Environmental and genetic factors affecting endosperm-based post-zygotic hybridization barriers

Renate Marie Alling

Dissertation for the degree of Philosophiae Doctor (PhD)

Section for Genetics and Evolutionary Biology (EVOGENE)

Department of Biosciences

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

2024

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2729

ISSN 1501-7710

Cover: UiO.
Print production: Graphic center, University of Oslo.

ACKNOWLEDGEMENTS

This thesis has been performed under the supervision of Professor Anne Krag Brysting and Professor Paul E. Grini, with co-supervision of Dr. Katrine N. Bjerkan and Dr. Anders Kristian Krabberød.

Anne and Paul have provided me with invaluable perspectives on my projects, offering dual insights, making me feel like Hannah Montana – experiencing the "Best of both worlds". Anders has generously shared his expertise in bioinformatics and patiently guiding me through UNIX coding. Katrine played an important role in the initial phase of my projects, contributing with her experience and knowledge, especially on how to work on impossible arctic mustards. Collectively, my supervisors have given me valuable lessons throughout my PhD, allowing me to grow as a scientist.

Special thanks to Jason R. Miller, Yuri van Ekelenburg, and Karina S. Hornslien for their assistance with the Informative Reads Pipeline. Ingrid Johansen and Marit Langrekken deserve commendation for handling my unique plant requirements and challenging requests.

Working at EVOGENE has been a very smooth experience, much thanks to our great staff, Chiara, Roy, Vegard and Cecilie. I am also very happy to have shared coffee breaks, lunch and engaging discussions with my friends and colleagues, Verena, Sergio, Anne, Yuri, Ida and Roberto.

I also want to thank my family and friends. Sissel and Julien, in particular, who have supported me throughout and been lifting me up when needed. And my family with much support throughout my never-ending education.

Finally, I would like to thank my partner Jeremias. Your boundless support, willingness to indulge in discussions about my projects and constant role as the devil's advocate, have given me comfort and allowed me to grow as a person. I will always be the Charon to your Pluto.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **I**

TABLE OF CONTENTS ..... **III**

LIST OF ABBREVIATIONS ..... **V**

LIST OF PAPERS ..... **VII**

ABSTRACT ..... **IX**

SAMMENDRAG ..... **XI**

INTRODUCTION **1**

1. SEED DEVELOPMENT AND MODEL SYSTEMS IN BRASSICACEAE 2
1.1 REPRODUCTION AND SEED DEVELOPMENT 2
1.2 PHYLOGENY AND IMPORTANCE OF THE BRASSICACEAE FAMILY 5
1.3 THE MODEL SYSTEM *ARABIDOPSIS* 7
1.4 THE MODEL SYSTEM *DRABA* 7

2. HYBRIDIZATION BARRIERS IN BRASSICACEAE 8
2.1 SPECIATION DRIVEN BY HYBRIDIZATION 8
2.2 HYBRIDIZATION IN *DRABA* 9
2.3 HYBRIDIZATION IN *ARABIDOPSIS* 10
2.4 THE ENDOSPERM-BASED POST-ZYGOTIC HYBRIDIZATION BARRIER 10
2.5 GENETIC FACTORS AFFECTING THE EPZB 11
2.6 TEMPERATURE EFFECT ON THE EPZB 12

3. EVOLUTIONARY ORIGIN AND ESTABLISHMENT OF GENOMIC IMPRINTING 13
3.1 EVOLUTIONARY ORIGIN 13
3.2 ESTABLISHING GENOMIC IMPRINTING 14
3.3 GENOMIC IMPRINTING IN ANGIOSPERMS 15
3.4 MADS-BOX TRANSCRIPTION FACTORS AND GENOMIC IMPRINTING 16

TABLE OF CONTENTS

AIM OF STUDY 19

RESULTS AND DISCUSSION 21

4. ANTAGONISTIC ENDOSPERM PHENOTYPES IN *ARABIDOPSIS* HYBRIDS AND ABSENCE OF EPZB IN *DRABA* 21

5. THE GENETIC NETWORK CAUSING AN EPZB IS EASILY AFFECTED BY USING DIFFERENT ACCESSIONS 25

6. SINGLE-GENE MUTATIONS AFFECTING THE HYBRID BARRIER 26

7. TEMPERATURE HAS A STRONG ANTAGONISTIC EFFECT AND POTENTIAL AS AN EPZB-ADJUSTING FACTOR 27

8. BOTH ENVIRONMENTAL AND GENETIC FACTORS AFFECT THE HYBRID BARRIER 30

9. MADS-BOX TYPE-I EXPRESSION LEVELS AT DIFFERENT DEVELOPMENTAL TIME POINTS MAY PREDICT EPZBS 32

10. GENOMIC IMPRINTING IN *DRABA NIVALIS* IS DOMINATED BY MATERNALLY EXPRESSED GENES 36

CONCLUDING REMARKS AND FUTURE PERSPECTIVES 41

REFERENCES 46

PAPERS I - III 65

LIST OF ABBREVIATIONS

A. alpina	*Arabis alpina*
A. arenosa	*Arabidopsis arenosa*
A. halleri	*Arabidopsis halleri*
A. kamchatica	*Arabidopsis kamchatica*
A. lyrata	*Arabidopsis lyrata*
A. suecica	*Arabidopsis suecica*
A. thaliana	*Arabidopsis thaliana*
ac	antipodal cells
AGL	*AGAMOUS-LIKE*
B. carinata	*Brassica carinata*
B. juncae	*Brassica juncae*
B. napus	*Brassica napus*
B. oleracea	*Brassica oleraceae*
B. rapa	*Brassica rapa*
BEG	Biparentally Expressed Gene
bp	Base pair
C	Cytosine
C. bursa-pastoris	*Capsella bursa-pastoris*
C. grandiflora	*Capsella grandiflora*
C. rubella	*Capsella rubella*
cc	central cell
CDKA;1	*CYCLIN DEPENDENT KINASE A;1*
CH3	methyl
CMT3	*CHROMOMETYLASE 3*
Col-0	Columbia 0
D. alpina	*Draba alpina*
D. fladnizensis	*Draba fladnizensis*
D. nivalis	*Draba nivalis*
D. subcapitata	*Draba subcapitata*
DAP	Days After Pollination
DME	*DEMETER*
DRM2	*DOMAINS REARRANGED METHYLTRANSFERASE 2*
e.g.	Exempli gratia, for example
EBN	Endosperm Balance Number
ec	egg cell
EE	Early Endosperm
ePZB	Endosperm-based post-zygotic hybridization barrier
FC	Fold Change
FIE	*FERTILIZATION INDEPENDENT ENDOSPERM*
FIS-PRC2	*FERTILIZATION INDEPENDENT SEED POLYCOMB COMPLEX 2*
G	Guanine

LIST OF ABBREVIATIONS

GFP	Green Fluorescent Protein
GO	Gene Ontology
GSC	General Seed Coat
H	adenine, cytosine, thymine
H. longifolia	*Heliophila longifolia*
H. matronalis	*Hesperis matronalis*
H2A	Histone 2 A
H3K27me3	Trimethylation of the 27th amino acid in the tail of histone 3
HDG	*HOMEOBOX-LEUCINE ZIPPER PROTEIN*
InDel	Insertion/Deletion
IRP	Informative Reads Pipeline
LEC1	*LEAFY COTYLEDON 1*
m	Maternal
M	MADS-box domain
MADS-box	MINICHROMOSOME MAINTENANCE 1, AGAMOUS, DEFICIENS and SERUM RESPONSE FACTOR
mbp	megabases
MEA	*MEDEA*
MEG	Maternally expressed genes
MET1	*METHYLTRANSFERASE 1*
MSI1	*MULTICOPY SUPRESSOR OF IRA1*
mya	Million years ago
p	Paternal
PEG	Paternally expressed genes
PHE1	*PHERES1*
PIF4	*PHYTOCHROME INTERACTING FACTOR 4*
S. alba	*Sinapis alba*
sc	synergid cell
SNP	SINGLE NUCLEOTIDE POLYMORPHISM
TE1	Total Endosperm 1
TFs	Transcription Factors
UPR	Unfolded Protein Response
Wa-1	Warschau 1
WGD	Whole Genome Duplication
WGT	Whole Genome Triplication
WISO	weak inbreeder/strong outbreeder
Ws-2	Wassilewskija 2
WT	Wild-Type

LIST OF PAPERS

Paper I

Structural evidence for MADS-box type I family expansion seen in new assemblies of *Arabidopsis arenosa* and *A. lyrata*

Bramsiepe J., Krabberød A.K., Bjerkan K.N., **Alling R.M.**, Johannessen I.M., Hornslien K.S., Miller J.R., Brysting A.K. and Grini P.E.

The Plant Journal: 116(3), 942–961. (2023). doi.org/10.1111/tpj.16401

Paper II

Genetic and environmental manipulation of *Arabidopsis* hybridization barriers uncovers antagonistic functions in endosperm cellularization

Bjerkan K.N., **Alling R.M.**, Myking I.V., Brysting A.K. and Grini P.E.

Shared first authorship

Frontiers in Plant Science, 14. (2023). doi.org/10.3389/fpls.2023.1229060

Paper III

Low parental conflict, no endosperm hybrid barriers, and maternal bias in genomic imprinting in selfing *Draba* species

Alling R.M., Bjerkan K.N., Bramsiepe J., Nowak M.D, Gustafsson A.L.S, Brochmann C., Brysting A.K. and Grini P.E.

bioRxiv, (2024) doi.org/10.1101/2024.01.08.574548

ABSTRACT

The seed of flowering plants (angiosperms) is divided into three main compartments, the seed coat, the embryo, and the endosperm. The mature endosperm can be compared to the placenta in humans, because it delivers nutrients to the embryo. Initially the syncytial endosperm works as a nutrient sink and transition into being a nutrient source requires cellularization of the endosperm nuclei. A shift in the endosperm cellularization time point has been demonstrated in certain hybrid seeds, causing an incomplete post-zygotic hybridization barrier. This shift can either cause delayed or precocious endosperm cellularization, leading to seed lethality. The genetic and molecular basis of this hybridization barrier is often explained by an unbalanced genomic ratio in the endosperm. Such imbalance can be caused by crosses between different ploidies, by differences in genomic imprinting, or by specific genetic loci.

In this thesis, we assembled the genomes of *Arabidopsis arenosa* and *A. lyrata* ssp. *petraea*, and used them to examine differences in genomic structure and genetic regulation in relation to *A. thaliana*. We showed that expression of MADS-box type-I genes in the seed during the transition to endosperm cellularization differed significantly, suggesting that the role of MADS-box type-I genes is not completely conserved between the three species (Paper I). We further examined the endosperm-based post-zygotic hybridization barrier in two specific *Arabidopsis* hybrids. Using *A. thaliana* as the maternal contributor and *A. arenosa* or *A. lyrata* ssp. *petraea* as pollen donors, the hybrid seed phenotypes displayed diametrically opposed cellularization phenotypes, with delayed and precocious endosperm cellularization, respectively. We exposed the crossed plants to a temperature gradient and mutated specific MADS-box type-I genes, revealing that both mutation of *agamous-like 35* (*agl35*) in *A. thaliana* and temperature have antagonistic effects in the two hybrids (Paper II). Furthermore, we expanded the investigation of the endosperm-based post-zygotic hybridization barrier into the less studied genus *Draba*. We suggest that low parental conflict in the three selfing species *D. nivalis*, *D. fladnizensis*, and *D. subcapitata* may explain the absence of developmental problems in hybrid seeds. In addition, a genomic imprinting study in *D. nivalis* resulted in numerous maternally expressed genes, which were enriched for biological processes that have previously been suggested to represent adaptation to the harsh arctic environment for this species (Paper III).

In sum, these papers unravel a complex basis for endosperm-based post-zygotic hybridization barriers, which seem to be affected by different genetic factors such as accessions, single-gene mutations and differences in the regulation of MADS-box type I genes. Our results also indicate that the barrier depends on mating system (outcrossing versus self-fertilization) and can be manipulated by environmental factors such as temperature. Altogether, these findings provide a good tool kit for studying and manipulating the endosperm-based post-zygotic hybridization barrier.

SAMMENDRAG

I blomsterplanter (angiospermer) er frøet delt inn i tre ulike vev, frøskall, embryo og endosperm (frøhvite). Ferdigutviklet endosperm kan sammenlignes med morkaken hos mennesker, fordi den leverer næring til embryoet. I startfasen av endospermutviklingen er cellekjernene ikke separert i egne celler (syncytium) og endospermen fungerer som et næringslager. Overgangen til å være en næringskilde krever cellularisering av endosperm-cellekjernene. Tidspunktet for endosperm-cellularisering blir i noen hybridfrø forskjøvet, slik at cellularisering enten skjer for tidlig eller for sent, som igjen fører til at frøet dør. Den genetiske og molekylære basisen for hybridiseringsbarrieren som forårsaker forskyvning av tidspunktet for endosperm-cellularisering forklares oftest med en ubalansert genomisk ratio i endospermen. Denne ubalansen kan forårsakes ved at artene som krysses har ulikt antall kromosomsett, forskjeller i genomisk imprinting eller av spesifikke genetiske loci.

I denne avhandlingen har vi sekvensert og sammenstilt genomene til *Arabidopsis arenosa* og *A. lyrata* ssp. *petraea*, og brukt disse til å undersøke forskjeller i genomstruktur og genetisk regulering sammenlignet med *A. thaliana*. Vi viser at visse MADS-boks type-I gener er uttrykt forskjellig under endosperm-cellularisering, noe som antyder at rollen til MADS-boks type-I gener mellom de tre artene ikke er fullstendig konservert (Artikkel I). Vi undersøkte også den endosperm baserte post-zygotiske hybridiserings-barrieren i to spesifikke *Arabidopsis* hybrider. Ved å krysse *A. thaliana* som mor og *A. arenosa* eller *A. lyrata* ssp. *petraea* som far, fikk vi to motsatte cellulariserings-fenotyper, med henholdsvis forsinket eller for tidlig cellularisering av endosperm. I tillegg til å karakterisere hybridfrøene til planter som har vokst under normale forhold, undersøkte vi også hybridfrø fra planter som var utsatt for en temperaturgradient og hybridfrø generert fra maternelle *A. thaliana* planter med enkeltmutasjoner i gener fra MADS-boks type-I familien. Dette avslørte at både *agamous-like 35* (*agl35)*-mutasjonen i *A. thaliana* og økt temperatur har motsatte effekter på de to hybridene (Artikkel II). Vi utvidet undersøkelsen av endosperm-baserte post-zygotiske hybridiserings-barrierer til en mindre kjent slekt kalt *Draba*. Vi foreslår at lav parental konflikt i de tre selvbefruktende artene *D. fladnizensis, D. nivalis* og *D. subcapitata* sannsynligvis kan forklare fraværet av hybridiseringsbarriere mellom artene. I tillegg undersøkte vi genomisk imprinting i *D. nivalis*. Denne analysen resulterte i et høyt antall maternelt uttrykte gener, som ofte tilhørte

biologiske prosesser som tidligere har vist seg å være viktige for tilpasninger til arktisk klima for denne arten (Artikkel III).

Samlet sett viser disse artiklene at den endosperm-baserte post-zygotiske hybridiserings-barrieren er styrt av et komplekst maskineri, som påvirkes av forskjellige genetiske faktorer slik som artsvarianter, enkeltmutasjoner eller forskjeller i regulering av MADS-boks type-I gener. Våre resultater indikerer at barrieren kan forårsakes av forskjeller i reproduksjonssystem (utkryssing versus selvbefruktning) og kan manipuleres av miljøfaktorer slik som temperatur. Samlet sett gir disse funnene et godt utgangspunkt for å studere og manipulere endosperm baserte post-zygotiske hybridiserings-barrierer.

INTRODUCTION

The importance of seeds is frequently under-communicated, despite their fundamental role as the primary basis for human sustenance, constituting 70% of the human caloric intake. Production of surplus food, beyond personal consumption, is enabled by seeds and is the basis for modern society. Seeds hold such immense significance that a dedicated global vault on Svalbard has been established to ensure long-term preservation and protection. As we currently know little about the effects of elevated temperatures and increasingly extreme weather conditions on seed development and plant propagation, such storage and protection measures are essential.

The seed in flowering plants (angiosperms) has a distinctive structure composed of three compartments: the seed coat, the endosperm, and the embryo. Each part of the seed serves different functions: the seed coat is a protective barrier against external factors that also transmit external cues to the interior of the seed (Radchuk & Borisjuk, 2014) and the endosperm serves as a nutrient storage for the embryo. The endosperm is an especially intriguing tissue, as it is triploid in a diploid species, with twice the maternal contribution compared to the paternal. During the syncytial stage of the endosperm, it works as a nutrient sink and after separating each nucleus into its own cell (cellularization) it switches to being a nutrient source, making the time point of endosperm cellularization essential for embryo viability (Lafon-Placette & Köhler, 2014).

A distinct hybridization barrier found throughout the angiosperm clade is manifested in some hybrids as a change in the time point of endosperm cellularization (Cornejo *et al.,* 2012; Lafon-Placette & Köhler, 2016; Oneal *et al.,* 2016; Petrén *et al.,* 2023; Tonosaki *et al.,* 2018). This so-called endosperm-based post-zygotic hybridization barrier (ePZB) is hypothesized to be caused by a disruption in the genomic ratio of the endosperm, which occurs in hybrids between two species of different ploidy. However, abnormal endosperm development is also observed in hybrids between species with the same ploidy. The term Endosperm Balance Number (EBN) was introduced to explain how effective ploidy may differ from the actual ploidy (Johnston *et al.,* 1980) and it was hypothesized that genomic imprinting could influence the effective ploidy (Bushell *et al.,* 2003; Haig & Westoby, 1991).

Genomic imprinting is an epigenetic phenomenon, which in plants is mainly identified in the endosperm, with a few exceptions in the embryo (Batista & Köhler, 2020; Hornslien *et al.,* 2019). Initially, studies on genomic imprinting were restricted to a few or single genes of interest. Today, studies on overall genomic imprinting within species emerge because of improved next-generation sequencing, which facilitates a rapid increase of both genome- and RNA sequencing. Interestingly, when comparing genomic imprinting across species, a very low overlap has been found, suggesting that it evolves at a rapid rate (Pignatta *et al.,* 2014; Zhang *et al.,* 2011).

As the majority of studies on the ePZB have focused on genomic imprinting and genetic factors to manipulate and understand the barrier, there are a few studies indicating that temperature increase, or heat stress, also has an impact on endosperm cellularization (Bjerkan *et al.,* 2019; Chen *et al.,* 2016; Folsom *et al.,* 2014; Paul *et al.,* 2020). Increasing our understanding of seed development, in terms of genetic regulation, hybridization, and the effects of abiotic factors, can contribute to crop improvement and help us prepare for future environmental changes.

1. Seed development and model systems in Brassicaceae

1.1 Reproduction and seed development

Reproduction in angiosperms depends on microgametophytes (pollen grains) and megagametophytes (embryo sacs). Each pollen grain contains two sperm cells and one vegetative cell (Dresselhaus *et al.,* 2016). The polygonum-type embryo sac, which is the most common type among angiosperms and also the prevalent type in Brassicaceae (Yadegari & Drews, 2004) contains three antipodal cells, a homodiploid central cell, an egg cell, and two synergids (Figure 1 A). The embryo sac is enclosed in the nucellus and further surrounded by the integuments, which form an opening, the micropyle, at the apex close to the egg cell (Bajon *et al.,* 1999; Rudall, 2021). When a pollen grain adheres to the stigma, it extends a pollen tube through the stigma, style, and eventually the ovule. When reaching the ovule, the pollen tube enters the embryo sac through the micropyle and penetrates one of the synergid cells, where it bursts and delivers two sperm cells to the embryo sac. One sperm cell fertilizes the egg cell and the other fertilizes the central cell, known as the double fertilization process (Dresselhaus & Franklin-Tong, 2013; Dresselhaus *et al.,* 2016). The egg- and one of the sperm cells produce

the embryo, whereas the homodiploid central cell and the other sperm cell produce the endosperm. Simultaneously, the integuments develop into the seed coat, and together these structures make up the seed (Figure 1 B). The different parts of the seed contain different ratios of parental genetic material. The embryo has an equal parental contribution (1maternal:1paternal), the endosperm has twice the maternal contribution compared to the paternal (2m:1p) and the seed coat is entirely of maternal origin (2m:0p).

