints

Embryo, endosperm and seed maternal tissues were dissected at 4-32 days after pollination. RNA was isolated and sequenced using Illumina NovaSeq 6000, generating 100-bp SE reads.



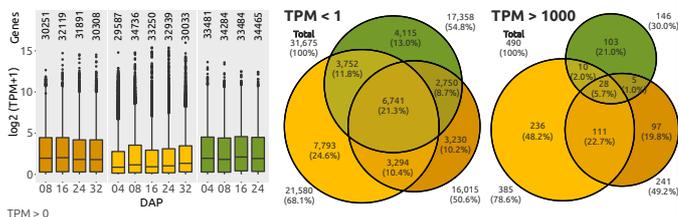
Seed tissue clusters

Heatmap of samples shows unique transcriptome of 4 DAP syncytial endosperm. Principal component analysis shows a clear separation of developmental timepoints (PC1) and tissue (PC2) indicating distinct transcript profiles activated in spatiotemporal manner.



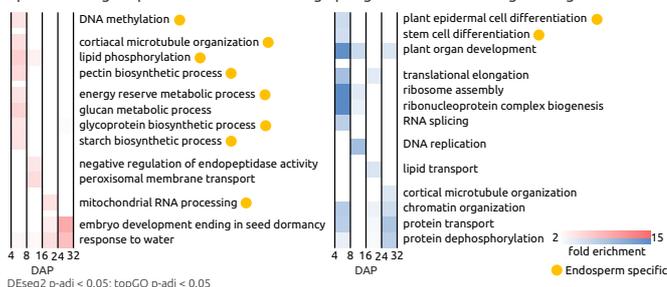
Extreme transcriptome of endosperm

High number of low expressed as well as high expressed genes is specific for endosperm.



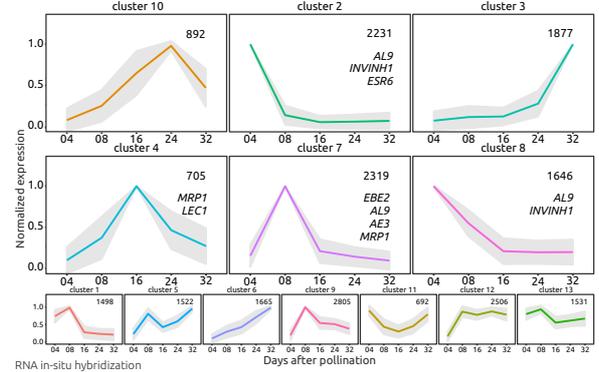
Functional analysis of DEGs

Specific biological processes enriched among up-regulated and down-regulated genes.



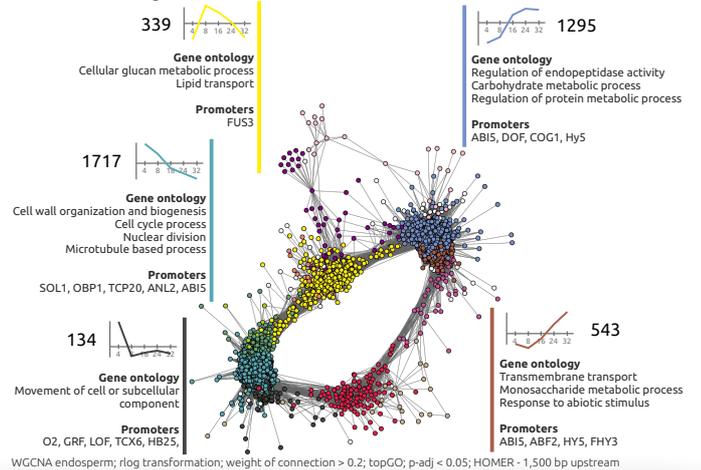
Gene clustering and marker gene selection

Clusters were investigated for known markers and could potentially serve as a source of new marker genes in the future.