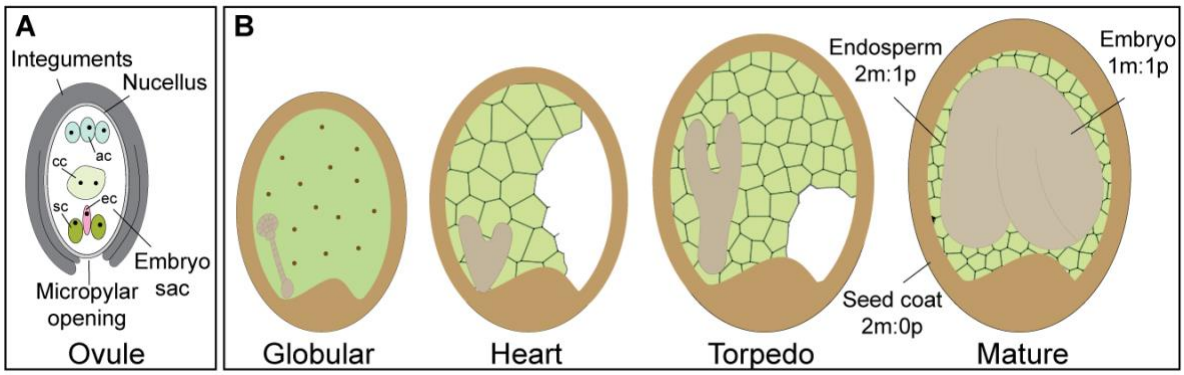


Figure 1: Illustration of an ovule together with the globular, heart, torpedo, and mature dicot seed. (A) The ovule encompasses the inner and outer integuments and the nucellus which encloses the embryo sac. The micropylar opening is close to the egg cell (ec), which lies adjacent to the two synergid cells (sc); the homodiploid central cell (cc) lies in the center and the three antipodal cells (ac) lie on the opposite side of the micropylar opening. **(B)** Illustration of four embryo stages found in dicot seeds, including the globular stage with uncellularized endosperm, heart, and torpedo with partial endosperm cellularization and a mature embryo with complete cellularized endosperm. The mature dicot seed consists of three separate tissues, with different parental genomic ratios as indicated. The seed coat has two maternal (m) copies and no paternal (p) copy, the endosperm has two maternal copies and one paternal copy, and the embryo has one copy from each parent. The chalazal region is not indicated in the illustration. Seed cartoons inspired by Hsieh *et al.* (2011).

During early seed development, the embryo acquires nutrients through the suspensor which connects the embryo to the maternal tissues and the endosperm (Kawashima & Goldberg, 2010). As the embryo reaches the heart stage (Figure 1 B), it transitions to nutrient uptake directly from the endosperm, and the suspensor degenerates (Morley-Smith *et al.*, 2008). Initially, the endosperm develops as a syncytium, with mitotic divisions devoid of cytokinesis, resulting in a large cell with multiple nuclei surrounding the central vacuole. During this stage, the endosperm acquires nutrients that are supplied from the phloem to the integument of the seed, where it moves symplastically throughout the outer integument and then apoplastically to the inner integument and further to the endosperm (Figure 2). Eventually, during endosperm development, each nucleus will be divided into its own cell, in a process called cellularization. After complete cellularization, the endosperm will supply the embryo with its nutrient repository (Lafon-Placette & Köhler, 2014).

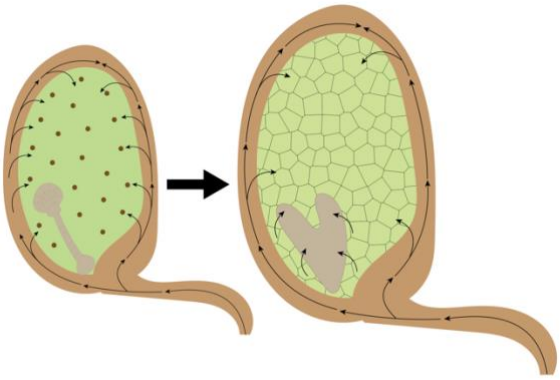


Figure 2: Illustration of nutrient allocation pathway before and after cellularization of the endosperm. Left: Early stage of seed development, where the endosperm works as a nutrient sink. The arrows show the sucrose pathway from the phloem to the integument of the seed, where it moves symplastically throughout the outer integument (not shown), and apoplastically to the inner integument and endosperm (shown for both seeds). The sucrose is further transported to the central vacuole of the endosperm, where it is hydrolyzed into hexoses. Right: Late stage of seed development, with cellularized endosperm working as a nutrient source. Endosperm cellularization causes a decrease in the central vacuole size (not shown), which results in a higher ratio of sucrose to hexose. This allows for transport of sucrose from the endosperm to the embryo.

Endosperm phase transition is essential for embryo viability. This has been demonstrated by mutants in *FERTILIZATION INDEPENDENT SEED POLYCOMB COMPLEX 2* (*FIS-PRC2*), which results in endosperm cellularization failure (Chaudhury *et al.,* 1997). Embryos in these seeds are arrested at the heart stage but can be rescued by isolating the embryo and cultivating them in vitro (Chaudhury *et al.,* 1997; Hehenberger *et al.,* 2012). Although embryo development can proceed without endosperm cellularization, Song *et al.,* (2021) showed that embryo maturation is dependent on endosperm-synthesized LEAFY COTYLEDON 1 (LEC1), suggesting that LEC1 is provided to the embryo before endosperm cellularization. The intricacy of embryo-endosperm interplay is further emphasized by two seemingly conflicting studies, examining seed development in single-fertilization situations. Xiong *et al.,* (2021) showed that the endosperm develops normally in the absence of egg cell fertilization, whereas Nowack *et al.,* (2006) showed that in the *CYCLIN DEPENDENT KINASE A 1* (*CDKA;1*) mutant, which results in fertilization of the egg cell alone, the homodiploid central cell still initiates endosperm development, suggesting that fertilization of the egg cell releases a signal for endosperm development to proceed. Aw *et al.,* (2010) showed that the *cdka;1* mutant predominantly delivers two sperm cells, with one fertilizing the egg cell, while the other sperm cell fails in nuclei fusion with the central cell. However, central cell division is still initiated, suggesting that sperm cell entry is sufficient to activate the progression of endosperm development.

1.2 Phylogeny and importance of the Brassicaceae family

Angiosperms constitute a major clade, encompassing approximately 300,000 distinct species, accounting for nearly 90% of all land plants. The diversity and abundance of angiosperms are especially remarkable considering their relatively recent origin, estimated to be around 140-250 million years ago (mya) (Beaulieu *et al.,* 2015; Foster & Ho, 2017; Magallón *et al.,* 2015), within a timespan of approximately 480 mya since the emergence of land plants (Strother & Foster, 2021). Charles Darwin, struggling to explain this rapid rise and early diversification, referred to it as an abominable mystery (Darwin, 1903).

Within the angiosperms, the Brassicaceae family originated around 25-30 mya (Hendriks *et al.,* 2023; Hohmann *et al.,* 2015). This successful family includes approximately 4140 species (German *et al.,* 2023), distributed worldwide, except for Antarctica (Lysak & Koch, 2011). It includes numerous economically important plant species, with diverse applications in medicine, agriculture, and scientific research (Raza *et al.,* 2020). One of the more recent Brassicaceae phylogenies (Hendriks *et al.,* 2023) divides the family into five supertribes: Camelinodae, Brassicodae, Hesperodae, Arabodae, and Heliophilodae (Figure 3), with Brassicodae (II) and Camelinodae (I) being the most extensively studied.

Brassicodae includes the prevalent tribe Brassiceae, with the genera *Sinapis* and *Brassica* (Figure 3). The most well-known species within *Sinapis* is *S. alba*, also known as the white mustard, which can be used as a food condiment and biofuel (Ciubota-Rosie *et al.,* 2013; Warwick, 2011). The Brassica genus comprises several crop species, including *B. oleraceae* encompassing cultivars like cauliflower, broccoli, cabbage, and Brussel sprouts, *B. rapa* which includes cultivars like turnip, bok choy, and rapini, *B. napus* which is also known as rapeseed, *B. juncae* which is used as food condiment, and *B. carinata* which can be used as a biofuel crop (Warwick, 2011). A whole genome triplication (WGT) event occurred in a common ancestor of Brassiceae (Lysak *et al.,* 2005) and there is evidence for recursive whole-genome duplications (WGD) in each of the *Brassica* genomes (Bowers *et al.,* 2003). These events result in complex genomes with numerous paralogs, and pseudogenes, which pose significant challenges in genetic research.

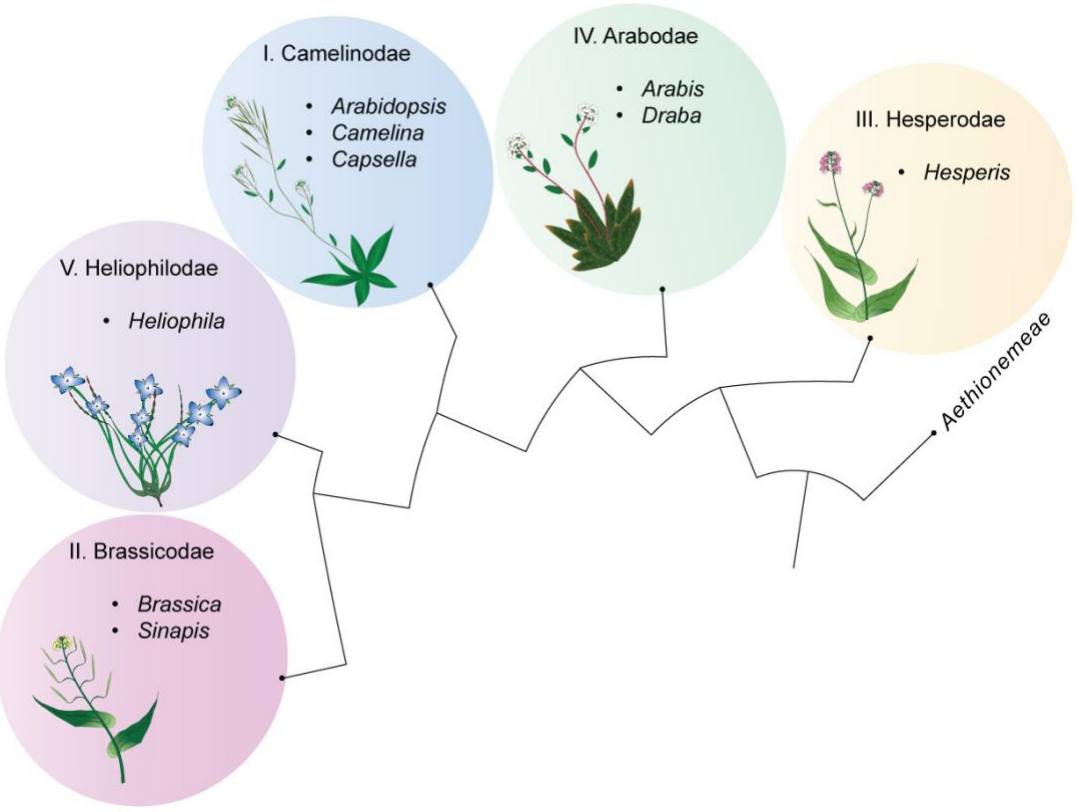


Figure 3: Brassicaceae phylogeny. The phylogenetic tree divides the Brassicaceae family into five supertribes: Camelinodae, Brassicodae, Hesperodae, Arabodae, and Heliophilodae with Aethionemeae as an outgroup. In Camelinodae (I), *Arabidopsis*, used in this thesis, is highlighted and illustrated by *A. thaliana*. Two other well-known genera within this supertribe, *Camelina* and *Capsella*, are also highlighted. Brassicodae (II) is mostly known for the genus *Brassica* and here illustrated by *B. rapa*. Another well-known genus, *Sinapis*, is also listed. In Hesperodae (II), the genus *Hesperis* is featured and illustrated by *H. matronalis*. Within Arabodae (IV), two genera are highlighted, *Draba*, which is used in this thesis and here illustrated by *D. nivalis*, and the well-known genus *Arabis*. Heliophilodae (V) is highlighted by the genus *Heliophila* and illustrated by *H. longifolia*. The schematic phylogenetic tree is based on the nuclear phylogeny published by Hendriks *et al.* (2023).

Camelinodae includes several important genera, with among others the genera *Arabidopsis*, *Camelina,* and *Capsella* (Figure 3). *Camelina sativa*, also known as false flax, used to be an important oilseed crop, however, today it is primarily used for biofuel (Kagale *et al.*, 2014). Within *Capsella* the three species *C. rubella*, *C. grandiflora,* and *C. bursa-pastoris* constitute an excellent model system due to different ploidies and mating strategies, with *C. rubella* being a diploid self-fertilizing plant, *C. grandiflora* a diploid outbreeder, and *C. bursa-pastoris* a tetraploid predominantly self-fertilizing plant (Nutt *et al.,* 2006). *Arabidopsis* is the most well-studied genus of all land plants and includes the principal model-plant species *A. thaliana*.

1.3 The model system *Arabidopsis*

A. thaliana makes a great model species because of several traits including self-fertilization, which reduces genetic variation between individual plants and allows for simple propagation of mutant lines, a small genome of approximately 135 megabases (Mbp), prolific seed production, short life cycle, large public collection of mutant lines and efficient transformation protocol with *Agrobacterium tumefaciens* (Clough & Bent, 1998; Woodward & Bartel, 2018). *A. thaliana* was the first plant to have its genome sequenced (The Arabidopsis Genome Initiative, 2000). Since then, the genome resource has been revised by The Arabidopsis Information Resource (TAIR), with TAIR10 and Araport11 being the latest genome and annotation versions available (Lamesch *et al.,* 2012). In addition, several genomes from different accessions have been released, enhancing our understanding of the variation present within the species (Jiao & Schneeberger, 2020; Lamesch *et al.,* 2012; Naish *et al.,* 2021; Wang *et al.,* 2022). Furthermore, genome assemblies of *A. arenosa*, *A. lyrata,* and *A. halleri* have been published, although with lower quality than the *A. thaliana* genome (Akama *et al.,* 2014; Bohutínská *et al.,* 2021; Briskine *et al.,* 2016; Hu *et al.,* 2011; Kolesnikova *et al.,* 2023). In contrast to *A. thaliana*, the other diploid *Arabidopsis* species, *A. arenosa*, *A. halleri*, and *A. lyrata* are outbreeding perennials, adapted to different environmental conditions (Yant & Bomblies, 2017). In addition, natural autotetraploid populations of *A. arenosa* and *A. lyrata* are found, and tetraploid populations of *A. arenosa* are shown to have adaptive flexibility towards more extreme environments (Konečná *et al.,* 2021). This variation in mating systems, ploidy, and environmental adaptations, coupled with available genomic resources, makes the genus *Arabidopsis* ideal for studying a wide range of plant-related questions. However, *Arabidopsis* falls short when considering certain scientific questions, including the special case of plant adaptation to the Arctic.

1.4 The model system *Draba*

Draba is a genus from the less studied Arabodae supertribe (Figure 3). The genus originated around 1.36-2.71 mya and has rapidly speciated since then (Al-Shehbaz *et al.,* 2006; Bailey *et al.,* 2006; Mulligan, 1976). It is predominantly located in alpine to arctic geographic locations and encompasses approximately 370 species, making it the largest genus within Brassicaceae (Warwick *et al.,* 2006). This complex genus exhibits frequent hybridization events, polyploidization, and occasionally apomixis, making reconstruction of a

comprehensive phylogeny difficult (Brochmann *et al.,* 1993; Jordon-Thaden *et al.,* 2013; Jordon-Thaden & Koch, 2008). The only published *Draba* genome is from the diploid *D. nivalis* (Nowak *et al.,* 2020). In addition to *D. nivalis*, two other diploid *Draba* species have been particularly studied, *D. fladnizensis* and *D. subcapitata*. They are all self-fertilizing plants with eight chromosome pairs, and low genetic diversity between them suggests recent divergence during Pleistocene (Grundt *et al.,* 2006, 2004; Jordon-Thaden & Koch, 2008; Mulligan, 1974, 1976). Alpine, sub-arctic, and arctic species belong to environments that experience the fastest rate of global warming, which makes *Draba* an especially vulnerable genus (Birkeland *et al.,* 2020, 2022; Rantanen *et al.,* 2022) and important as a potential model system for addressing ramifications of elevated temperatures on alpine, sub-arctic, and arctic species.

2. Hybridization barriers in Brassicaceae

2.1 Speciation driven by hybridization

Darwin saw natural selection as a mechanism for evolutionary change and adaptation within populations, which could eventually lead to new species (Darwin, 1859). The underlying mechanisms allowing for this divergence were unknown until population genetics showed how new variations could be introduced into populations by mutations and recombination (Provine, 1977). Despite an increasing understanding of the speciation process, the definition of a species has remained problematic and several species concepts have been proposed (Mallet, 2013).

The most well-known species concept is the biological species concept, stating that: “species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups” (Mayr, 1970). The important trait of a species, as outlined by the biological species concept, involves protection from gene flow through physiological isolating mechanisms, established either prior to (pre-zygotic) or following fertilization (post-zygotic) (Dobzhansky, 1937). Two main shortcomings of the biological species concept are the omission of asexual species, and its lack of acknowledging the occurrence and evolutionary potential of hybridization. When the concept was introduced, hybridization was regarded as an infrequent occurrence and consequently insignificant to take into consideration. However, genomic data reveals ample evidence of horizontal gene transfer and interbreeding among species (Mallet, 2013; Mallet *et al.,* 2016), which is especially

prevalent in plants. Gene flow eventually leading to introgression of advantageous traits (Novikova *et al.,* 2016; Suarez-Gonzalez *et al.,* 2016) and speciation by hybridization (Stebbins, 1959; Vallejo-Marín *et al.,* 2015; Yakimowski & Rieseberg, 2014) are important factors in plant evolution, allowing for increased variation in these sessile species.