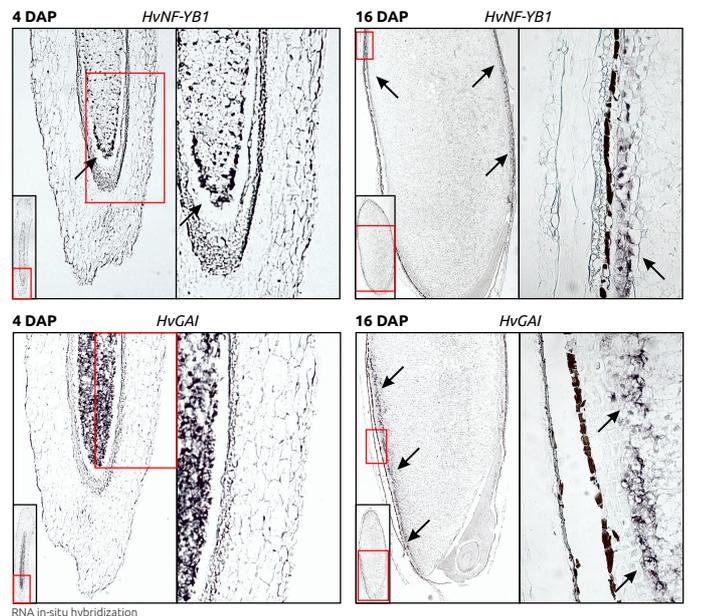
Possible relations, functions and regulators

Modules of highly correlated genes were clustered based on expression profile in endosperm, and related to each other. Gene ontology enrichment and motif enrichment provide information about their function and regulators.



Validation of endosperm marker expression

Transcription factor NF-YB is aleurone marker, with peak of expression at 16 DAP. DELLA protein GAI is marker for 4 DAP endosperm. Arabidopsis homolog restrain cell proliferation and expansion.



Palacký University Olomouc
Faculty of Science
Department of Cell Biology and Genetics



and

Institute of Experimental Botany of the Czech Academy of Sciences
Centre of Plant Structural and Functional Genomics



**Analysis of transcriptome dynamics in developing
barley seeds**

Summary of the Ph.D. Thesis

Martin Kovačik

P1527 – Molecular and Cellular Biology

Olomouc 2024

Supervisor: Assoc. Prof. Dr. Aleš Pečinka

Co-supervisor: Ing. Anna Nowicka Ph.D.

The Ph.D. thesis was carried out at the Department of Cellular Biology and Genetics, Faculty of Science, Palacký University Olomouc, between 2018 and 2024.

Candidate: **Mgr. Martin Kovačik**

Supervisor: **Assoc. prof. Aleš Pečinka, Ph.D.**

Reviewers: **doc. Ing. Petr Smýkal, Ph.D.**

Palacký University Olomouc, Olomouc, Czech Republic

Helene Robert Boisivon, Ph.D.

Central European Institute of Technology, Brno, Czech Republic

The evaluation of this Ph.D. thesis was written by

.....

The summary of the Ph.D. thesis was sent for distribution on

The oral defense will take place on in front of the commission for the Ph.D. study of the study program Molecular and Cellular Biology in the conference room at the Institute of Experimental Botany ASCR, Šlechtitelů 31,779 00 Olomouc – Holicе.

The Ph.D. thesis is available in the Library of the Biological Departments of the Faculty of Science at Palacký University Olomouc, Šlechtitelů 27, 779 00 Olomouc-Holicе.

Prof. RNDr. Zdeněk Dvořák, DrSc. et Ph.D.

Chairman of the Commission for the Ph.D. Thesis of the Study Program Molecular and Cellular Biology
Department of Cellular Biology and Genetics, Faculty of Science, Palacký University Olomouc

Table of content

1	Introduction.....	4
2	Aims of the thesis.....	5
3	Materials and methods	6
4	Summary of results	10
5	Summary	11
6	References.....	12
7	List of author's publications	15
8	Souhrn (Summary in Czech).....	17

1 Introduction

Cereal seeds represent an important source of food, feed, and industrial materials for humans and domestic animals, accounting for more than 60% of the global food resources. These seeds are the most valuable product of plant production. Botanically, seeds emerge from the fertilization of ovules. This process begins with the delivery of two haploid sperm cells by the pollen tube to the embryo sac (Peris et al., 2010). One sperm cell fertilizes the haploid egg cell, creating the diploid embryo, while the second sperm cell merges with the diploid central cell to produce the triploid endosperm (West and Harada, 1993; Goldberg et al., 1994). Following fertilization, the seed comprises three main parts: the embryo, endosperm, and the seed maternal tissues. While the embryo and endosperm are genetically related, they differ in their ploidy levels and the contribution of parental genomes. Specifically, the embryo carries one maternal and one paternal genome, whereas the endosperm includes two maternal and one paternal genome, leading to their distinct developmental trajectories (Brown et al., 1999; Kiesselbach, 1999; Chandler et al., 2008; Peris et al., 2010).

Barley (*Hordeum vulgare* L. subsp. *vulgare*) has emerged as a significant model organism for scientific research, providing insights into organ development, genetic diversity, and epigenetics (Baker et al., 2015; Jayakodi et al., 2020; Thiel et al., 2021; Hertig et al., 2023). Its diploid genome, and close evolutionary relationship with other cereal crops render it an excellent model for studying seed development and other fundamental biological processes.

Advances in high-throughput sequencing technologies have significantly boosted the potential for developmental studies of individual seed tissues. Nevertheless, the compact and adherent nature of these tissues poses challenges for their separation and analysis (Sreenivasulu et al., 2010). Understanding the regulatory mechanisms governing seed development in cereals is essential for plant breeding and yield enhancement. Although various studies have explored the transcriptional and chromatin dynamics in seed development in model systems such as *Arabidopsis* and maize, comprehensive information on these processes in barley has remained incomplete.

2 Aims of the thesis

2.1 Development of protocol for dissection of endosperm, embryo and seed maternal tissues during seed development of barley (*Hordeum vulgare*)

The first aim was to develop a protocol for the reproducible isolation of tissues from developing seeds. The first task involved determining the day of pollination (DOP) in self-pollinated spikes, due to the low success rate of manual pollination. The second task was to manually isolate whole seed tissues in high purity. Dissecting high-purity tissues poses a challenge due to their small size, compact structure, and tight adhesion of the seed tissues.

2.2 Transcriptome analysis of dissected tissues during barley seed development

The second aim of the thesis was to perform transcriptomic analysis of developing barley seeds using high-throughput sequencing. The initial step was to establish the bioinformatic pipeline. The subsequent step was determining transcripts per million reads (TPM), identifying significantly differentially expressed genes (DEGs) within the tissues, and cluster them according to their expression profile. The final step was focused on identifying regulatory motifs important for different tissues and developmental stages, thereby suggesting the involvement of specific groups of transcription factors.

2.3 Analysis of PRC2 complex expression, H3K27me3 distribution and identification of imprinted genes during barley seed development

The third aim of this thesis was to validate the activity of the PRC2 complex involved in the establishment of genomic imprinting and to analyze the deposition of the H3K27me3 modification across the genome. The following step was to compare transcriptomic and epigenomic data from different tissues to assess the direct impact of H3K27me3 modification on gene expression. The last objective was to identify potential imprinted genes through a comparative search for evolutionary conserved imprinted genes among cereals, followed by their validation using Sanger sequencing.

3 Materials and methods

3.1 Plant material and growth conditions

This study used six-rowed spring barley, *Hordeum vulgare* subsp. *vulgare* cv. Morex. The analysis of imprinted genes also included additional wild barley, *Hordeum vulgare* subsp. *spontaneum* strain HOR 12560. Seeds were germinated for 3 days at 25°C in darkness and then grown in soil under a long-day regime (16 h day 20°C, 8 h night 16°C; light intensity 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$; humidity 60%) until flowering. Days after pollination (DAP) were determined by identifying the day of self-pollination, based on anthers and stigma morphology. Developing seeds at 4, 8, 16, 24 and 32 DAP were harvested from the middle of the central row of the spikelets in at least three replicates. Embryo, endosperm and seed maternal tissues (SMTs) were manually dissected, their purity verified by flow-cytometric ploidy measurement as per Kovacik et al. (2020), and tissues from up to 10 seeds were pooled per biological replicate. The 4 DAP embryos and 32 DAP SMTs were excluded due to small size and dry nature, respectively. Dissected tissues were frozen in liquid nitrogen and stored at -80°C until use.