2.2 Hybridization in *Draba*

Hybridization in *Draba* has not been extensively studied, however the few studies that have been conducted identified reproductive barriers between species. Brochmann (1993) focused specifically on the polyploid species of the *D. alpina* complex, which includes 13 polyploid and three diploid species, and found primarily post-zygotic barriers with partly sterile F1 individuals. However, crosses between the three diploid species *D. fladnizensis*, *D. nivalis*, and *D. subcapitata* from the same study resulted in no adult F1 individuals suggesting that earlier hybridization barriers are at play. Crosses between *D. fladnizensis* and *D. nivalis* suggested a complete pre-zygotic barrier with no fertilized ovules, whereas crosses between *D. fladnizensis* and *D. subcapitata* had approximately 23% seed set with non-viable seeds and crosses between *D. nivalis* and *D. subcapitata* resulted in a 50% seed set with a 14.3% germination rate (Brochmann *et al.,* 1993). Similar interspecies crosses have previously been attempted between these three diploid species, although these studies report entirely aborted seeds (Heilborn, 1927; Mulligan, 1974). Despite these reported strong interspecific hybridization barriers, *D. fladnizensis* and *D. nivalis*/*D. subcapitata* have been suggested as parental species of the hexaploid *D. lactea* (Brochmann *et al.,* 1992; Böcher, 1966), implying that viable seeds from these crosses could be obtained, although not yet demonstrated in any experiments (Brochmann *et al.,* 1993; Heilborn, 1927; Jørgensen, 1958; Mulligan, 1974). Grundt et al. (2006) found post-zygotic hybridization barriers in intraspecies crosses using the same three diploid species, *D. fladnizensis*, *D. nivalis*, and *D. subcapitata*. Crosses between different populations from separate or even the same geographic region resulted in a high occurrence of infertile individuals, revealing the presence of numerous cryptic species within each taxonomic species (Grundt *et al.,* 2006). The presence of cryptic species lacks overall genetic support, suggesting that the evolution of hybridization barriers happened rapidly involving limited parts of the genome (Gustafsson *et al.,* 2014; Skrede *et al.,* 2009, 2008).

2.3 Hybridization in *Arabidopsis*

Hybridization has occurred multiple times between *Arabidopsis* species in nature, which on two occasions resulted in new allotetraploid species: *A. kamchatica*, which is a hybrid between diploid *A. lyrata* and *A. halleri* ssp. *gemmifera* (Shimizu-Inatsugi *et al.*, 2009) and *A. suecica*, which is a hybrid between *A. thaliana* and diploid *A. arenosa* (Jakobsson *et al.*, 2006; O'Kane *et al.*, 1996). The latter hybrid is interesting because an incomplete post-zygotic hybridization barrier has been found between maternal *A. thaliana* and paternal diploid *A. arenosa*, whereas the reciprocal cross shows a complete pre-zygotic hybridization barrier (Bjerkan *et al.*, 2019; Burkart-Waco *et al.*, 2015). Gene flow and incorporation of adaptive traits through introgression between *A. lyrata* and *A. arenosa* have also been reported, although these two species exhibit strong hybridization barriers (Jørgensen *et al.*, 2011; Schmickl & Koch, 2011). *A. lyrata* and *A. arenosa* are especially interesting because they have mainly separate diploid and autotetraploid populations, and the ploidy of the species has an effect on the strength of the hybridization barrier. Reciprocal crosses between diploid individuals result in almost no viable seeds and similar results are obtained when reciprocally crossing diploid *A. lyrata* to tetraploid *A. arenosa*. However, reciprocal crosses between diploid *A. arenosa* and tetraploid *A. lyrata*, or reciprocal crosses between tetraploid individuals, result in almost complete germination of hybrid seeds (Lafon-Placette *et al.*, 2017). The inviable hybrid seeds from crosses between *A. lyrata* and *A. arenosa* and between maternal *A. thaliana* and paternal diploid *A. arenosa*, exhibited developmental problems specifically within the endosperm (Bjerkan *et al.*, 2019; Lafon-Placette *et al.*, 2017).

2.4 The endosperm-based post-zygotic hybridization barrier

The ePZB causes inviable hybrid seeds because of phase-transition problems in the endosperm. This can either be precocious or delayed cellularization of the endosperm (Figure 4), ultimately affecting embryo viability (Cooper & Brink, 1942). The barrier is found in several genera throughout the angiosperm clade, including *Oryza* (Ishikawa *et al.*, 2011; Sekine *et al.*, 2013; Tonosaki *et al.*, 2018; Wang *et al.*, 2018; H.-Y. Zhang *et al.*, 2016), *Arabidopsis* (Lafon-Placette & Köhler, 2016), *Capsella* (Dziasek *et al.*, 2021; Rebernig *et al.*, 2015), *Arabis* (Petrén *et al.*, 2023), *Solanum* (Cornejo *et al.*, 2012; Florez-Rueda *et al.*, 2016; Johnston &

Hanneman, 1982; Roth *et al.*, 2019), and *Mimulus* (Flores-Vergara *et al.*, 2020; Kinser *et al.*, 2021; Oneal *et al.*, 2016), suggesting that it is controlled by a conserved mechanism. Abnormal endosperm development can be induced by a disruption in the genomic ratio, which occurs in hybrids between two species of different ploidy. However, abnormal endosperm development is also observed in hybrids between species with the same ploidy. The hybridization barrier between *A. lyrata* and *A. arenosa* is found both in crosses with the same ploidy and varying ploidy. In addition, the endosperm phenotype varies with the cross-direction. The endosperm in the cross using diploid *A. lyrata* as the maternal contributor and diploid *A. arenosa* as pollen donor has delayed cellularization, whereas the reciprocal cross has precocious cellularization (Lafon-Placette *et al.*, 2017). Similarly, when crossing *A. thaliana* as the maternal part to diploid *A. arenosa* as a pollen donor, the endosperm has delayed cellularization (Bjerkan *et al.*, 2019). The EBN hypothesis can be used to account for such observations, by explaining how the effective ploidy may differ from the actual ploidy (Johnston *et al.*, 1980), and how crosses between species with the same ploidy can still result in disruption of the genomic ratio in the endosperm. This imbalance can lead to paternal or maternal excess, with paternal excess displaying delayed cellularization and maternal excess displaying precocious cellularization (Haig & Westoby, 1991).

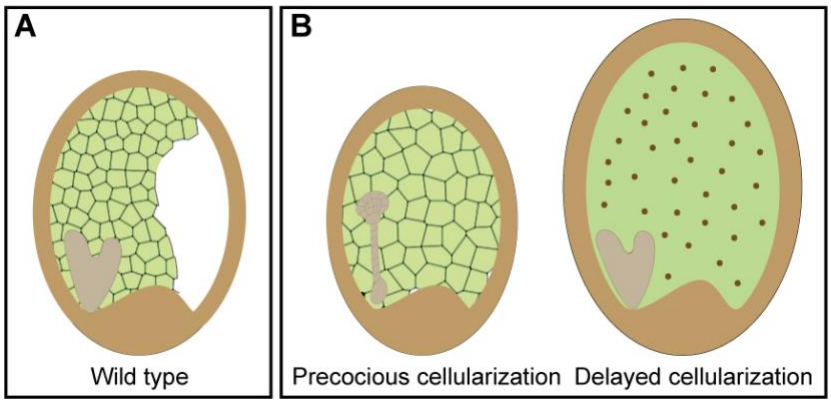


Figure 4: Illustration of wild-type seed and seeds with precocious or delayed cellularization of the endosperm. (A) Wild-type seed as found in *A. thaliana* at six days after pollination (DAP). **(B)** Precocious cellularization of the endosperm, with an embryo at the globular stage and delayed cellularization of the endosperm with an embryo at the heart stage.

2.5 Genetic factors affecting the ePZB

In the cross between *A. thaliana* and *A. arenosa*, the strength of the ePZB varies when using different accessions. Burkart-Waco *et al.*, (2011) showed that diploid *A. arenosa* used as pollen donor to 53 different accessions of *A. thaliana* resulted in 0-30% live seeds, displaying a wide range of variation. When (Bjerkan *et al.*, 2019) crossed *A. thaliana* and diploid *A.*

arenosa, they found an accessions-based maternal effect. The germination rate ranged from approximately 30 to 50% when using four different *A. arenosa* accessions, and from 10 to 80% when using four different *A. thaliana* accessions. Certain mutants have also shown an effect on the ePZB strength in the same cross, with an 1-10% increased rate of live seeds when using single mutant backgrounds of MADS-BOX type-I transcription factor genes, *agamous-like 62* (*agl62*) and *agl90* in the maternal *A. thaliana* (Walia *et al.*, 2009). Interestingly, (Bjerkan *et al.*, 2019) did not find a significant change in the germination rate of the *A. thaliana* × *A. arenosa* cross when using the same mutants, arguing the use of different *A. arenosa* accessions could overshadow the effects.

2.6 Temperature effect on the ePZB

Research on how temperature affects early seed development is limited to a few studies. In rice, moderate heat stress accelerates the cellularization event, whereas severe heat shock delays cellularization (Chen *et al.*, 2016; Folsom *et al.*, 2014; Paul *et al.*, 2020). Temperature effects on seed set and embryo development have been found in interspecific *Hordeum* crosses, but these studies dissected the embryo to grow on media, excluding any endosperm effect (Molnár-Láng & Sutka, 1994; Pickering, 1984; Sitch & Snape, 1987; Thörn, 1992; Tiara & Larter, 1978). Additionally, Bjerkan *et al.*, (2019) demonstrated a significantly decreased germination rate of hybrid seeds between *A. thaliana* and *A. arenosa* when the temperature was elevated from 18°C to 22°C. They also investigated whether the single mutants *agl23*, *agl34*, *agl35*, *agl36*, *agl62*, *agl90* or the double mutant *agl36 agl62* had an effect on temperature-sensitivity in the *A. thaliana* × *A. arenosa* hybrid cross exposed to 18°C and 22°C. The single mutants *agl23*, *agl36*, *agl62*, and *agl90* continued to show significant differences in germination rates between the two temperatures. However, *agl34*, *agl35*, and the double mutant *agl36 agl62* neutralized the temperature effect, with the strongest neutralization in the *agl35* mutant. It was therefore suggested that AGL35 is involved in relieving hybridization barriers and sensing temperature changes (Bjerkan *et al.*, 2019).

3. Evolutionary origin and establishment of genomic imprinting

3.1 Evolutionary origin

Genomic imprinting is proposed as a factor influencing effective ploidy, by contributing to a disruption in the genomic ratio of the endosperm in crosses between species with either matching or different ploidy (Bushell *et al.*, 2003; Haig & Westoby, 1991). It is characterized by the parent-of-origin specific expression of alleles, with either maternally expressed genes (MEGs) or paternally expressed genes (PEGs) (Batista & Köhler, 2020; Hornslien *et al.*, 2019). Two main theories are suggested to explain the occurrence of imprinting in the endosperm, namely the "kinship theory" (Haig & Westoby, 1991) and the "maternal-offspring co-adaptation theory" (Wolf & Hager, 2006). The kinship theory is based on a difference of interest in resource allocation to the embryo between the paternal and maternal parts. In outbreeding species, the female will be fertilized by different male individuals and it is beneficial with an equal resource allocation to the developing embryos, whereas the male counterpart would benefit from increased resource allocation to its own progeny. This theory fits with the observed trends in crosses between *Arabidopsis* lines with different ploidy levels, where an increased parental dosage results in an increased size of the developing seed, whereas an increased maternal dosage results in a decreased size (Scott *et al.*, 1998). The parental conflict predicted by the "kinship theory" is the basis for the weak inbreeder/strong outbreeder (WISO) hypothesis, which predicts that the genomes of more outbred species will have a higher effective ploidy compared to the genomes of more inbred species. Crosses between species with different mating systems will therefore result in an imbalance in resource allocation within the hybrid seeds (Brandvain & Haig, 2005, 2018; Raunsgard *et al.*, 2018). In species with nuclear endosperm development, this imbalance in resource allocation will be caused by precocious or delayed endosperm cellularization (Bjerkan *et al.*, 2019; Lafon-Placette *et al.*, 2018; Lafon-Placette & Köhler, 2014). One problem with the kinship theory is that it does not explain the abundance of MEGs compared to PEGs. The maternal-offspring co-adaptation theory (Wolf & Hager, 2006), on the other hand, explains how the inherited maternal alleles in the offspring interact favorably with the mother's genotype, which could result in an increase of MEGs in the endosperm.

3.2 Establishing genomic imprinting

Genomic imprinting is enabled by differential epigenetic modifications on the maternal and paternal genes, and these modifications are located either as direct modifications of DNA bases or changes in chromatin condensation guided by histone modifications (Batista & Köhler, 2020; Law & Jacobsen, 2010).

DNA methylation is the addition of a methyl (CH3) group directly on the DNA strand, typically on cytosine (C) bases in a CG sequence context, where G denotes guanine. Cytosine methylations also occur in a so-called asymmetric CHH- and CHG sequence context, where H denotes adenine, cytosine, or thymine (Henderson & Jacobsen, 2007). Methylation of cytosine is mediated by specific methyltransferases, including METHYLTRANSFERASE 1 (MET1), which maintains symmetric cytosine methylation in CG sites (Finnegan & Dennis, 1993). Non-CG methylations are redundantly generated by CHROMOMETHYLASE 3 (CMT3) and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Cao & Jacobsen, 2002; Chan *et al.,* 2004). DNA methylation can be reversed, and DEMETER (DME) DNA glycosylases excise methylated cytosines and replace them with unmethylated cytosines (Gehring *et al.,* 2006). The current model for genomic imprinting by DNA methylation is based on the antagonistic relationship between MET1 and DME, where DME removes DNA methylation which is not replaced by MET1 in the central cell as it is downregulated in this tissue, resulting in the expression of the maternal allele only (MEGs). DNA methylation has been documented in endosperm-specific genes with transposon sequences in the promoter region and it has been postulated that imprinting evolved as a consequence of transposon silencing. Imprinted genes with transposable elements within the promoter region include genes such as *MEDEA* (*MEA*), *FLOWERING WAGENINGEN*, *PHERES 1* (*PHE1*), *HOMEOBOX-LEUCINE ZIPPER PROTEIN 3* (*HDG3*), *HDG9*, and *AGL36*, but not all imprinted genes have foreign DNA in the promoter regions (Gehring *et al.,* 2009; Kinoshita *et al.,* 2007; Raissig *et al.,* 2011; Shirzadi *et al.,* 2011).

Histone methylation refers to the addition of methyl group(s) to specific amino acids in the histone tail. There are several common histone modifications, but only one is known to be involved in genomic imprinting, namely the trimethylated lysine residue 27 on histone H3 (H3K27me3) (Rodrigues & Zilberman, 2015). This trimethylation is performed by a Polycomb Repressive Complex 2 (PRC2) and in the endosperm of *Arabidopsis* this is specifically called

FERTILIZATION INDEPENDENT SEED (FIS)-PRC2. This complex is composed of four subunits, namely MEA, FIS2, FERTILIZATION INDEPENDENT ENDOSPERM (FIE) and MULTICOPY SUPPRESSOR OF IRA1 (MSI1) (Grossniklaus *et al.,* 1998; Kiyosue *et al.,* 1999; Köhler *et al.,* 2003a). In the case of PEGs, the maternal allele is DNA hypomethylated while the paternal allele is hypermethylated and the latter case interferes with the binding of FIS-PRC2, resulting in H3K27me3 methylation on the maternal allele alone (Deleris *et al.,* 2012; Weinhofer *et al.,* 2010). Addition of H3K27me3 methylations on the paternal allele in MEGs has been registered, although how the FIS-PRC2 complex distinguishes the paternal from the maternal allele is not known (Bai & Settles, 2015).

3.3 Genomic imprinting in angiosperms

Prior to the cost-effective and efficient advancements in sequencing technology, most research on genomic imprinting was focused on single genes hypothesized to be imprinted (Jullien *et al.,* 2006; Kinoshita *et al.,* 2003; Köhler *et al.,* 2005; Tiwari *et al.,* 2008; Vielle-Calzada *et al.,* 1999). While these studies delved deeply into the specific gene of interest, they provided limited insight into imprinting as a comprehensive phenomenon. A better overview of the number of imprinted genes and whether this is a conserved mechanism between angiosperm species was obtained when whole-genome imprinting studies emerged. This surprisingly showed that imprinting varied immensely, even between closely related species (Pignatta *et al.,* 2014; Zhang *et al.,* 2011), but it also revealed that genomic imprinting is found widespread throughout the angiosperm clade. Genomic imprinting has been investigated in genera such as *Zea* (Dong *et al.,* 2023; Zhang *et al.,* 2014), *Oryza* (Wang *et al.,* 2018; H.-Y. Zhang *et al.,* 2016), *Sorghum* (M. Zhang *et al.,* 2016), *Fragaria* (Liu *et al.,* 2021), *Solanum* (Florez-Rueda *et al.,* 2016; Roth *et al.,* 2019), *Ricinus* (Xu *et al.,* 2014), *Mimulus* (Kinser *et al.,* 2021), *Arabidopsis* (Del Toro-De León & Köhler, 2019; Hornslien *et al.,* 2019; Klosinska *et al.,* 2016; Picard *et al.,* 2021; Pignatta *et al.,* 2014; van Ekelenburg *et al.,* 2023; Wolff *et al.,* 2011), *Brassica* (Liu *et al.,* 2018; Rong *et al.,* 2021; Yoshida *et al.,* 2018) and *Capsella* (Hatorangan *et al.,* 2016; Lafon-Placette *et al.,* 2018), each of them containing numerous cases of imprinted genes. Most of these imprinting studies have isolated RNA from whole seeds, including the maternal seed coat (2m:0p), the embryo (1m:1p), and the endosperm tissue (2m:1p) (Figure 1), which results in a maternal bias when studying gene expression. However, some studies utilized methods for avoiding maternal bias, such as seed dissection into seed coat, endosperm, and embryo (Gehring *et al.,* 2011; Pignatta *et al.,* 2014) and endosperm nuclei

sorting (Del Toro-De León & Köhler, 2019; Picard *et al.,* 2021; van Ekelenburg *et al.,* 2023). In addition, (van Ekelenburg *et al.,* 2023) used domain-specific GFP-markers in order to examine differences in imprinting between endosperm domains. Despite these efforts, the overlap between studies is low, suggesting that mechanisms establishing imprinting of genes are subjected to rapid evolution. It is further hypothesized that genes that have been recently duplicated are more commonly regulated by genomic imprinting to control gene dosage (Yoshida & Kawabe, 2013). Furthermore, recent gene duplications make the identification of distinct orthologs between species more challenging, which could explain why identified imprinted genes suggest a low level of conservation.

3.4 MADS-box transcription factors and genomic imprinting

The MADS-box gene family encodes transcription factors (TFs), which have a shared DNA binding domain in the N-terminal region called the MADS-box (M). The gene family is divided into two lineages, Type-I and Type-II, based on their phylogenetic relationships (Alvarez-Buylla *et al.,* 2000). Notably, the Type-I lineage is highly divergent compared to the Type-II lineage and has undergone numerous recent duplication events. This lineage can be further classified into the subgroups Mα, Mβ, and Mγ. The Type-II lineage, which is subdivided into MIKC and Mδ, is highly conserved and includes many genes important for regulating floral organ formation (Irish, 2010; Pařenicová *et al.,* 2003). Previous studies have shown that the Type-I lineage includes several genes regulated by genomic imprinting and that many are specifically expressed during endosperm development (Bjerkan *et al.,* 2019; Köhler *et al.,* 2003b; Masiero *et al.,* 2011; Shirzadi *et al.,* 2011; Zhang *et al.,* 2018). However, imprinting of certain Type-II genes has been found, albeit at a much lower frequency, consistent with the hypothesis that highly duplicated genes are more often regulated by genomic imprinting (Del Toro-De León & Köhler, 2019; Picard *et al.,* 2021; Pignatta *et al.,* 2014). The high duplication rate of Type-I genes results in genetic redundancy and impedes functional studies. There are therefore only a few Type-I genes with a clear function, including *AGL23*, *PHE1*, *AGL61*, *AGL80,* and *AGL62*. AGL23 is involved in the biogenesis of organelles during embryo development (Colombo *et al.,* 2008) but has been identified as a PEG and is mainly paternally expressed (Hornslien *et al.,* 2019; Wolff *et al.,* 2011). Another PEG is PHE1 which regulates the expression of both imprinted genes and genes important in endosperm development (Batista *et al.,* 2019a; Köhler *et al.,* 2003b). AGL61 is important during female gametophyte

development and interacts with AGL80, which together are suggested to regulate a set of genes important during central cell development (Bemer *et al.,* 2008), while AGL62 regulates cellularization of the endosperm (Kang *et al.,* 2008). Despite difficulties in functional studies of the MADS-box type I genes, the expression in the endosperm together with the high imprinting rate of this gene family makes it an interesting group to further investigate in relation to endosperm development.

AIM OF STUDY

The primary objective of this thesis is to investigate the intricate mechanisms underlying endosperm-based post-zygotic hybridization barriers. This was achieved through three approaches:

(1) Generation of genome resources for *A. arenosa* and *A. lyrata* ssp. *petraea*, which together with new seed-specific transcriptomes are used to identify and examine the expression of MADS-box type-I genes. Since the MADS-box type-I family includes several genes regulated by genomic imprinting with specific expression in the endosperm, we expect to find differences in the expression of these genes, which may have the potential to explain the presence of endosperm-based post-zygotic hybridization barriers (Paper I).

(2) Investigation of endosperm-based post-zygotic hybridization barriers in two *Arabidopsis* hybrids, *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea* by exposing the crossed plants to a temperature gradient, using different *A. thaliana* accessions or by introducing single-gene mutations in the maternal *A. thaliana*. Based on reciprocal crosses between *A. arenosa* and *A. lyrata* ssp. *petraea*, we hypothesize that *A. thaliana* × *A. lyrata* ssp. *petraea* might have an opposite endosperm phenotype compared to *A. thaliana* × *A. arenosa*. If this hypothesis holds true, we further hypothesize that temperature, accessions, and single-gene mutations have opposed effects on the hybridization barrier phenotype in the two hybrids (Paper II).

(3) Expand the investigation of endosperm-based post-zygotic hybridization barriers and genomic imprinting to strong selfing species of the genus *Draba* belonging to the supertribe Arabodae of Brassicaceae. Due to low parental conflict in the selfing *Draba* species, we hypothesize an absence of endosperm-based post-zygotic hybridization barriers. We also hypothesize a lower number of conserved imprinted genes in *Draba* compared to other Brassicaceae species, as genomic imprinting studies are dominated by the more distantly related Camelinodae species. However, we hypothesize that a higher overlap may be found between *Draba* and other species with a selfing mating system (Paper III).

RESULTS AND DISCUSSION

In the work presented in this thesis, we have used established and new model systems from the genera *Arabidopsis* and *Draba*. We have used *A. arenosa* accession MJ09-4 from Pusté Pole in Slovakia, *A. lyrata* ssp. *petraea* accession MJ09-11 from Pernitz in Austria and four different accessions of *A. thaliana*, Col-0, C24, Ws-2, and Wa-1. We have used *D. fladnizensis* from three different populations originating from Norway, Alaska, and Greenland, *D. nivalis* with two different populations from Norway and one population each from Alaska and Greenland, and one *D. subcapitata* population from Greenland (Figure 5).

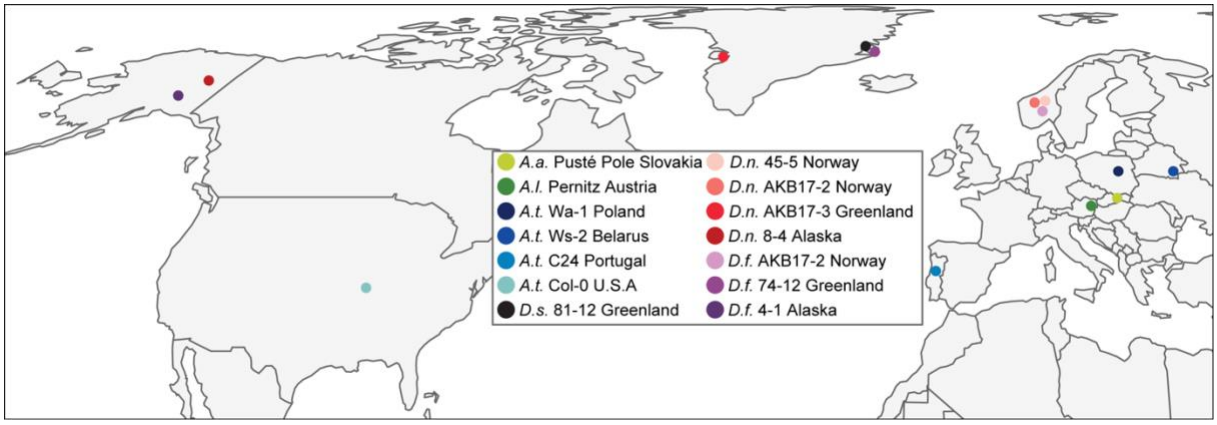


Figure 5: Geographic location of populations or accessions used in this thesis. Every species is indicated with its own color, and if different populations or accessions are used, a gradient of the species' color is utilized to indicate their different locations. The legend includes population name, country of origin, and abbreviations of species names: *Arabidopsis arenosa* (*A.a.*), *A. lyrata* ssp. *petraea* (*A.l.*), *A. thaliana* (*A.t.*), *Draba subcapitata* (*D.s.*), *D. nivalis* (*D.n.*), and *D. fladnizensis* (*D.f.*). Geographical coordinates for *A.t.* accessions are available on The Arabidopsis Information Resource (TAIR), *Draba* populations in Paper III, and *A.a.* and *A.l.* in Paper I.

4. Antagonistic endosperm phenotypes in *Arabidopsis* hybrids and absence of ePZB in *Draba*

Previous work on the ePZB within the *Arabidopsis* genus has mainly focused on the species combinations *A. arenosa* × *A. lyrata* or *A. thaliana* × *A. arenosa*. The results observed in these studies show delayed or precocious endosperm cellularization in crosses between *A. arenosa* and *A. lyrata*, depending on the ploidy and cross direction (Lafon-Placette *et al.*, 2017), as well as delayed cellularization in crosses between maternal *A. thaliana* and *A. arenosa* pollen donors (Bjerkan *et al.*, 2019; Walia *et al.*, 2009). Based on these results, we hypothesized that crosses between maternal *A. thaliana* and *A. lyrata* ssp. *petraea* pollen donor would result in

precocious endosperm development. We therefore performed this cross at 18°C and quantified germination rate as a measure of barrier strength as well as investigated the endosperm cellularization phenotype.

Endosperm nuclei were counted in seeds from *A. thaliana* self-crosses, *A. thaliana* × *A. arenosa*, and *A. thaliana* × *A. lyrata* ssp. *petraea* crosses. This revealed that the average number of nuclei in *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds (240 nuclei per seed) were significantly lower than in *A. thaliana* × *A. arenosa* hybrid seeds (340 nuclei per seed). Despite this, the germination rate was significantly higher for *A. lyrata* ssp. *petraea* hybrid seeds compared to *A. arenosa* hybrid seeds, with mean germination rates of approximately 63% and 40%, respectively. However, the germination rate of the *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds were still well below 100%, meaning that a post-zygotic hybridization barrier was present. To further investigate endosperm cellularization phenotypes in the hybrid crosses, we used feulgen-stained whole-mount embedded seeds to visualize the endosperm developmental stage. Hybrid seeds from the *A. thaliana* × *A. arenosa* cross showed no sign of cellularization as the embryo was still at an early developmental stage (Figure 6). However, hybrid seeds from the *A. thaliana* × *A. lyrata* ssp. *petraea* cross at a similar embryo stage displayed a higher frequency of precocious endosperm cellularization when compared to *A. thaliana* self and was in strong contrast to the endosperm phenotype in *A. thaliana* × *A. arenosa* hybrid seeds (Figure 6).

To broaden our perspective of the ePZB, we examined the presence of a barrier in three selfing *Draba* species *D. fladnizensis*, *D. nivalis*, and *D. subcapitata*, by performing interspecific crosses within three geographical regions (Alaska, Greenland, and Norway). As *D. subcapitata* was only sampled from Greenland, crosses within Alaska and Norway were between *D. fladnizensis* and *D. nivalis*. Despite previous studies showing either complete pre-zygotic or incomplete seed-based post-zygotic hybridization barriers, we achieved a germination rate with a median above 75% in all crosses. There was no significant difference in germination rates of hybrid seeds compared to midparent values, except for the *D. nivalis* × *D. subcapitata* hybrid, which had significantly higher germination rate compared to its midparent value, altogether suggesting that an ePZB is not present. Nevertheless, we observed significant differences between the paternal selfed seeds and the hybrid seeds from the reciprocal *D. nivalis* × *D. fladnizensis* crosses from Alaska and the *D. fladnizensis* × *D. subcapitata* hybrid seeds from Greenland. In addition, we observed that both the maternal and

paternal parent had significantly lower germination rates compared to the *D. nivalis* × *D. subcapitata* hybrid seeds from Greenland (Paper III). This implies that the maternal contribution has a larger influence on factors affecting germination rate, compared to the paternal component.

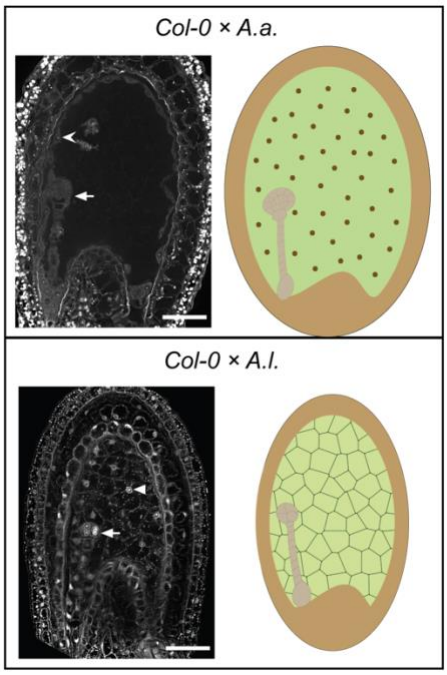


Figure 6: Endosperm phenotypes identified in two *Arabidopsis* hybrids. Confocal images of Feulgen-stained *A. thaliana* Col-0 (Col-0) crossed to *A. arenosa* (*A.a.*) or *A. lyrata* ssp. *petraea* (*A.l.*) six days after pollination emphasizing endosperm cellularization. Open arrowheads point to syncytial endosperm nuclei, closed arrowheads point to cellularized endosperm nuclei, and full arrows point to the embryo. In Col-0 × *A.a.*, the endosperm is mainly syncytial, whereas in Col-0 × *A.l.*, the hybrid seeds exhibit precocious endosperm cellularization at an early globular stage. The confocal images are accompanied by illustrations emphasizing the endosperm phenotypes. All crosses are female × male. Scale bar = 50µm.

The absence of ePZB between these *Draba* species is in support of the WISO hypothesis, as low parental conflict is expected in selfing species (Brandvain & Haig, 2005, 2018). However, the WISO hypothesis also predicts a higher effective ploidy in the outbreeding *A. arenosa* and *A. lyrata* ssp. *petraea* species compared to the selfing *A. thaliana*. Accordingly, we should expect a paternal excess and delayed endosperm cellularization in crosses using *A. thaliana* as the maternal part and *A. arenosa* or *A. lyrata* ssp. *petraea* as the pollen donor. This fits well with the observed endosperm phenotype in the *A. thaliana* × *A. arenosa* hybrid seeds but is the opposite of what we observe in the *A. thaliana* × *A. lyrata* ssp. *petraea*. This suggests that mating systems alone cannot predict the occurrence of ePZB or the resulting endosperm phenotype.

Further investigation of endosperm developmental progression was facilitated by endosperm markers expressed either before cellularization, Early Endosperm (EE), or after cellularization, Total Endosperm 1 (TE1). Both reporter lines are available in *A. thaliana* (Col-0), and in addition, we genetically modified *D. nivalis* individuals from the Norwegian

population (Grimsdalen) with the pAtTE1-GFP reporter from *A. thaliana*. Examination of EE and TE1 in hybrid crosses between *A. thaliana* and *A. arenosa* or *A. lyrata* ssp. *petraea* showed opposite phenotypes, corresponding to the endosperm developmental progression (Figure 6). The pAtEE-GFP marker showed a prolonged expression in the *A. thaliana* × *A. arenosa* hybrid seeds (Figure 7), in line with the observation that endosperm cellularization is delayed or absent. The pAtTE1-GFP reporter showed a premature expression in the *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds (Figure 7), corresponding with the observed precocious endosperm phenotype. In *A. thaliana* seeds, the EE marker showed decreased expression after six days after pollination (DAP), corresponding to the micropylar onset of cellularization, and few seeds expressed the marker line until eight DAP. Congruently, the pAtTE1-GFP reporter started expression at eight DAP and continued until seed maturation. A similar expression of the pAtTE1-GFP reporter was found in *D. nivalis* seeds, which developed at a slightly slower rate and therefore had weak expression from 10 DAP until seed maturation (Paper III). The timely expression of the pAtTE1-GFP reporter at the correct developmental stage indicates that corresponding mechanisms operate endosperm development and cellularization in *D. nivalis* and *A. thaliana*.

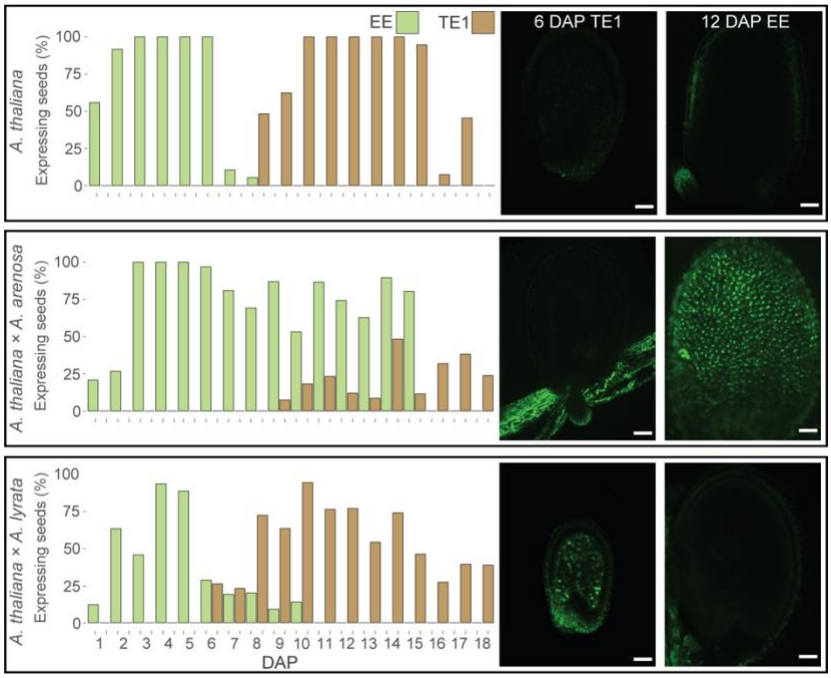


Figure 7: Expression and confocal images of early and late endosperm markers in *Arabidopsis* hybrid seeds. Percentage of seeds expressing proAT5G09370 >> H2A-GFP (pAtEE-GFP) and proAT4G00220 >> H2A-GFP (pAtTE1-GFP) at 1-18 days after pollination (DAP) together with confocal images of pAtTE1-GFP at six DAP and pAtEE-GFP at 12 DAP in *A. thaliana*, *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea*. pAtEE-GFP expression in the *A. thaliana* × *A. arenosa* hybrid seeds was not documented after 15 DAP. All crosses are female × male. Scale bar = 50µm.

5. The genetic network causing an ePZB is easily affected by using different accessions

Previous studies have indicated a substantial impact of accessions on the strength of the ePZB in crosses between *A. thaliana* and *A. arenosa*, especially when using different *A. thaliana* accessions (Bjerkan *et al.,* 2019; Walia *et al.,* 2009). To investigate whether this was the case for the *A. thaliana* × *A. lyrata* ssp. *petraea* cross, we expanded our experimental set-up to include the accessions C24, Ws-2 (Wassilewskija), and the tetraploid Wa-1 (Warschau) in addition to Col-0. Using colchicine tetraploidized strains of *A. thaliana* in crosses to *A. arenosa* had previously been reported to give an almost complete bypass of the ePZB (Josefsson *et al.,* 2006; Walia *et al.,* 2009). Furthermore, using the natural tetraploid Wa-1 in crosses with *A. arenosa* Strecno, resulted in approximately 50% normal seed production (Burkart-Waco *et al.,* 2011). We therefore expected the Wa-1 accession to have an ameliorative effect on the ePZB when crossed to *A. arenosa* Pusté Pole. Surprisingly, we observed that the C24 accession resulted in a stronger bypass of the ePZB in the *A. thaliana* × *A. arenosa* cross, compared to the tetraploid Wa-1 accession. In contrast, when crossed to *A. arenosa* Strecno, C24 resulted in approximately 17% normal seed production (Burkart-Waco *et al.,* 2011), showing a weaker bypass of the ePZB compared to Wa-1. In the *A. thaliana* × *A. lyrata* ssp. *petraea* cross, both using C24 and Wa-1 gave severely decreased seed viability compared to Col-0 and Ws-2. Using Ws-2 had the opposite effect of C24 and Wa-1, with severely decreased seed viability in *A. thaliana* × *A. arenosa* crosses and moderately increased seed viability in *A. thaliana* × *A. lyrata* ssp. *petraea* crosses. The stronger effect using C24 compared to Wa-1 was unexpected, but we hypothesize that using a different *A. arenosa* population and the use of a natural tetraploid population rather than a synthetic tetraploid strain causes this effect.

Overall, this implies that genetic differences between accessions could potentially have larger effects on the ePZB strength than the ploidy of the parents. Furthermore, our results suggest that C24 exhibits higher EBN compared to Col-0 and Ws-2. This supports the hypothesis that *A. arenosa* has a higher EBN compared to *A. lyrata* ssp. *petraea* (Lafon-Placette & Köhler, 2016), leading to enhanced seed viability when C24 or Wa-1 is crossed to *A. arenosa*, and decreased seed viability when crossed to *A. lyrata* ssp. *petraea*. Altogether, these results show that the genetic factors affecting the EBN and the ePZB are intricate and finely tuned, making it difficult to properly predict the effect that accession, and sometimes ploidy, can have on the barrier.

6. Single-gene mutations affecting the hybrid barrier

Utilizing specific genes or gene families, with conserved functions in endosperm development, holds the potential for reproducible results across different species in circumventing the ePZB. MADS-box type-I genes are emerging candidate genes for this purpose, as they are often expressed in the endosperm, and several are found to be imprinted. We aimed to investigate the effect of the two MADS-box type-I genes, *AGL40* and *AGL35*, along with the maternally expressed *MEA*, a component of the FIS-PRC2 complex, in *A. thaliana* × *A. arenosa*/*A. lyrata* ssp. *petraea* crosses. Although MADS-box type-I genes are often highly duplicated and subsequently redundant, *AGL40* and *AGL35* have one copy each in *A. thaliana*, *A. arenosa*, *A. lyrata* ssp. *petraea* and *A. halleri* (Paper I). Furthermore, both *AGL40* and *AGL35* are expressed in the chalazal endosperm (Bemer *et al.,* 2010b) and are downregulated during the transition from syncytial to endosperm cellularization (Paper I). Interestingly, *AGL40* and *AGL35* are upregulated in the incompatible diploid *A. thaliana* × *A. arenosa* hybrid, potentially contributing to the hybridization barrier (Walia *et al.,* 2009). *MEA* is important in endosperm development, and mutations in this gene have shown overproliferation of endosperm nuclei and a delay of cellularization (Grossniklaus *et al.,* 1998; Guitton *et al.,* 2004; Kinoshita *et al.,* 1999; Köhler *et al.,* 2003b). When considering gene expression patterns, copy number, and gene function, *AGL40*, *AGL35*, and *MEA* are ideal candidates for further examination of their potential to bypass the ePZB in the *A. thaliana* × *A. arenosa*/*A. lyrata* ssp. *petraea* crosses.