3.2 RNA extraction, sequencing, and data processing

For sequencing, total RNA was isolated using RNeasy® Plant Mini Kit (QIAGEN) and Spectrum™ Plant Total RNA Kit (Sigma-Aldrich). The RNA quality was checked using Bioanalyzer 2100 (Agilent) and samples with RNA integrity number > 6.8 were processed into library preparation using NEBNext® Ultra™ RNA Library Prep Kit (Illumina). The mRNA-enriched libraries were sequenced on Illumina HiSeq2500 as single-end 100 bp RNA-seq reads. The raw reads were trimmed for adaptor, aligned to the *H. vulgare* cv. Morex reference genome v3 (Mascher, 2021), and aligned reads were assigned to genes. Differential expression analysis, principal component analysis and Venn diagrams were prepared using DESeq2 v.1.24.0 (Love et al., 2014) and eulerr v.6.1.0 (Larsson, 2020) libraries in R software (R Core Team, 2020).

3.3 Clustering analyses

Unique DEGs from all tested combinations withing each tissue were clustered using k-means algorithm. For WGCNA, weighted gene co-expression network was constructed for each tissue using the read counts in R. Hierarchical clustering was performed for grouping

the genes with highly similar co-expression patterns. Each co-expression module was represented by expression profile of module eigengene. Both k-means clustering and WGCNA were done in R (Langfelder and Horvath, 2008; R Core Team, 2020).

3.4 A seed view in Barley ePlant

The Barley ePlant framework (Thiel et al., 2021) was modified to accept V3 Barley gene identifiers (Mascher, 2021). Transcriptomic data were databased on the Bio-Analytic Resource for Plant Biology (Toufighi et al., 2005) and linked with specific parts of the SVG image depicting the parts of the seed that were sampled.

3.5 GO term enrichment and annotation of transcription factors

GO enrichment analysis was performed using the topGO v.2.44.0 package in R (Alexa and Rahnenfuhrer, 2021), redundant GO terms were filtered using REVIGO (Supek et al., 2011) and general terms were filtered using size selection (Yon Rhee et al., 2008). Transcription factors were classified based on the presence of specific domains according to PlantTFDB (Jin et al., 2017).

3.6 Cis-motif identification and clustering

Sequences 1500 bp upstream from the predicted start codon of all WGCNA module genes were used for cis-motif identification and enrichment analysis. The analysis of known and *de novo* motifs was carried out using HOMER suite (Heinz et al., 2010). Collections of identified motifs in each WGCNA module were post-filtered for plant motifs and clustered using the RSAT (Castro-Mondragon et al., 2017).

3.7 RNA in situ hybridization

Barley seeds were fixed with 4% paraformaldehyde (w/v) for 1 h under vacuum. Vacuum was broken every 10 min and applied again. Subsequently seeds were transferred into fresh fixative and incubated overnight at 4°C, dehydrated using ethanol series, cleared by ROTIHistol series, and embedded into Paraplast. Longitudinal dorsoventral sections of 10 µm were cut with microtome and attached to adhesion slides. RNA probes were prepared by amplification of DNA probe from cDNA and subsequent amplification by T7 RNA polymerase in presence of DIG-UTP. Slides with sections were prepared for hybridization and hybridizations were carried out as described in Kovacik et al. (2024). After washing and digestion of unbound RNA, immunological detection was performed using DIG antibodies

coupled with alkaline phosphatase followed by staining procedure with 4-Nitro blue tetrazolium chloride and 5-Bromo-4-chloro-3-indolyl-phosphate. Hybridization signal was analyzed by light microscopy.

3.8 Identification of chromatin genes in barley

Subset of 47 *A. thaliana* genes encoding histones and 12 encoding PRC2 complex subunits were used for homology search using BLAST+ (Camacho et al., 2009). The resulting hits were confirmed with reciprocal homology search and candidates were further filtered by E-value (≤ 0.01) and additional parameters including gene length and alignment coverage of the hit.