Selfed *A. thaliana* with the heterozygous *mea* mutation has a reduced germination rate of 60%, meaning that *mea* itself contributes to a decrease in seed viability. A comparable decline in germination rate is observed when using the same heterozygous *mea* mutant rather than wild-type *A. thaliana* in hybrid crosses to *A. arenosa*/*A. lyrata* ssp. *petraea*. Despite the delayed endosperm cellularization phenotype of the *mea* mutant, the expected ameliorative effect in the *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid and an exacerbated effect in *A. thaliana* × *A. arenosa* were not observed. This means the *mea* mutant failed to enhance or bypass the ePZB in these two hybrids (Paper II). In contrast to *mea*, *agl35-1* and *agl40-1* do not affect the germination rate of selfed *A. thaliana* and we could therefore use homozygous mutant lines. Hybrid crosses between *agl40-1* and *A. arenosa* showed no significant change when compared to crosses between wild-type *A. thaliana* and *A. arenosa*. However, in *A. lyrata* ssp. *petraea* hybrid crosses, the germination rate was significantly reduced when using *agl40-1* as a

maternal contributor. Crosses between *agl35-1* and *A. arenosa* have previously been shown to strengthen the hybrid barrier (Bjerkan *et al.,* 2019). This was confirmed in our study, showing a significant reduction in germination rate when using the *agl35-1* mutant compared to wild-type *A. thaliana* in hybrid crosses to *A. arenosa*. Interestingly, we found the opposite trend in hybrid crosses to *A. lyrata* ssp. *petraea*, with a significant increase in germination rate when using the *agl35-1* mutant compared to wild-type *A. thaliana* (Paper II). As discussed below, a different expression pattern of *AGL35* before and after endosperm cellularization in the parental species, may explain the different effects the *agl35* mutant has on the two *Arabidopsis* hybrids.

Although the *agl35-1* mutant could be used to enhance or bypass the ePZB phenotype in *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea* crosses, respectively, predicting the effects of single mutants based on their regulation and function in *A. thaliana* is difficult. Imprinting can be perturbed in hybrids, exemplified by *PHE1*, a PEG in *A. thaliana* which shifts to a MEG in the *A. thaliana* × *A. arenosa* hybrid (Josefsson *et al.,* 2006). Similarly, *AGL36* shifts from a MEG in *A. thaliana* to a BEG in the *A. thaliana* × *A. arenosa* hybrid (Bjerkan *et al.,* 2019). However, further knowledge about the copy number and expression of these genes in all parental species could make predictions more precise.

7. Temperature has a strong antagonistic effect and potential as an ePZB-adjusting factor

The optimal temperature range for *A. thaliana* is between 8 and 28°C, and different laboratories use different temperatures for experimental work on this species (Lloyd *et al.,* 2018). Although temperatures within the optimal range have no detrimental effect on *A. thaliana*, it will affect developmental progress. In different studies on the *A. thaliana* × *A. arenosa* hybrids, the crossed plants have been exposed to either 18°C or 22°C, resulting in a discrepancy in germination rates that cannot be explained by genetic differences (Bjerkan *et al.,* 2019; Burkart-Waco *et al.,* 2015; Walia *et al.,* 2009). A significant decrease in germination rate was for instance observed when the *A. thaliana* × *A. arenosa* cross was exposed to 22°C compared to 18°C (Bjerkan *et al.,* 2019). These results warranted further examination of a more extreme temperature effect on the ePZB in both hybrid crosses. Initially we investigated the temperature effect at 8°C and 28°C, revealing almost no germination in either *A. thaliana* × *A. arenosa* or *A. thaliana* × *A. lyrata* ssp. *petraea*. This observation suggests that these

temperatures cause stress in the hybrid seeds adding to the effect of the ePZB. Using 14°C and 26°C, we observed a strong significant difference in both hybrids. In line with previous research, the *A. thaliana* × *A. arenosa* hybrid seeds exhibited a significant decrease in germination rate when exposed to increased temperature (Figure 8) (Bjerkan *et al.,* 2019). Interestingly, the opposite was found for *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds, which had increased germination rate with increased temperature (Figure 8). This suggests that decreasing temperature during hybrid seed development can bypass an ePZB exhibiting delayed cellularization and that increasing temperature can bypass an ePZB exhibiting precocious cellularization. The antagonistic effect of temperature on hybrid seeds with opposite endosperm phenotypes suggests that the ePZB is controlled by a genetic mechanism sensitive to environmental conditions.

The temperature effect might be explained by the obstruction of abscisic acid (ABA) breakdown in the desiccating endosperm in cooler temperatures (Chen *et al.,* 2021). As a response to dehydration-stress, there is an elevation in the levels of ABA and this response has been triggered in *A. thaliana* inter-ploidy crosses with paternal excess, which exhibit delayed endosperm cellularization. It was also found that endosperm cellularization in the *A. thaliana* inter-ploidy seeds was induced by increasing ABA levels, resulting in improved seed viability (Xu *et al.,* 2023). Higher ABA levels in cooler temperatures could therefore explain why the *A. thaliana* × *A. arenosa* hybrid seeds have higher germination rates at 14°C compared to 26°C.

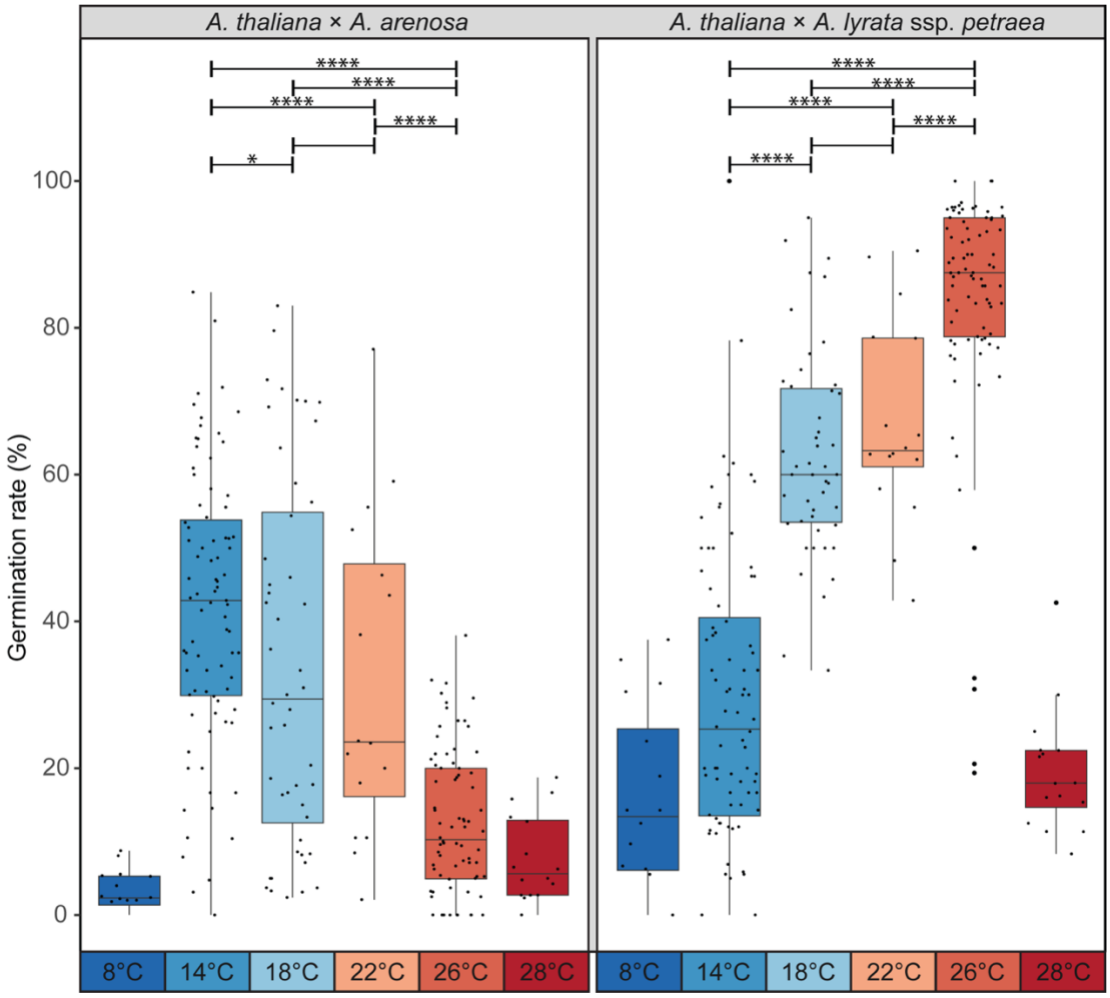


Figure 8: Temperature affects the germination rate of hybrid seeds. Germination rate of *A. thaliana* (Col-0) × *A. arenosa* (*A.a.*) hybrid seeds at 8°C, 14°C, 18°C, 22°C, 26°C and 28°C. Biological replicates (siliques): 8°C, n = 16 (811 seeds); 14°C, n = 83 (3243 seeds); 18°C, n = 48 (2610 seeds); 22°C, n = 16 (796 seeds); 26°C, n = 78 (2713 seeds); 28°C, n = 16 (663 seeds). Germination rate of *A. thaliana* (Col-0) × *A. lyrata* (*A.l.*) hybrid seeds at 8°C, 14°C, 18°C, 22°C, 26°C and 28°C. Biological replicates (siliques): 8°C, n = 16 (345 seeds); 14°C, n = 84 (1676 seeds); 18°C, n = 47 (1540 seeds); 22°C, n = 16 (533 seeds); 26°C, n = 85 (2043 seeds); 28°C, n = 16 (699 seeds). Box plot contains scattered data points representing germination rates observed per silique. Outliers are plotted as large data points. Significance is indicated for the comparisons between 14°C, 18°C, 22°C and 26°C within each hybrid (Welch's t-test: ***P* ≤ 0.01; ****P* ≤ 0.001; *****P* ≤ 0.0001).

Another possible explanation considers the high-temperature induced *PHYTOCHROME INTERACTING FACTOR 4* (*PIF4*) (Koini *et al.,* 2009). *PIF4* is upregulated when exposed to increased temperature in *A. thaliana*, which results in increased auxin biosynthesis and signaling that indirectly results in induction of *EXPANSIN* genes (Koini *et al.,* 2009; Quint *et al.,* 2016). Furthermore, several *EXPANSIN* genes are involved in the syncytium formation in *A. thaliana* roots upon nematode infection (Wieczorek *et al.,* 2006). *PIF4* and the *EXPANSIN* genes involved in syncytial formation are all expressed in the seed (Mergner *et al.,* 2020; Nakabayashi *et al.,* 2005), suggesting that increased temperature could result in a prolonged syncytial stage of the endosperm if regulated by these genes. Batista *et al.* (2019b)

showed that increased auxin biosynthesis inhibits endosperm cellularization, which further supports this hypothesis. An increased temperature could therefore prolong the syncytial phase, counteracting the precocious endosperm phenotype in the *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds, while exacerbating the delayed endosperm cellularization *A. thaliana* × *A. arenosa* hybrid seeds. If this is a conserved mechanism, one can use shifts in temperature to overcome the ePZB in other hybrid crosses.

8. Both environmental and genetic factors affect the hybrid barrier

The ePZB in *Arabidopsis* is influenced by various factors, both genetic and environmental, as shown in this work and previous literature, and summed up in Figure 9. Parent-of-origin specific effects have been demonstrated by variation in the genetic background, caused by choice of accessions, different ploidies, or single-gene mutations. In addition, temperature has a significant effect on the ePZB and can be used to both enhance or ameliorate the hybrid barrier strength. Furthermore, accession, temperature, and the single-gene mutant *agl35-1* demonstrate robust and antagonistic effects on the two hybrids, *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea*.

Identifying single mutants, which have an effect on the ePZB, has proven difficult, and several MADS-box type-I mutants have been investigated. Expression of MADS-box type-I genes has been examined in hybrid seeds from the incompatible homoploid *A. thaliana* × *A. arenosa* cross and the compatible tetraploid *A. thaliana* × diploid *A. arenosa* cross. This showed that several MADS-box type-I genes were upregulated in the seeds from the incompatible cross, suggesting that these genes might be responsible for the hybrid barrier (Walia *et al.,* 2009). However, knock-out mutants of the genes showed no change in germination rate of seeds from the homoploid *A. thaliana* × *A. arenosa* cross, with the exception of *agl35* (Bjerkan *et al.,* 2019). It is also challenging to extrapolate the effects of single mutant genes, such as *agl35*, to other species because of difficulties in identifying proper orthologs, particularly in genes with high duplication rates. Knowledge about accession effects is also difficult to transfer to other hybrid crosses and would need to be individually assessed for each species combination.

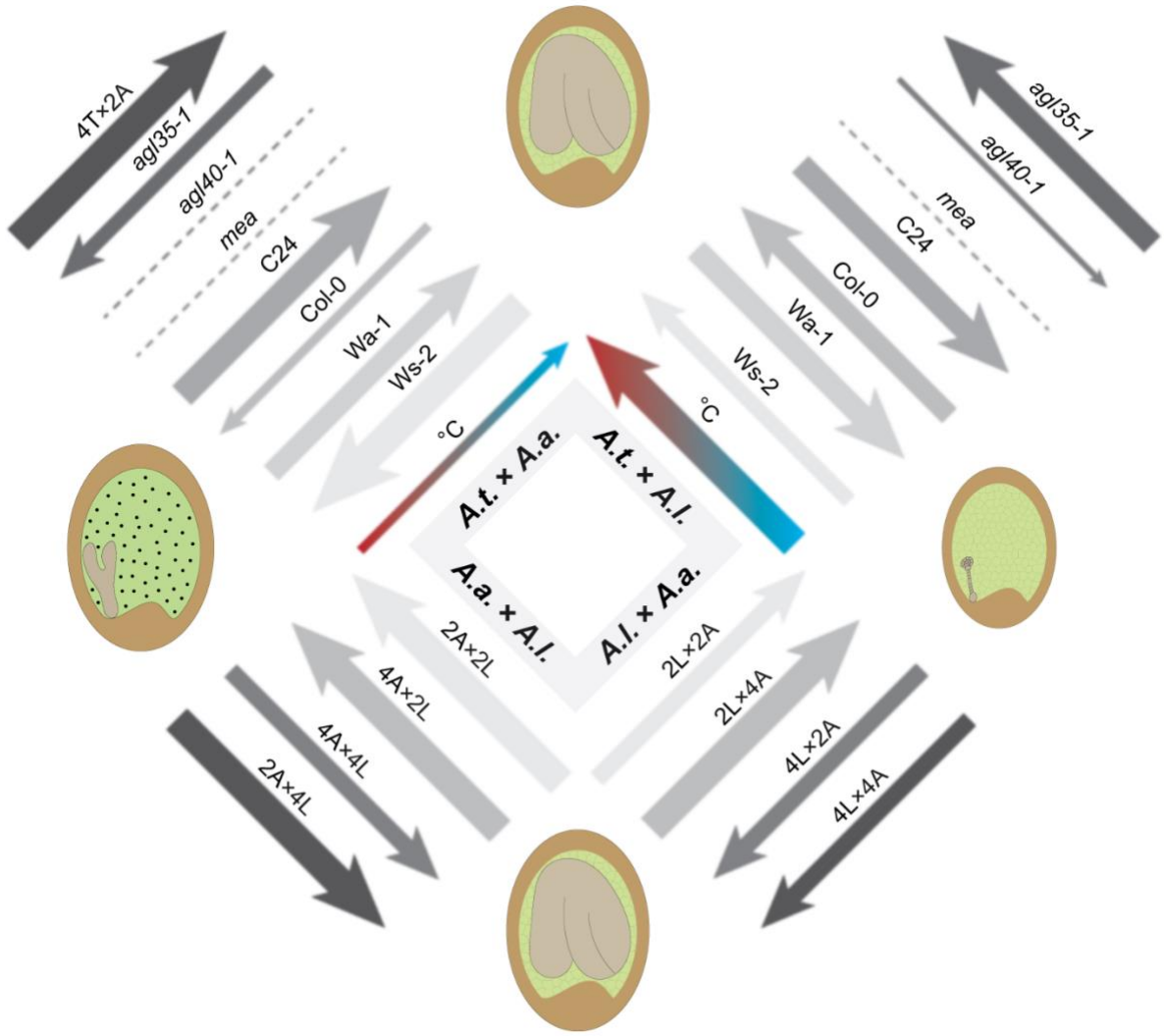


Figure 9: Different factors affecting the endosperm-based post-zygotic hybridization barrier in *Arabidopsis* hybrids. Dashed lines denote insignificant effects, while arrows denote significant effects, pointing towards the direction of influence on the hybrid seed phenotype. The width of the arrow reflects the magnitude of these effects. Colored arrows denote temperature gradient, with red being warmer and blue being colder. In *A. thaliana* (*A.t.*/T) × *A. arenosa* (*A.a.*/A) hybrid seeds, *agl35-1*, Col-0, and Ws-2 result in an exacerbated effect of the delayed endosperm barrier phenotype, whereas decreased temperature, tetraploidization of *A.t.* (Josefsson *et al.,* 2006; Walia *et al.,* 2009), C24 and Wa-1 ameliorate the effect. In *A.t.* × *A. lyrata* ssp. *petraea* (*A.l.*/L) hybrid seeds, *agl40-1*, C24, and Wa-1 enhance the frequency of precocious endosperm cellularization, whereas increased temperature, *agl35-1*, Col-0, and Ws-2 ameliorate the effect. In *A.l.* × *A.a.*, the diploid *A.l.* is incompatible with both the diploid and tetraploid *A.a.*, whereas the tetraploid *A.l.* is compatible with both the diploid and tetraploid *A.a.* In *A.a.* × *A.l.* hybrid seeds, both diploid and tetraploid *A.a.* are incompatible with diploid *A.l.*, and conversely the diploid and tetraploid *A.a.* are compatible with tetraploid *A.l.* (Lafon-Placette *et al.,* 2017).

In contrast to single mutant genes and accession effects, the effect of temperature manipulation on the ePZB has a much higher potential to be transferred even to species distantly related to *Arabidopsis*. Another effective tool for affecting the barrier is varying the ploidy of the parental species, which has been thoroughly tested in reciprocal *A. arenosa* × *A. lyrata* ssp. *petraea* crosses, in addition to the *A. thaliana* × *A. arenosa* cross. Change of ploidy has resulted in an effective bypass or strengthening of the hybridization barrier and is transferable to other diploid species, through the production of synthetic tetraploid strains.

9. MADS-box type-I expression levels at different developmental time points may predict ePZBs

Having knowledge on duplication or presence of MADS-box type-I genes in species exhibiting ePZB can contribute to identifying promising candidates to overcome the barrier. Although there is much information available on MADS-box type-I genes in *A. thaliana*, there is little information available for *A. arenosa* and *A. lyrata* ssp. *petraea*. It is therefore difficult to predict the effects of single-gene mutations in hybrids between *A. thaliana* and *A. arenosa* or *A. lyrata* ssp. *petraea*. To acquire information on MADS-box type-I genes in *A. arenosa* and *A. lyrata* ssp. *petraea*, high-quality genome sequences are required. Current available whole genome resources are of varying quality, and stem from other populations or subspecies that vary genetically from our populations, which is not optimal when examining e.g. gene copy numbers. This prompted us to perform de-novo whole-genome sequencing of *A. arenosa* and *A. lyrata* ssp. *petraea* individuals from our populations. This included one individual of *A. arenosa* from Pusté Pole in Slovakia and one individual of *A. lyrata* ssp. *petraea* from Pernitz in Austria (Figure 5), with estimated genome sizes of 201 Mbp and 179 Mbp, respectively, confirmed by flow-cytometry data (Paper I). The assembled sequence lengths were 188 Mbp for *A. lyrata* ssp. *petraea* and 153 Mbp for *A. arenosa* with chromosome-length scaffolds (Paper I).

The predicted proteomes of *A. arenosa* and *A. lyrata* ssp. *petraea* were compared to *A. thaliana* (Araport11), showing more than 21,000 shared orthogroups and very few species-specific orthogroups (Figure 10). As expected, orthogroups shared between the two recently split sister species, *A. arenosa* and *A. lyrata* ssp. *petraea*, were two to three times higher than shared orthogroups between *A. thaliana* and *A. arenosa* or *A. lyrata* ssp. *petraea*.

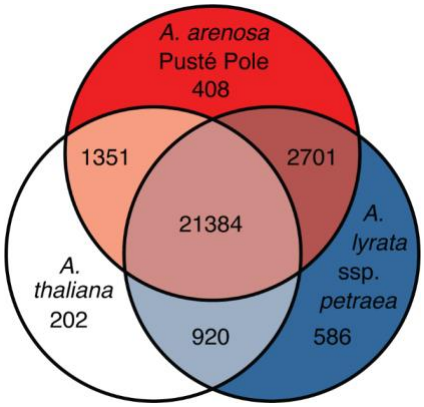


Figure 10: Venn diagram showing overlap of predicted orthologs between *Arabidopsis lyrata* ssp. *petraea, A. arenosa* Pusté Pole, and *A. thaliana*. Orthogroups were predicted using OrthoFinder. A majority of orthologs are common for all three species; in comparison few species-specific genes were identified.

The assembled genomes and proteome predictions were used, in combination with other available genomes of sufficient quality, in a detailed study of MADS-box TFs within the *Arabidopsis* genus, including *A. arenosa* Pusté Pole, *A. lyrata* ssp. *petraea* Pernitz, *A. arenosa* Strecno, *A. lyrata* ssp. *lyrata* and *A. halleri*. A phylogenetic gene tree, constructed based on all the predicted MADS-box genes together with MADS-box genes from *A. thaliana*, demonstrated large differences in duplications of several MADS-box type-I genes. Differences in copy number were not only found between species, but also between subspecies or populations of the same species. The *PHE2* clade (Mγ-1b) demonstrated a species-specific duplication with two copies in both *A. lyrata* subspecies, whereas *A. thaliana* and both *A. arenosa* populations only had one copy (Paper I). The *PHE1* clade (Mγ-1a) showed a subspecies-specific duplication, with six copies in *A. lyrata* ssp. *petraea*, and three copies in *A. lyrata* ssp. *lyrata*. A population-specific expansion could be found in the *DIA* clade (Mα-4), with one copy in *A. arenosa* Strecno and two copies in *A. arenosa* Pusté Pole (Paper I).

Duplications of MADS-box type-I genes have previously been suggested to arise because of small-scale duplication events, rather than whole-genome duplications (De Bodt *et al.*, 2003; Nam *et al.*, 2004; Pařenicová *et al.*, 2003). Such duplications will often be silenced or lost from the genome, because of low selection pressure due to redundancy. However, some duplications result in neofunctionalization and it has been suggested that redundancy amongst the MADS-box type-I genes may not be as common as previously expected (Bemer *et al.*, 2010a). The observed differences in ortholog-copies of MADS-box type-I genes between populations of the same species may explain how using different accessions or populations of the same species can have such a large effect on the ePZB when crossed to other species.

To further examine MADS-box type-I genes, we performed an RNAseq analysis of *A. arenosa* and *A. lyrata* ssp. *petraea* seeds at nine and 15 DAP, corresponding to before and after endosperm cellularization in these species. The expression of MADS-box type-I genes was analyzed at each stage, to see whether any of the subgroups, Mα, Mβ, and Mγ, changed after cellularization of the endosperm. Using available RNAseq data from a previous study (Bjerkan *et al.,* 2019), we did the same analysis with *A. thaliana* at six and nine DAP, which correspond to before and after endosperm cellularization in this species. Interestingly, we found that Mγ and half of the Mα genes were highly expressed before cellularization and declined substantially after cellularization in both *A. thaliana* and *A. arenosa*, whereas *A. lyrata* ssp. *petraea* showed increased expression of several Mγ and Mα genes after cellularization (Figure 11). This expression pattern is intriguing, not just because of the distinction between *A. lyrata* ssp. *petraea* to *A. thaliana*/*A. arenosa*, but also due to the established interaction patterns between the Mγ and Mα subclades. The Mα subclade can be divided into two monophyletic clades, consisting of Mα-1 to Mα-11 and Mα-12 to Mα-18. The genes in the first group, Mα-1 to Mα-11, interact solely with Mγ proteins, except for Mα-5 and Mα-9 genes, which have shown protein interaction with the Mβ subclade. The genes, which primarily interact with Mγ proteins, are expressed during early endosperm development. The genes of the second group, Mα-12 to Mα-18, interact primarily with Mβ proteins, although Mα-13, Mα-14, Mα-17, and Mα-19 genes have also shown Mγ interactions (de Folter *et al.,* 2005; Qiu & Köhler, 2022).

AGL35, which showed antagonistic effects in the *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea* hybrids when mutated in the *A. thaliana* background, is an Mγ gene. In *A. thaliana* and *A. arenosa AGL35* expression significantly decreases after endosperm cellularization. In contrast, *AGL35* expression in *A. lyrata* ssp. *petraea* slightly increases after endosperm cellularization. This difference in regulation may explain the observed differences in hybrid seed germination rates when *A. thaliana AGL35* is mutated (Figure 11). However, the same expressional pattern is observed with *AGL40* (Figure 11), which only resulted in a significant decline in germination rate for the *A. thaliana* × *A. lyrata* ssp. *petraea* cross when mutated in the *A. thaliana* background. As *AGL35* and *AGL40* are both single-copy genes in all the examined *Arabidopsis* species, redundancy cannot explain the different effects on the ePZB. Typically MADS-box Type-I TFs heterodimerize, and while interaction between Mβ and Mγ is considered unlikely, Mα interacts with both Mβ and Mγ (de Folter *et al.,* 2005; Qiu & Köhler, 2022). The flexibility of Mα TFs can potentially allow Mα orthologs to bind more easily to their interaction partners when *AGL40* is mutated, and

therefore cause a weaker effect when mutated, compared to the Mγ gene *AGL35*. Furthermore, interaction between AGL35 and AGL62 has been shown in a yeast-to-hybrid study (Trigg *et al.,* 2017) and both are expressed in the chalazal endosperm (Bemer *et al.,* 2010b). This is interesting, because AGL62 is important in suppressing cellularization during the syncytial stage (Kang *et al.,* 2008). Suppressing the function of AGL62 should therefore exacerbate the precocious endosperm phenotype in the *A. thaliana* × *A. lyrata* ssp. *petraea* cross and give a decreased germination rate. However, the *agl35-1* mutant results in an increased germination rate in the *A. thaliana* × *A. lyrata* ssp. *petraea* cross. Another yeast-to-hybrid study showed that AGL80, which together with *DIA* regulates a set of genes during central cell development (Bemer *et al.,* 2008), also interacts with AGL62 (de Folter *et al.,* 2005). A possible explanation could be that when interacting with AGL35, the function of AGL62 is suppressed, while it is active when interacting with AGL80. However, this does not explain why *AGL35* has a higher expression before cellularization than after in both *A. thaliana* and *A. arenosa*.

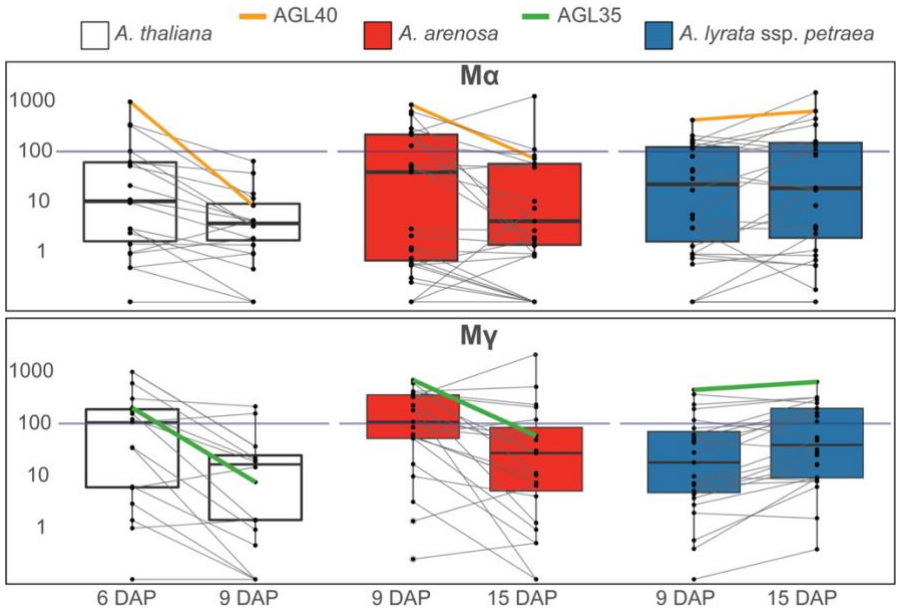


Figure 11: Expression of Mγ and Mα genes before and after cellularization in *Arabidopsis thaliana*, *A. arenosa* and *A. lyrata* ssp. *petraea*. Top: Box plot of Mα-gene expression based on mean read counts normalized with DESeq2 (Love *et al.,* 2014) on the y-axis for *A. thaliana* (white), *A. arenosa* (red), and *A. lyrata* ssp. *petraea* (blue). Bottom: An equivalent box plot for Mγ-genes. RNA was isolated from whole seeds at six and nine days after pollination (DAP) for *A. thaliana*, corresponding to before and after cellularization of the endosperm, and at nine and 15 DAP for *A. arenosa* and *A. lyrata* ssp. *petraea*, which coincides with the same developmental stages. *AGL40* and *AGL35* are highlighted with orange and green, respectively.

AGL35, which showed antagonistic effects in the *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea* hybrids when mutated in the *A. thaliana* background, is an

Mγ gene. In *A. thaliana* and *A. arenosa AGL35* expression significantly decreases after endosperm cellularization. In contrast, *AGL35* expression in *A. lyrata* ssp. *petraea* slightly increases after endosperm cellularization. This difference in regulation may explain the observed differences in hybrid seed germination rates when *A. thaliana AGL35* is mutated (Figure 11). However, the same expressional pattern is observed with *AGL40* (Figure 11), which only resulted in a significant decline in germination rate for the *A. thaliana* × *A. lyrata* ssp. *petraea* cross when mutated in the *A. thaliana* background. As *AGL35* and *AGL40* are both single-copy genes in all the examined *Arabidopsis* species, redundancy cannot explain the different effects on the ePZB. Typically MADS-box Type-I TFs heterodimerize, and while interaction between Mβ and Mγ is considered unlikely, Mα interacts with both Mβ and Mγ (de Folter *et al.,* 2005; Qiu & Köhler, 2022). The flexibility of Mα TFs can potentially allow Mα orthologs to bind more easily to their interaction partners when *AGL40* is mutated, and therefore cause a weaker effect when mutated, compared to the Mγ gene *AGL35*. Furthermore, interaction between AGL35 and AGL62 has been shown in a yeast-to-hybrid study (Trigg *et al.,* 2017) and both are expressed in the chalazal endosperm (Bemer *et al.,* 2010b). This is interesting, because AGL62 is important in suppressing cellularization during the syncytial stage (Kang *et al.,* 2008). However, suppressing the function of AGL62 should therefore exacerbate the precocious endosperm phenotype in the *A. thaliana* × *A. lyrata* ssp. *petraea* cross and give a decreased germination rate, while the *agl35-1* mutant results in an increased germination rate in the *A. thaliana* × *A. lyrata* ssp. *petraea* cross. Another yeast-to-hybrid study showed that AGL80, which together with *DIA* regulates a set of genes during central cell development (Bemer *et al.,* 2008), also interacts with AGL62 (de Folter *et al.,* 2005). A possible explanation could be that when interacting with AGL35, the function of AGL62 is suppressed, while it is active when interacting with AGL80. However, this does not explain why *AGL35* has a higher expression before cellularization than after in both *A. thaliana* and *A. arenosa*.

10. Genomic imprinting in *Draba nivalis* is dominated by maternally expressed genes

To investigate genomic imprinting in *D. nivalis*, we used two populations originating from Alaska and Norway. To exclude developmental differences between the parental populations or in the intraspecific hybrids, we investigated seeds at seven and 12 DAP (Paper III). Both in seeds from selfed parental populations and in reciprocal interpopulation crosses, the embryo at seven DAP was primarily in the globular stage with uncellularized endosperm.

A *D. nivalis* transgenic line carrying a pAtTE1-GFP reporter, expressed only after onset of cellularization in *A. thaliana*, was observed from 10 DAP in *D. nivalis*. Having a globular embryo and uncellularized endosperm at seven DAP, expression of the pAtTE1-GFP reporter from 10 DAP suggests this reporter works in a similar manner in *D. nivalis* as *A. thaliana*, expressing after onset of cellularization. In our analysis, we therefore generated whole seed transcriptomes from seven DAP seed stages, representing a globular embryo stage with syncytial endosperm (Paper III). RNAseq data from both homozygous parents, *D. nivalis* AKB17-2 from Norway and *D. nivalis* 8-4 from Alaska (Figure 5), in addition to reciprocal intraspecific crosses between these populations, were used to investigate parent-of-origin allele-specific expression. To examine variation between the two populations, the intraspecific hybrids, and the biological replicates, we performed a principal component analysis (PCA). Biological replicates showed high similarity and the parental lines were clearly separated along the first principal component (PC1), which explained 86% of the total variation. The intraspecific hybrids had scores similar to the maternal parent along PC1, but were separated from the maternal parent along the PC2 axis, explaining 6% of the variation. Overall, this indicates that the intraspecific hybrids show high similarity to the maternal parent. As RNA was isolated from whole seeds, seed-coat expressing genes will be present in the dataset. Therefore, a general seed coat (GSC) filter based on enriched seed coat genes in *A. thaliana* was applied to the dataset (Belmonte *et al.,* 2013; Hornslien *et al.,* 2019).

To identify imprinted genes, we first established which read pairs had single nucleotide polymorphisms (SNPs) or insertions/deletions (InDels) between the parental populations (informative reads). Then informative read pairs from the intraspecific hybrid seeds were mapped to the parental transcriptomes, to establish parent-of-origin specific gene expression. The resulting intraspecific hybrid dataset was plotted for distribution of maternal:paternal Fold-change (log2 FC) using a histogram plot (Figure 12 B). Interestingly, a large bias towards maternal expression was found, although seed coat enriched genes had been removed by the above-mentioned GSC-filter and all maternal reads were halved to compensate for the 2:1 maternal:paternal contribution to the endosperm.

Due to the maternal bias observed in the whole dataset (Figure 12 B), a higher threshold was set for MEGs (log2 FC = 3) compared to PEGs (log2 FC = -1). Furthermore, we excluded genes with population-specific expression by focusing on imprinted genes found in both cross directions. This resulted in numerous MEGs (4115) and very few PEGs (10) (Figure 12 A).

Given the high number of MEGs identified, we considered that maternal bias could have been caused by self-fertilization, as each biological replicate consists of several seeds derived from manual pollination of the stigma, and potentially could include self-fertilized seeds. However, in the case of self-fertilization, some seeds would still be intraspecific hybrids and the ratio of intraspecific hybrid seeds to self-fertilized seeds should differ in each biological replicate, resulting in large variation between them, which is not the case. In addition, significant PEGs with almost no expression from the maternal parent were identified, altogether suggesting that self-fertilization is not causing the maternal expression bias in the dataset. It is also possible that not all GSC-genes were removed, as this filtering step was based on the expression pattern of *A. thaliana* orthologs. However, we anticipate that the majority of GSC-genes were removed with this filter and the substantial maternal expression observed is difficult to explain as being the result of a few remaining GSC-genes.

Comparing identified MEGs from *D. nivalis* with orthologous MEGs identified in other Brassicaceae species, we found that approximately 10% of *D. nivalis* MEGs overlapped with orthologous MEGs reported by one or more Brassicaceae. This overlap is considerably lower compared to what is found between other Brassicaceae species, which show overlaps of MEGs between 30-75% (Paper III). Surprisingly, despite identifying only ten PEGs, three overlapped with other Brassicaceae species, demonstrating a moderately high degree of conserved PEGs, comparable to what is found between studies within the same species (Paper III).

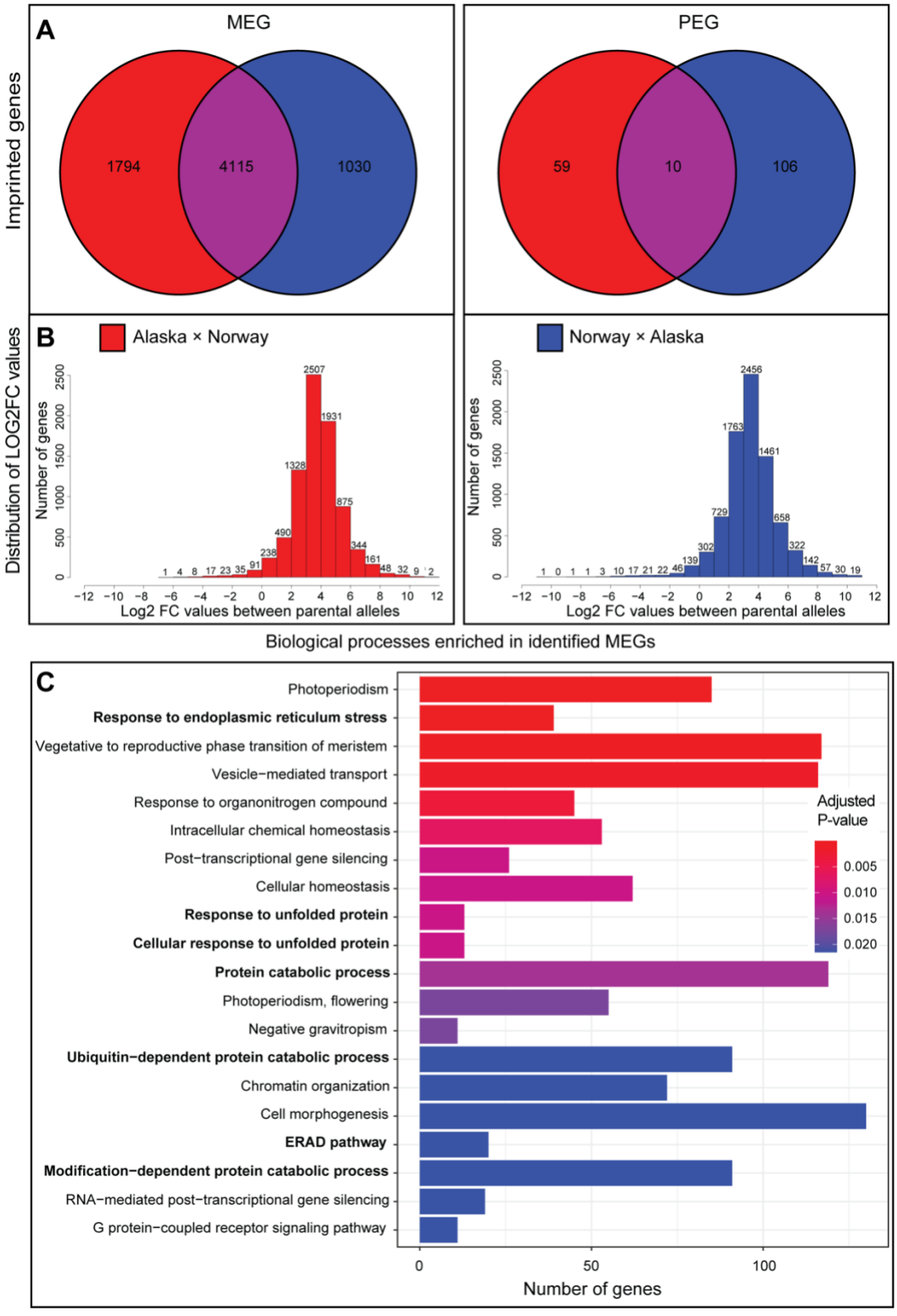


Figure 12: Maternally expressed genes (MEGs) and paternally expressed genes (PEGs) identified in *Draba nivalis* together with a gene ontology analysis of the MEGs. (A) Venn diagrams showing MEGs and PEGs of intraspecific *D. nivalis* hybrids. **(B)** Histograms showing the distribution of log2 fold change (FC) values in each of the two *D. nivalis* hybrids. X-axis: Negative values indicate paternal expression bias and positive values indicate maternal expression bias. **(C)** Biological processes enriched in identified MEGs, based on *A. thaliana* orthologs. Adjusted P-value was calculated using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995) and illustrated using a heatmap, with red indicating more significantly enriched compared to blue. GO-terms associated with potential protein-related stress are written in bold. Crosses are indicated as maternal × paternal.