3.9 ImmunoFISH, microscopy, and imagen analysis

Nuclei from 24 DAP embryos and 8 and 24 DAP endosperm were flow-sorted as described in Nowicka et al. (2023) and incubated with the rabbit anti-H3K27me3 primary antibody (1:200; Abcam, 195477) at 4°C overnight and secondary goat anti-rabbit-Alexa Fluor 488 antibody (1:300, Molecular Probes, A11008) at 37°C for 90 min. Barley centromeres were detected with a synthetic 28-mer oligonucleotide (5'-AGGGAGA-3')₄ *CEREBA* probe labeled at the 5' end with Cy3 (Eurofins). Telomeres were visualized with a synthetic 28-mer oligonucleotide probe (5'-CCCTAAA-3')₄ labeled at the 5' end with Cy5 (Eurofins). The images were acquired with an AxioImager Z2 microscope (Zeiss, Oberkochen, Germany) equipped with a pE-4000 LED illuminator light source (CoolLED), laser-free confocal spinning disk device (DSD2, Andor, Belfast, UK) and with $\times 100/1.4$ NA Oil M27 Plan-Apochromat (Zeiss) objective. Image stacks of 40-80 slides depending from the C-value of the nucleus, on average, with 0.2 μm z-step were taken separately for each fluorochrome using the appropriate excitation and emission filters. The images were saved as maximum intensity projection and the nucleus surface, immunosignals, and centromere and telomere 3D modeling were generated using Imaris 9.7 software. Fluorescence intensity measurements were performed in FIJI.

3.10 Analysis of imprinted genes

Cereal imprinted genes were extracted from published works (Luo et al., 2011; Waters et al., 2011; Zhang et al., 2011; Chen et al., 2018; Yang et al., 2018) and their overlaps were visualized using Venn diagrams in R. The Morex and HOR 12560 strains were reciprocally crossed, 8 DAP endosperm was manually dissected as described in Kovacik et al. (2020),

total RNA was isolated using Spectrum™ Plant Total RNA Kit (Sigma-Aldrich), and reverse transcribed into cDNA using RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). The PCR amplicons of 200–1,000 bp harboring informative SNPs were gel purified using GeneJet Gel Extraction Kit (Thermo Fisher Scientific) and subjected to Sanger sequencing followed by *in silico* analysis using SnapGene v6.2 software (GSL Biotech LLC).

3.11 H3K27me3 ChIP-seq and data analysis

Approximately 2 g of 16 DAP endosperm tissue from cv. Morex were cross-linked in 1% (w/v) formaldehyde under vacuum, quenched by glycine, washed, and frozen in liquid nitrogen. ChIP followed the method by Gendrel et al. (2005), with the following modifications. Isolated nuclei were resuspended in Nuclei lysis buffer, incubated at 4°C for 20 min, and sonicated using a Biorupter Plus (Diagenode) at 4°C. Sheared chromatin was diluted 1:4 with ChIP dilution buffer and 600 µL aliquots of diluted chromatin were incubated with 7 µL of the rabbit anti-trimethyl-Histone H3 (Lys27) antibody (Millipore, 07-449) in a rotator at 4°C overnight. Samples without antibody were used as negative controls. The next day, 40 µL of the Dynabeads Protein A (Invitrogen) were added to each tube, and the samples were further incubated for 2 hours, followed by washing and elution of immune complexes. Control chromatin aliquots ('input DNA') were taken prior to immunoprecipitation. Reverse crosslinking was performed for all samples, and DNA was extracted and purified using the ChIP DNA Clean & Concentrator™ kit (Zymo Research). Sequencing libraries were prepared and 150 bp pair-end reads were sequenced using Illumina NovaSeq 6000 platform (Illumina) by Novogene.

ChIP-seq data from our study and H3K27me3 ChIP-seq data from Baker et al. (2015) were adaptor-trimmed, aligned to Morex reference genome, and deduplicated with GATK (der Auwera and O'Connor, 2020). MACS2 (Zhang et al., 2008) performed peak calling; peaks with fold enrichment ≥ 5 were analyzed for differential signal intensity in endosperm vs. shoot using MAnorm2 (Tu et al., 2021). Testing was performed in genomic intervals of 2000 bp and intervals with differential signal intensities localized in coding regions or 2000 bp upstream were related to genes. GO term enrichment analysis used g:Profiler (Raudvere et al., 2019) barley Morex V3 GO annotation from Ensembl Plants.

4 Summary of results

4.1 Development of protocol for dissection of endosperm, embryo and seed maternal tissues during seed development of barley (*Hordeum vulgare*)

The developed protocol has enabled the isolation of high-purity barley seed tissues, validated through rigorous quality control analysis using flow cytometry. Its effectiveness has been proven across several two- and six-row spring barley cultivars, and the protocol can be easily adapted for use with other members of the *Triticeae* tribe, such as wheat, oat, rye, and triticale.

4.2 Transcriptome analysis of dissected tissues during barley seed development

The transcriptome atlas we have created, using the current version of the barley genome assembly and annotation, provided a solid base for subsequent biological investigations into the key factors involved in barley grain development. Although our study primarily focused on endosperm tissues, the dataset offers equal resolution also for the embryo and SMTs. We have investigated the expression of transcription factors and the presence of their regulatory motifs among the expressed genes, suggesting the G-box and P-box motifs as essential for barley grain development. Through reciprocal BLAST, we have identified barley homologs of well-known marker genes for individual endosperm domains from other cereals and confirmed the specificity of these markers in barley.

4.3 Analysis of PRC2 complex expression, H3K27me3 distribution and identification of imprinted genes during barley seed development

Within our last aim, we delved into the role of epigenetic processes and molecular factors in seed development, specifically examining the expression of histones, subunits of the Polycomb Repressive Complex 2 (PRC2), and the genome-wide distribution of the H3K27me3 modification. We observed that several PRC2 complex components were expressed at low levels, which corresponded with a significant reduction in H3K27me3 modification, observable both cytologically and molecularly. Furthermore, we conducted a basic comparative analysis to identify evolutionarily conserved imprinted genes in barley. The discovery of a relatively low number of imprinted genes shared by at least two major cereal species suggests that imprinted genes in cereals have undergone rapid evolution.

5 Summary

The thesis presented here focuses on the spatiotemporal transcriptome analysis of dissected barley seed tissues, generating valuable data for functional research into barley grain development. It provides several unique resources that will significantly advance the field of barley genomic research.

This thesis provides description of the protocol for the isolation of high-purity barley seed tissues, validated through rigorous quality control analysis using a flow cytometer. The application of this protocol facilitated the development of a comprehensive gene expression atlas for developing barley seeds, offering high-resolution insights into the embryo, endosperm, and seed maternal tissues across five developmental stages.

Analysis of this dataset has revealed a list of tissue-specific genes that have the potential to serve as developmental markers for individual seed tissues, and transcription factors critical for developmental transitions. Particular emphasis was placed on marker genes for specific endosperm domains, which are instrumental for studying endosperm differentiation. Moreover, presented work shed light on the role of epigenetic processes and molecular factors in seed development, including the expression of PRC2 complex subunits and the genome-wide distribution of the H3K27me3 modification. Finally, this study laid down the basis for future analyses focused on genomic imprinting through the identification of evolutionarily conserved imprinted genes in barley.

6 References

- Alexa, A. and Rahnenfuhrer, J.** (2021). topGO: Enrichment analysis for gene ontology.
- der Auwera, G.A. and O'Connor, B.D.** (2020). Genomics in the cloud: using Docker, GATK, and WDL in Terra (O'Reilly Media).
- Baker, K., Dhillon, T., Colas, I., Cook, N., Milne, I., Milne, L., Bayer, M., and Flavell, A.J.** (2015). Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J* **84**: 111–124.
- Brown, R.C., Lemmon, B.E., Nguyen, H., and Olsen, O.-A.** (1999). Development of endosperm in *Arabidopsis thaliana*. *Sex Plant Reprod* **12**: 32–42.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L.** (2009). BLAST+: Architecture and applications. *BMC Bioinformatics* **10**: 1–9.
- Castro-Mondragon, J.A., Jaeger, S., Thieffry, D., Thomas-Chollier, M., and van Helden, J.** (2017). RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res* **45**: e119–e119.
- Chandler, J., Nardmann, J., and Werr, W.** (2008). Plant development revolves around axes. *Trends Plant Sci* **13**: 78–84.
- Chen, C. et al.** (2018). Characterization of imprinted genes in rice reveals conservation of regulation and imprinting with other plant species. *Plant Physiol* **177**: 1754–1771.
- Gendrel, A.V., Lippman, Z., Martienssen, R., and Colot, V.** (2005). Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat Methods* **2**: 213–218.
- Goldberg, R.B., De Paiva, G., and Yadegari, R.** (1994). Plant embryogenesis: zygote to seed. *Science* **266**: 605–614.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K.** (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576.
- Hertig, C., Rutten, T., Melzer, M., Schippers, J.H.M., and Thiel, J.** (2023). Dissection of Developmental Programs and Regulatory Modules Directing Endosperm Transfer Cell and Aleurone Identity in the Syncytial Endosperm of Barley. *Plants* **12**.

- Jayakodi, M. et al.** (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**: 284–289.
- Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G.** (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* **45**: D1040–D1045.
- Kiesselbach, T.A.** (1999). The structure and reproduction of corn (Cold spring harbor laboratory press).
- Kovacik, M. et al.** (2024). The transcriptome landscape of developing barley seeds. *Plant Cell*: in press.
- Kovacik, M., Nowicka, A., and Pecinka, A.** (2020). Isolation of high purity tissues from developing barley seeds. *J Vis Exp*: e61681.
- Langfelder, P. and Horvath, S.** (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*: 559.
- Larsson, J.** (2020). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses.
- Love, M.I., Huber, W., and Anders, S.** (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
- Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A.** (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet* **7**: e1002125.
- Mascher, M.** (2021). Pseudomolecules and annotation of the third version of the reference genome sequence assembly of barley cv. Morex [Morex V3].
- Nowicka, A., Ferková, L., Said, M., Kovacik, M., Zwyrtková, J., Baroux, C., and Pecinka, A.** (2023). Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues. *J Exp Bot* **74**: 2527–2541.
- Peris, C.I.L., Rademacher, E.H., and Weijers, D.** (2010). Green Beginnings – Pattern Formation in the Early Plant Embryo. In *Curr Top Dev Biol* (Elsevier Inc.), pp. 1–27.
- R Core Team** (2020). R: A language and environment for statistical computing.
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J.** (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**: W191–W198.
- Sreenivasulu, N., Borisjuk, L., Junker, B.H., Mock, H.-P., Rolletschek, H., Seiffert, U., Weschke, W., and Wobus, U.** (2010). Barley Grain Development. In *Int Rev Cell Mol Biol*, pp. 49–89.

- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T.** (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**: e21800.
- Thiel, J. et al.** (2021). Transcriptional landscapes of floral meristems in barley. *Sci Adv* **7**: eabf0832.
- Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J.** (2005). The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *Plant J* **43**: 153–163.
- Tu, S., Li, M., Chen, H., Tan, F., Xu, J., Waxman, D.J., Zhang, Y., and Shao, Z.** (2021). MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Res* **31**: 131–145.
- Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M.** (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**: 4221–4233.
- West, M.A.L. and Harada, J.J.** (1993). Embryogenesis in higher plants: An overview. *Plant Cell* **5**: 1361–1369.
- Yang, G., Liu, Z., Gao, L., Yu, K., Feng, M., Yao, Y., Peng, H., Hu, Z., Sun, Q., Ni, Z., and Xin, M.** (2018). Genomic imprinting was evolutionarily conserved during wheat polyploidization. *Plant Cell* **30**: 37–47.
- Yon Rhee, S., Wood, V., Dolinski, K., and Draghici, S.** (2008). Use and misuse of the gene ontology annotations. *Nat Rev Genet* **9**: 509–515.
- Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Zhao, H., Guan, H., Hu, X., Jiao, Y., Song, W., and Lai, J.** (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A* **108**: 20042–20047.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Shirley, X.S.** (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: 1–9.

7 List of author's publications

7.1 Original publications

Kovacik, M., Nowicka, A., & Pecinka, A. (2020). Isolation of high purity tissues from developing barley seeds. *JoVE (Journal of Visualized Experiments)*, (164), e61681.

Kovacik, M., Nowicka, A., Zwyrtková, J., Strejčková, B., Vardanega, I., Esteban, E., Pasha, A., Kaduchová, K., Krautsova, M., Červenková, M., Šafář, J., Provart, N., J., Simon, R., & Pecinka, A. (2024). The transcriptome landscape of developing barley seeds. *The Plant Cell*, in press.

7.2 Co-author publications

Nowicka, A., **Kovacik, M.**, Tokarz, B., Vrána, J., Zhang, Y., Weigt, D., Dolezel, J., & Pecinka, A. (2021). Dynamics of endoreduplication in developing barley seeds. *Journal of Experimental Botany*, 72(2), 268-282.

Nowicka, A., Sahu, P. P., **Kovacik, M.**, Weigt, D., Tokarz, B., Krugman, T., & Pecinka, A. (2021). Endopolyploidy variation in wild barley seeds across environmental gradients in Israel. *Genes*, 12(5), 711.

Nowicka, A., Ferková, L., Said, M., **Kovacik, M.**, Zwyrtková, J., Baroux, C., & Pecinka, A. (2023). Non-Rabl chromosome organization in endoreduplicated nuclei of barley embryo and endosperm tissues. *Journal of Experimental Botany*, 74(8), 2527-2541.

Pavlu, S., Nikumbh, S., **Kovacik, M.**, An, T., Lenhard, B., Simkova, H., & Navratilova, P. (2024). Core promoterome of barley embryo. *Computational and Structural Biotechnology Journal*, 23, 264-277.

7.3 Published abstracts

Kovačik, M., Nowicka, A., Pecinka, A., 2019. Analysis of transcriptome landscape in developing barley seeds. In: Book of abstracts of the Plant Biotechnology: Green for Good V., Olomouc, Czech Republic, June 10–13, poster 23.

Kovačik, M., Nowicka, A., Vardanega, I., Provart, N. J., Simon, R., Pecinka, A., 2019. Transcriptome landscape of endosperm in developing barley seeds. In: Book of abstracts of the 26th International Conference on Sexual Plant Reproduction, Prague, Czech Republic, 2022, June 20–24, poster 203.

Kovačik, M., Nowicka, A., Vardanega, I., Zwyrkova, J., Provart, N. J., Simon, R., Pecinka, A., 2019. Developing an atlas of gene expression during barley grain development. In: Book of abstracts of the 13th International Barley Genetics Symposium (IBGS 13), Riga, Latvia, 2022, July 3–7, talk.

8 Souhrn (Summary in Czech)

Tato práce se zaměřuje na časoprostorovou analýzu transkriptomu disektovaných pletiv ze semen ječmene, čímž generuje cenná data pro funkční analýzu vývoje ječmenného zrna. Poskytuje několik unikátních zdrojů, které významně posunou oblast výzkumu genomu ječmene.

Tato disertační práce popisuje protokol pro izolaci pletiv vysoké čistoty ze semen ječmene, ověřený prostřednictvím přísné kontroly kvality analýzou pomocí průtokového cytometru. Použití tohoto protokolu usnadnilo vývoj komplexního atlasu genové exprese pro vyvíjející se ječmenná zrna, který nabízí detailní vhledy do embrya, endospermu a obalových pletiv zrna v pěti vývojových stádiích.

Analýza tohoto datasetu odhalila seznam pletivově specifických genů, které mají potenciál sloužit jako vývojové markery pro jednotlivá pletiva, a transkripční faktory klíčové pro vývojové přechody. Zvláštní důraz byl kladen na genové markery pro specifické domény endospermu, které jsou nástrojem pro studium jeho diferenciaci. Navíc, prezentovaná práce osvětlila roli epigenetických procesů a molekulárních faktorů ve vývoji zrna, včetně exprese Polycomb represivního komplexu 2 a genomové distribuce histonové modifikace H3K27me3. Závěrem, tato studie položila základy pro budoucí analýzy zaměřené na genomový imprinting skrze identifikaci evolučně konzervovaných imprintovaných genů v ječmeni.