To further investigate the identified MEGs, a gene ontology (GO) analysis was performed, which showed an enrichment of processes involved in protein modification and folding, including genes involved in the unfolded protein response (UPR) (Figure 12 C). The UPR is activated if the processing capacity of the endoplasmatic reticulum is overloaded, resulting in accumulation of unfolded proteins, and is therefore an indication of stress (Fanata *et al.,* 2013). Nowak *et al.* (2020) found that genes involved in protein modification and other stress-related processes were highly duplicated and strongly expressed in *D. nivalis*, and hypothesized that the strong expression and high duplication rate of these genes could be an adaptation to the harsh arctic environment (Nowak *et al.,* 2020). In support of the co-adaptation theory, which states that maternal inherited alleles in the offspring interact favorably with the maternal genotype (Wolf & Hager, 2006), species exposed to stressful environments could benefit from a strong maternal expression bias to make the progeny pre-adapted to the environment of the maternal plant (Veselá *et al.,* 2021). Taken together, this could indicate that the maternal bias observed in *D. nivalis* is an environmental adaptation.

The mating system of plants has been shown to have an impact on genomic imprinting in *Arabidopsis* and *Capsella*, reflected in a higher number and expression level of PEGs in outbreeding species compared to selfing species within the same genus. However, the number and expression of MEGs did not show any significant change (Klosinska *et al.,* 2016; Lafon-Placette *et al.,* 2018). The numbers of identified MEGs in our study are 5-fold or higher than other selfing species that have been investigated for imprinted expression in the endosperm. The highest number of identified MEGs previously reported was 781 in the selfing *A. thaliana* (Del Toro-De León & Köhler, 2019), which implies that the mating system alone is insufficient in explaining the high number of identified imprinted genes in *D. nivalis*.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The overarching aim of this study was to investigate mechanisms contributing to ePZB, by examining genetic and environmental factors causing and affecting the barrier. We have shown that the expression of MADS-box type-I genes before and after endosperm cellularization differ significantly between *A. arenosa* and *A. lyrata* ssp. *petraea* and we have identified conserved MADS-box type-I genes with only one ortholog in different *Arabidopsis* species (Paper I). Furthermore, we confirm that *A. thaliana* × *A. arenosa* hybrid seeds exhibit delayed endosperm cellularization and show that *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds have precocious endosperm cellularization. These endosperm phenotypes can be ameliorated or exacerbated using different single-gene mutations, different temperature conditions or different accessions/populations. We found the observed effect on endosperm cellularization and ultimately seed survival to be diametrically opposed in the two *Arabidopsis* hybrid crosses (Paper II).

According to the WISO hypothesis, crosses between the maternal selfer (*A. thaliana*) and the paternal outbreeder (*A. arenosa*/*A. lyrata* ssp. *petraea*) should cause a paternal excess (Brandvain & Haig, 2005). However, this is only observed for the *A. thaliana* × *A. arenosa* hybrid cross, suggesting the two outbreeders have different effective ploidy. A higher effective ploidy in *A. arenosa* compared to *A. lyrata* is supported by the endosperm phenotypes observed in reciprocal crosses between the two species (Lafon-Placette *et al.,* 2017), showing delayed endosperm cellularization with paternal *A. arenosa* and precocious endosperm cellularization with paternal *A. lyrata*. We also found that expression of Mα and Mγ MADS-box type-I genes before and after cellularization was oppositely regulated in *A. arenosa* and *A. lyrata* ssp. *petraea*, respectively (Figure 11). Whether these differences can be linked to the observed ePZB when these two species are crossed to *A. thaliana* remains to be determined. However, the genome resource reported here (Paper I) facilitates the analysis of parent-of-origin specific expression in *A. arenosa* x *A. lyrata* ssp. *petraea* hybrid seeds, and greatly enhances the feasibility of such analysis.

Despite the high duplication rate in MADS-box type-I genes, some genes had one clear ortholog in all *Arabidopsis* species investigated, including *AGL35* and *AGL40*. Both genes showed significant effects on the hybrid seed germination rate when mutated in the maternal

A. thaliana background and crossed to paternal *A. lyrata* ssp. *petraea*. In addition, *agl35-1* showed an antagonistic effect, with increased germination rate of the *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds and decreased germination rate of the *A. thaliana* × *A. arenosa* hybrid seeds. In addition to the *agl35-1* mutant, we show that temperature and different *A. thaliana* accessions cause antagonistic effects in the *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid crosses. With increased temperature, the *A. thaliana* × *A. lyrata* ssp. *petraea* cross exhibited an almost complete bypass of the barrier, and we speculate that increased temperature causes a prolonged syncytial stage of the endosperm, which would explain why increased temperature is detrimental to the *A. thaliana* × *A. arenosa* hybrid. A thorough examination of the temperature-effect in other species with ePZB is necessary to conclude whether this is an effect specific for these *Arabidopsis* hybrids or the endosperm phenotype. Furthermore, differential expression analysis of hybrid seeds together with parental specific expression at two different temperatures from the three parental species, *A. thaliana*, *A. arenosa,* and *A. lyrata* ssp. *petraea*, and the two hybrids *A. thaliana* × *A. arenosa* and *A. thaliana* × *A. lyrata* ssp. *petraea*, would elucidate the genetic basis for the temperature effect. Using such a dataset, one could also investigate the regulation of the temperature-dependent *PIF4* gene and the downstream *EXPANSIN* genes, to see whether these are differentially expressed at this developmental time point.

When considering the germination rate of hybrid seeds, the accession effect correlates well with ploidy and effective ploidy. This is shown when using the tetraploid Wa-1 accession, which increases the maternal contribution and therefore decreases the paternal excess in the hybrid crosses between *A. thaliana* and paternal *A. arenosa* and increases the maternal excess when crossed to paternal *A. lyrata* ssp. *petraea*. In consistance with this, we showed an increased germination rate in the *A. thaliana* × *A. arenosa* hybrid seeds and a decreased germination rate in the *A. thaliana* × *A. lyrata* ssp. *petraea* hybrid seeds. However, the diploid C24 accession showed a stronger effect on the germination rate in these two *Arabidopsis* hybrids than Wa-1, indicating that C24 has a higher effective ploidy than the tetraploid Wa-1. This suggests that differences in accession background can have larger effects on the ePZB than the ploidy and underscores that the ePZB is intricately controlled by several genetic factors. To better understand the impact of different accessions on the ePZB, conducting a differential expression analysis to examine variations in gene expression before and after

endosperm cellularization between different accessions could yield valuable insights, especially with a focused study on MADS-box type-I expression.

In contrast to indications from previous studies, we show that hybrids between the selfing *Draba* species show no signs of an ePZB and identified a high number of MEGs in *D. nivalis*, deviating markedly from what has been reported from previous studies on genomic imprinting (Paper III). The absence of an ePZB in selfing *Draba* species was not unexpected considering previous hypotheses that differences in mating system may cause an ePZB in closely related species (Brandvain & Haig, 2005; Lafon-Placette *et al.,* 2018; Petrén *et al.,* 2023). However, the cross between the selfing *A. thaliana* and the outbreeding *A. lyrata* ssp. *petraea*, contradicts this theory with the precocious endosperm phenotype, which suggests maternal excess rather than paternal excess. This means that mating systems alone cannot explain or predict the outcome in hybrid crosses.

We identified a large number of MEGs in *D. nivalis* in the genomic imprinting study, 3-4 times higher than what is found in any other study. The numerous MEGs identified cannot be explained by the mating system alone, as similar results have not been found in other selfing species. However, enrichment of protein-stress related GO-terms in the identified MEGs, which previously have been indicated as arctic-related adaptation (Nowak *et al.,* 2020), suggests that this could be related to the arctic environment and growth conditions. Future research on genomic imprinting in other selfing arctic plant species could potentially confirm or reject the hypothesis that the high maternal bias reflects a combination of mating system and environmental adaptation.

Since genomic imprinting generally shows low conservation across species, conducting an imprinting study in arctic vs non-arctic populations, e.g. in the selfing *A. alpina* from arctic populations and outbreeding *A. alpina* populations from central Italy or Greece (Laenen *et al.,* 2018), would give valuable insight into whether selfing and arctic adaptations causes an increase in the number of MEGs. To uncouple environmental adaptations from mating systems, one could conduct an imprinting study in an arctic outbreeding species. However, an outbreeding arctic species with a sequenced genome is currently not available (Colella *et al.,* 2020; Nowak *et al.,* 2020; Wullschleger *et al.,* 2015). Furthermore, performing a new genomic imprinting study in *D. nivalis* with endosperm nuclei sorting, could verify that maternal imprinted expression

originates from the endosperm alone. This could easily be done with the newly established pAtTE1-GFP reporter line in *D. nivalis* (Paper III).

In a wider perspective, the findings in this thesis provide evidence for a dynamic and complex endosperm-based post-zygotic hybridization barrier, closely associated with and impacted by regulation of MADS-box type-I genes before and after cellularization. We hypothesize that increased temperature can be used to bypass the ePZB when hybrid seeds exhibit precocious endosperm cellularization, and that decreased temperature can bypass the ePZB when hybrid seeds show delayed endosperm cellularization. If this holds true, it provides a unique opportunity in plant breeding research, allowing for introduction of genetic variation by hybridization by simple adjustments in temperature. Furthermore, we show that genetic factors such as ploidy and parental accessions with different effective ploidies affect the ePZB in a diametrical manner when crossing *A. thaliana* to *A. arenosa* or *A. lyrata* ssp. *petraea*, emphasizing the role of ploidy and effective ploidy on the ePZB. We also suggest that environmental adaptations and mating systems influence genomic imprinting, but this must be confirmed in other species systems and with higher tissue sampling accuracy, for instance nuclei sorting or single nuclei sequencing. Altogether, the findings in this study provide great potential for bypassing the ePZB, which can have a substantial impact in plant breeding and present a promising new way for sustainable food production.

REFERENCES

Akama, S., Shimizu-Inatsugi, R., Shimizu, K. K., & Sese, J. (2014). Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid *Arabidopsis*. Nucleic Acids Research, 42(6), e46–e46.

Al-Shehbaz, I. A., Beilstein, M. A., & Kellogg, E. A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Systematics and Evolution, 259(2), 89–120.

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martínez-Castilla, L., & Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proceedings of the National Academy of Sciences, 97(10), 5328–5333.

Aw, S. J., Hamamura, Y., Chen, Z., Schnittger, A., & Berger, F. (2010). Sperm entry is sufficient to trigger division of the central cell but the paternal genome is required for endosperm development in *Arabidopsis*. Development, 137(16), 2683–2690.

Bai, F., & Settles, A. M. (2015). Imprinting in plants as a mechanism to generate seed phenotypic diversity. Frontiers in Plant Science (Bd. 5, s. 780). https://www.frontiersin.org/article/10.3389/fpls.2014.00780

Bailey, C. D., Koch, M. A., Mayer, M., Mummenhoff, K., O'Kane, S. L., Jr, Warwick, S. I., Windham, M. D., & Al-Shehbaz, I. A. (2006). Toward a global phylogeny of the Brassicaceae. Molecular Biology and Evolution, 23(11), 2142–2160.

Bajon, C., Horlow, C., Motamayor, J. C., Sauvanet, A., & Robert, D. (1999). Megasporogenesis in *Arabidopsis thaliana* L.: an ultrastructural study. Sexual Plant Reproduction, 12(2), 99–109.

Batista, R. A., Figueiredo, D. D., Santos-González, J., & Köhler, C. (2019a). Auxin regulates endosperm cellularization in *Arabidopsis*. Genes & Development, 33(7–8), 466–476.

Batista, R. A., & Köhler, C. (2020). Genomic imprinting in plants—revisiting existing models. Genes & Development, 34(1–2), 24–36.

Batista, R. A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D. D., & Köhler, C. (2019b). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. eLife, 8, e50541.

REFERENCES

Beaulieu, J. M., O'Meara, B. C., Crane, P., & Donoghue, M. J. (2015). Heterogeneous rates of molecular evolution and diversification could explain the triassic age estimate for angiosperms. Systematic Biology, 64(5), 869–878.

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., Hashimoto, M., Fei, J., Harada, C. M., Munoz, M. D., Le, B. H., Drews, G. N., Brady, S. M., Goldberg, R. B., & Harada, J. J. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. Proceedings of the National Academy of Sciences of the United States of America, 110(5), E435-44.

Bemer, M., Gordon, J., Weterings, K., & Angenent, G. C. (2010a). Divergence of recently duplicated M{gamma}-type MADS-box genes in *Petunia*. Molecular Biology and Evolution, 27(2), 481–495.

Bemer, M., Heijmans, K., Airoldi, C., Davies, B., & Angenent, G. C. (2010b). An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. Plant Physiology, 154(1), 287–300.

Bemer, M., Wolters-Arts, M., Grossniklaus, U., & Angenent, G. C. (2008). The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in *Arabidopsis* ovules. The Plant Cell, 20(8), 2088–2101.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B, Statistical methodology, 57(1), 289–300.

Birkeland, S., Gustafsson, A. L. S., Brysting, A. K., Brochmann, C., & Nowak, M. D. (2020). Multiple genetic trajectories to extreme abiotic stress adaptation in arctic Brassicaceae. Molecular Biology and Evolution, 37(7), 2052–2068.

Birkeland, S., Slotte, T., Krag Brysting, A., Gustafsson, A. L. S., Rhoden Hvidsten, T., Brochmann, C., & Nowak, M. D. (2022). What can cold-induced transcriptomes of arctic Brassicaceae tell us about the evolution of cold tolerance? Molecular Ecology, 31(16), 4271–4285.

Bjerkan, K. N., Hornslien, K. S., Johannessen, I. M., Krabberød, A. K., van Ekelenburg, Y. S., Kalantarian, M., Shirzadi, R., Comai, L., Brysting, A. K., Bramsiepe, J., & Grini, P. E. (2019). Genetic variation and temperature affects hybrid barriers during interspecific hybridization. The Plant Journal: for Cell and Molecular Biology. https://doi.org/10.1111/tpj.14523.

Bohutínská, M., Handrick, V., Yant, L., Schmickl, R., Kolář, F., Bomblies, K., & Paajanen, P. (2021). De novo mutation and rapid protein (co-)evolution during meiotic adaptation in *Arabidopsis arenosa*. Molecular Biology and Evolution, 38(5), 1980–1994.

Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature, 422(6930), 433–438.

Brandvain, Y., & Haig, D. (2005). Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. The American Naturalist, 166(3), 330–338.

Brandvain, Y., & Haig, D. (2018). Outbreeders pull harder in a parental tug-of-war. Proceedings of the National Academy of Sciences of the United States of America, 115(45), 11354–11356.

Briskine, R. V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., & Shimizu, K. K. (2016). Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. Molecular Ecology Resources, 17(5), 1025–1036.

Brochmann, C. (1993). Reproductive strategies of diploid and polyploid populations of arctic *Draba* (Brassicaceae). Plant Systematics and Evolution, 185, 55–83.

Brochmann, C., Borgen, L., & Stedje, B. (1993). Crossing relationships and chromosome numbers of Nordic populations of *Draba* (Brassicaceae), with emphasis on the *D. alpina* complex. Nordic Journal of Botany, 13(2), 121–147.

Brochmann, C., Soltis, P. S., & Soltis, D. E. (1992). Recurrent formation and polyphyly of Nordic polyploids in *Draba* (Brassicaceae). American Journal of Botany, 79(6), 673–688.

Burkart-Waco, D., Josefsson, C., Dilkes, B., Kozloff, N., Torjek, O., Meyer, R., Altmann, T., & Comai, L. (2011). Hybrid incompatibility in *Arabidopsis* is determined by a multiple-locus genetic network. Plant Physiology, 158(2), 801–812.

Burkart-Waco, D., Ngo, K., Lieberman, M., & Comai, L. (2015). Perturbation of parentally biased gene expression during interspecific hybridization. PloS One, 10(2), e0117293.

Bushell, C., Spielman, M., & Scott, R. J. (2003). The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. The Plant Cell, 15(6), 1430–1442.

Böcher, T. W. (1966). Experimental and cytological studies on plant species: XI. North Atlantic tetraploids of the *Campanula rotundifolia* complex (Bd. 3). JSTOR.

REFERENCES

Cao, X., & Jacobsen, S. E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes. Proceedings of the National Academy of Sciences of the United States of America, 99 Suppl 4, 16491–16498.

Chan, S. W.-L., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C., & Jacobsen, S. E. (2004). RNA silencing genes control de novo DNA methylation. Science, 303(5662), 1336 LP – 1336.

Chaudhury, A. M., Ming, L., Miller, C., Craig, S., Dennis, E. S., & Peacock, W. J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences, 94(8), 4223–4228.

Chen, C., Begcy, K., Liu, K., Folsom, J. J., Wang, Z., Zhang, C., & Walia, H. (2016). Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. Plant physiology, 171(1), 606 LP – 622.

Chen, X., Yoong, F.-Y., O'Neill, C. M., & Penfield, S. (2021). Temperature during seed maturation controls seed vigour through ABA breakdown in the endosperm and causes a passive effect on DOG1 mRNA levels during entry into quiescence. The New Phytologist, 232(3), 1311–1322.

Ciubota-Rosie, C., Macoveanu, M., Fernández, C. M., Ramos, M. J., Pérez, A., & Moreno, A. (2013). *Sinapis alba* seed as a prospective biodiesel source. Biomass and Bioenergy, 51, 83–90.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal: For Cell and Molecular Biology, 16(6), 735–743.

Colella, J. P., Talbot, S. L., Brochmann, C., Taylor, E. B., Hoberg, E. P., & Cook, J. A. (2020). Conservation genomics in a changing arctic. Trends in Ecology & Evolution, 35(2), 149–162.

Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M. M., & Colombo, L. (2008). *AGL23*, a type I MADS-box gene that controls female gametophyte and embryo development in *Arabidopsis*. The Plant Journal: For Cell and Molecular Biology, 54(6), 1037–1048.

Cooper, D. C., & Brink, R. A. (1942). The endosperm as a barrier to interspecific hybridization in flowering plants. Science, 95(2455), 75–76.

Cornejo, P., Camadro, E. L., & Masuelli, R. W. (2012). Molecular bases of the postzygotic barriers in interspecific crosses between the wild potato species *Solanum acaule* and *Solanum commersonii*. Genome / National Research Council Canada, 55(8), 605–614.

Darwin, C. (1859). The origin of species, (1872). London: reprinted by D. Appleton and Company.

Darwin, C. (1903). More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters (Bd. 2). D. Appleton.

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouzé, P., Theißen, G., & Van de Peer, Y. (2003). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. Journal of Molecular Evolution, 56(5), 573–586.

de Folter, S., Immink, R. G. H., Kieffer, M., Parenicová, L., Henz, S. R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M. M., Davies, B., & Angenent, G. C. (2005). Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. The Plant Cell, 17(5), 1424–1433.

Del Toro-De León, G., & Köhler, C. (2019). Endosperm-specific transcriptome analysis by applying the INTACT system. Plant Reproduction, 32(1), 55–61.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., & Jacobsen, S. E. (2012). Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. PLoS Genetics, 8(11), e1003062.

Dobzhansky, T. (1937). Genetic nature of species differences. The American Naturalist, 71(735), 404–420.

Dong, X., Luo, H., Bi, W., Chen, H., Yu, S., Zhang, X., Dai, Y., Cheng, X., Xing, Y., Fan, X., Zhu, Y., Guo, Y., & Meng, D. (2023). Transcriptome-wide identification and characterization of genes exhibit allele-specific imprinting in maize embryo and endosperm. BMC Plant Biology, 23(1), 470.

Dresselhaus, T., & Franklin-Tong, N. (2013). Male–female crosstalk during pollen germination, tube growth and guidance, and double fertilization. Molecular Plant, 6(4), 1018–1036.

Dresselhaus, T., Sprunck, S., & Wessel, G. M. (2016). Fertilization mechanisms in flowering plants. Current Biology: CB, 26(3), R125–R139.

Dziasek, K., Simon, L., Lafon-Placette, C., Laenen, B., Wärdig, C., Santos-González, J., Slotte, T., & Köhler, C. (2021). Hybrid seed incompatibility in *Capsella* is connected to chromatin condensation defects in the endosperm. PLoS Genetics, 17(2), e1009370.

Fanata, W. I. D., Lee, S. Y., & Lee, K. O. (2013). The unfolded protein response in plants: A fundamental adaptive cellular response to internal and external stresses. Journal of Proteomics, 93, 356–368.

Finnegan, E. J., & Dennis, E. S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. Nucleic Acids Research, 21(10), 2383–2388.

REFERENCES

Flores-Vergara, M. A., Oneal, E., Costa, M., Villarino, G., Roberts, C., De Luis Balaguer, M. A., Coimbra, S., Willis, J., & Franks, R. G. (2020). Developmental analysis of *Mimulus* seed transcriptomes reveals functional gene expression clusters and four imprinted, endosperm-expressed genes. Frontiers in Plant Science, 11, 132.

Florez-Rueda, A. M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U., & Städler, T. (2016). Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. Molecular Biology and Evolution, 33(11), 2935–2946.

Folsom, J. J., Begcy, K., Hao, X., Wang, D., & Walia, H. (2014). Rice Fertilization-Independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. Plant Physiology, 165(1), 238–248.

Foster, C. S. P., & Ho, S. Y. W. (2017). Strategies for partitioning clock models in phylogenomic dating: Application to the angiosperm evolutionary timescale. Genome Biology and Evolution, 9(10), 2752–2763.

Gehring, M., Bubb, K. L., & Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science, 324(5933), 1447–1451.

Gehring, M., Huh, J. H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell, 124(3), 495–506.

Gehring, M., Missirian, V., & Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PloS One, 6(8), e23687.

German, D. A., Hendriks, K. P., Koch, M. A., Lens, F., Lysak, M. A., Bailey, C. D., Mummenhoff, K., & Al-Shehbaz, I. A. (2023). An updated classification of the Brassicaceae (Cruciferae). PKV Research Journal, 220, 127–144.

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M. A., & Gagliano, W. B. (1998). Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in *Arabidopsis*. Science, 280(5362), 446 LP – 450.

Grundt, H. H., Kjølner, S., Borgen, L., Rieseberg, L. H., & Brochmann, C. (2006). High biological species diversity in the arctic flora. Proceedings of the National Academy of Sciences, 103(4), 972–975.

Grundt, H. H., Popp, M., Brochmann, C., & Oxelman, B. (2004). Polyploid origins in a circumpolar complex in *Draba* (Brassicaceae) inferred from cloned nuclear DNA sequences and fingerprints. Molecular Phylogenetics and Evolution, 32(3), 695–710.

Guitton, A.-E., Page, D. R., Chambrier, P., Lionnet, C., Faure, J.-E., Grossniklaus, U., & Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. Development , 131(12), 2971–2981.

Gustafsson, A. L. S., Skrede, I., Rowe, H. C., Gussarova, G., Borgen, L., Rieseberg, L. H., Brochmann, C., & Parisod, C. (2014). Genetics of cryptic speciation within an arctic mustard, *Draba nivalis*. PloS One, 9(4), e93834.

Haig, D., & Westoby, M. (1991). Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 333(1266), 1–13.

Hatorangan, M. R., Laenen, B., Steige, K. A., Slotte, T., & Köhler, C. (2016). Rapid evolution of genomic imprinting in two species of the Brassicaceae. The Plant Cell, 28(8), 1815–1827.

Hehenberger, E., Kradolfer, D., & Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. Development, 139(11), 2031–2039.

Heilborn, O. (1927). Chromosome numbers in *Draba*. Hereditas, 9(1-3), 59–68.

Henderson, I. R., & Jacobsen, S. E. (2007). Epigenetic inheritance in plants. Nature, 447(7143), 418–424.

Hendriks, K. P., Kiefer, C., Al-Shehbaz, I. A., Bailey, C. D., Hooft van Huysduynen, A., Nikolov, L. A., Nauheimer, L., Zuntini, A. R., German, D. A., Franzke, A., Koch, M. A., Lysak, M. A., Toro-Núñez, Ó., Özüdoğru, B., Invernón, V. R., Walden, N., Maurin, O., Hay, N. M., Shushkov, P., … Lens, F. (2023). Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. Current Biology: CB, 33(19), 4052-4068.e6.

Hohmann, N., Wolf, E. M., Lysak, M. A., & Koch, M. A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. The Plant Cell, 27(10), 2770–2784.

Hornslien, K. S., Miller, J. R., & Grini, P. E. (2019). Regulation of parent-of-origin allelic expression in the endosperm. Plant Physiology, 180(3), 1498 LP – 1519.

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., Hashimoto, M., Kirkbride, R. C., Harada, J. J., Zilberman, D., & Fischer, R. L. (2011). Regulation of imprinted gene expression in *Arabidopsis* endosperm. Proceedings of the National Academy of Sciences, 108(5), 1755 LP – 1762.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J. D., Ossowski, S., Ottilar, R. P., Salamov, A.

A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., … Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nature Genetics, 43(5), 476–481.

Irish, V. F. (2010). The flowering of *Arabidopsis* flower development. The Plant Journal: for Cell and Molecular Biology, 61(6), 1014–1028.

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., & Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. The Plant Journal: for Cell and Molecular Biology, 65(5), 798–806.

Jakobsson, M., Hagenblad, J., Tavaré, S., Säll, T., Halldén, C., Lind-Halldén, C., & Nordborg, M. (2006). A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. Molecular Biology and Evolution, 23(6), 1217–1231.

Jiao, W.-B., & Schneeberger, K. (2020). Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. Nature Communications, 11(1), 989.

Johnston, S. A., den Nijs, T. P. M., Peloquin, S. J., & Hanneman, R. E. (1980). The significance of genic balance to endosperm development in interspecific crosses. Theoretical and Applied Genetics, 57(1), 5–9.

Johnston, S. A., & Hanneman, R. E. (1982). Manipulations of endosperm balance number overcome crossing barriers between diploid *Solanum* species. Science, 217(4558), 446–448.

Jordon-Thaden, I. E., Al-Shehbaz, I. A., & Koch, M. A. (2013). Species richness of the globally distributed, arctic–alpine genus *Draba* L. (Brassicaceae). Alpine Botany, 123(2), 97–106.

Jordon-Thaden, I. E., & Koch, M. A. (2008). Species richness and polyploid patterns in the genus *Draba* (Brassicaceae): a first global perspective. Plant Ecology & Diversity, 1, 255–263.

Jorgensen, C. A. (1958). The flowering plants of Greenland. A taxonomical and cytological survey. K. Dan. Vidensk. Selsk. Biol. Skrift., 9, 1–172.

Josefsson, C., Dilkes, B., & Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Current Biology: CB, 16(13), 1322–1328.

Jullien, P. E., Kinoshita, T., Ohad, N., & Berger, F. (2006). Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. The Plant Cell, 18(6), 1360–1372.

Jørgensen, M. H., Ehrich, D., Schmickl, R., Koch, M. A., & Brysting, A. K. (2011). Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). BMC Evolutionary Biology, 11(1), 1–13.

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., Spillane, C., Robinson, S. J., Links, M. G., Clarke, C., Higgins, E. E., Huebert, T., Sharpe, A. G., & Parkin, I. A. P. (2014). The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nature Communications, 5(1), 3706.

Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A., & Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. The Plant Cell, 20(3), 635–647.

Kawashima, T., & Goldberg, R. B. (2010). The suspensor: not just suspending the embryo. Trends in Plant Science, 15(1), 23–30.

Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (1999). Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. The Plant Cell, 11(10), 1945–1952.

Kinoshita, Tetsu, Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L., & Kakutani, T. (2003). One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. Science, 303(5657), 521–523.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W. J. J., Koornneef, M., & Kakutani, T. (2007). Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. The Plant Journal: for Cell and Molecular Biology, 49(1), 38–45.

Kinser, T. J., Smith, R. D., Lawrence, A. H., Cooley, A. M., Vallejo-Marín, M., Conradi Smith, G. D., & Puzey, J. R. (2021). Endosperm-based incompatibilities in hybrid monkeyflowers. The Plant Cell, 33(7), 2235–2257.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (1999). Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in *Arabidopsis*. Proceedings of the National Academy of Sciences, 96(7), 4186 LP – 4191.

Klosinska, M., Picard, C. L., & Gehring, M. (2016). Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. Nature Plants, 2(10), 16145.

Koini, M. A., Alvey, L., Allen, T., Tilley, C. A., Harberd, N. P., Whitelam, G. C., & Franklin, K. A. (2009). High temperature-mediated adaptations in plant architecture require the bHLH transcription factor PIF4. Current Biology: CB, 19(5), 408–413.

REFERENCES

Kolesnikova, U. K., Scott, A. D., Van de Velde, J. D., Burns, R., Tikhomirov, N. P., Pfordt, U., Clarke, A. C., Yant, L., Seregin, A. P., Vekemans, X., Laurent, S., & Novikova, P. Y. (2023). Transition to self-compatibility associated with dominant S-allele in a diploid siberian progenitor of allotetraploid *Arabidopsis kamchatica* revealed by *Arabidopsis lyrata* genomes. Molecular Biology and Evolution, 40(7), msad122.

Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R. R., Bollmann-Giolai, A., Flis, P., Salt, D. E., Parisod, C., Yant, L., & Kolář, F. (2021). Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. Nature Communications, 12(1), 4979.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., & Gruissem, W. (2003a). *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. The EMBO Journal, 22(18), 4804–4814.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., & Grossniklaus, U. (2003b). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Genes & Development, 17(12), 1540–1553.

Köhler, C., Page, D. R., Gagliardini, V., & Grossniklaus, U. (2005). The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of *PHERES1* by parental imprinting. Nature Genetics, 37(1), 28–30.

Laenen, B., Tedder, A., Nowak, M. D., Toräng, P., Wunder, J., Wötzel, S., Steige, K. A., Kourmpetis, Y., Odong, T., Drouzas, A. D., Bink, M. C. A. M., Ågren, J., Coupland, G., & Slotte, T. (2018). Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*. Proceedings of the National Academy of Sciences, 115(4), 816–821.

Lafon-Placette, C., Hatorangan, M. R., Steige, K. A., Cornille, A., Lascoux, M., Slotte, T., & Köhler, C. (2018). Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. Nature Plants, 4(6), 352–357.

Lafon-Placette, C., Johannessen, I. M., Hornslien, K. S., Ali, M. F., Bjerkan, K. N., Bramsiepe, J., Glöckle, B. M., Rebernig, C. A., Brysting, A. K., Grini, P. E., & Köhler, C. (2017). Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. Proceedings of the National Academy of Sciences, 114(6), E1027 LP-E1035.

Lafon-Placette, C., & Köhler, C. (2014). Embryo and endosperm, partners in seed development. Current Opinion in Plant Biology, 17, 64–69.

Lafon-Placette, C., & Köhler, C. (2016). Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. Molecular Ecology, 25(11), 2620–2629.

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., & Huala, E. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research, 40(D1), D1202–D1210.

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews. Genetics, 11(3), 204–220.

Liu, J., Li, J., Liu, H.-F., Fan, S.-H., Singh, S., Zhou, X.-R., Hu, Z.-Y., Wang, H.-Z., & Hua, W. (2018). Genome-wide screening and analysis of imprinted genes in rapeseed (*Brassica napus* L.) endosperm. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes, 25(6), 629–640.

Liu, Y., Jing, X., Zhang, H., Xiong, J., & Qiao, Y. (2021). Identification of imprinted genes based on homology: An example of *Fragaria vesca*. Genes, 12(3). https://doi.org/10.3390/genes12030380

Lloyd, A., Morgan, C., H. Franklin, F. C., & Bomblies, K. (2018). Plasticity of meiotic recombination rates in response to temperature in *Arabidopsis*. Genetics, 208(4), 1409–1420.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550.

Lysak, M. A., & Koch, M. A. (2011). Phylogeny, genome, and karyotype evolution of crucifers (Brassicaceae). I R. Schmidt & I. Bancroft (Red.), Genetics and Genomics of the Brassicaceae (s. 1–31). Springer New York.

Lysak, M. A., Koch, M. A., Pecinka, A., & Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. Genome Research, 15(4), 516–525.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L., & Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. The New Phytologist, 207(2), 437–453.

Mallet, J. (2013). Species, Concepts of (s. 679–691).

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? BioEssays: news and reviews in molecular, cellular and developmental biology, 38(2), 140–149.

Masiero, S., Colombo, L., Grini, P. E., Schnittger, A., & Kater, M. M. (2011). The emerging importance of Type I MADS box transcription factors for plant reproduction. The Plant Cell, 23(3), 865–872.

Mayr, E. (1970). Populations, species, and evolution: an abridgment of animal species and evolution (Bd. 19). Harvard University Press.

Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., Samaras, P., Richter, S., Shikata, H., Messerer, M., Lang, D., Altmann, S., Cyprys, P., Zolg, D. P., Mathieson, T., Bantscheff, M., Hazarika, R. R., Schmidt, T., Dawid, C., … Kuster, B. (2020). Mass-spectrometry-based draft of the *Arabidopsis* proteome. Nature, 579(7799), 409–414.

Molnár-Láng, M., & Sutka, J. (1994). The effect of temperature on seed set and embryo development in reciprocal crosses of wheat and barley. Euphytica/ Netherlands Journal of Plant Breeding, 78(1), 53–58.

Morley-Smith, E. R., Pike, M. J., Findlay, K., Köckenberger, W., Hill, L. M., Smith, A. M., & Rawsthorne, S. (2008). The transport of sugars to developing embryos is not via the bulk endosperm in oilseed rape seeds. Plant Physiology, 147(4), 2121–2130.

Mulligan, G. A. (1974). Cytotaxonomic studies of *Draba nivalis* and its close allies in Canada and Alaska. Canadian Journal of Botany, 52(8), 1793–1801.

Mulligan, G. A. (1976). The genus *Draba* in Canada and Alaska: key and summary. Canadian Journal of Botany, 54(12), 1386–1393.

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A. J., Abramson, B. W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., Yelina, N., Hartwick, N., Colt, K., Smith, L. M., Ton, J., Kakutani, T., Martienssen, R. A., Schneeberger, K., Lysak, M. A., … Henderson, I. R. (2021). The genetic and epigenetic landscape of the *Arabidopsis* centromeres. Science, 374(6569), eabi7489.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., & Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. The Plant Journal: For Cell and Molecular Biology, 41(5), 697–709.

Nam, J., Kim, J., Lee, S., An, G., Ma, H., & Nei, M. (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. Proceedings of the National Academy of Sciences, 101(7), 1910–1915.

Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., Guggisberg, A., Paape, T., Schmid, K., & Fedorenko, O. M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nature Genetics, 48(9), 1077–1082.

Nowack, M. K., Grini, P. E., Jakoby, M. J., Lafos, M., Koncz, C., & Schnittger, A. (2006). A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nature Genetics, 38(1), 63–67.

Nowak, M. D., Birkeland, S., Mandáková, T., Roy Choudhury, R., Guo, X., Gustafsson, A. L. S., Gizaw, A., Schrøder-Nielsen, A., Fracassetti, M., Brysting, A. K., Rieseberg, L., Slotte, T., Parisod, C., Lysak, M. A., & Brochmann, C. (2020). The genome of *Draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic. Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13280

Nutt, P., Ziermann, J., Hintz, M., Neuffer, B., & Theißen, G. (2006). *Capsella* as a model system to study the evolutionary relevance of floral homeotic mutants. Plant Systematics and Evolution, 259(2), 217–235.

O'Kane, S. L., Jr, Schaal, B. A., & Al-Shehbaz, I. A. (1996). The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. Systematic Botany, 559–566.

Oneal, E., Willis, J. H., & Franks, R. G. (2016). Disruption of endosperm development is a major cause of hybrid seed inviability between *Mimulus guttatus* and *Mimulus nudatus*. The New Phytologist, 210(3), 1107–1120.

Pařenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., Angenent, G. C., & Colombo, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: New openings to the MADS world. The Plant Cell, 15(7), 1538–1551.

Paul, P., Dhatt, B. K., Sandhu, J., Hussain, W., Irvin, L., Morota, G., Staswick, P., & Walia, H. (2020). Divergent phenotypic response of rice accessions to transient heat stress during early seed development. Plant Direct, 4(1), e00196.

Petrén, H., Thosteman, H., Stift, M., Toräng, P., Ågren, J., & Friberg, M. (2023). Differences in mating system and predicted parental conflict affect post-pollination reproductive isolation in a flowering plant. Evolution; International Journal of Organic Evolution, 77(4), 1019–1030.

Picard, C. L., Povilus, R. A., Williams, B. P., & Gehring, M. (2021). Transcriptional and imprinting complexity in *Arabidopsis* seeds at single-nucleus resolution. Nature Plants, 7(6), 730–738.

Pickering, R. A. (1984). The influence of genotype and environment on chromosome elimination in crosses between *Hordeum vulgare* L. × *Hordeum bulbosum* L. Plant Science Letters, 34(1), 153–164.

Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W., & Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. eLife, 3, e03198.

Provine, W. B. (1977). Role of mathematical population geneticists in the evolutionary synthesis of the 1930's and 40's. I Mathematical Models in Biological Discovery (s. 2–31). Springer Berlin Heidelberg.

Qiu, Y., & Köhler, C. (2022). Endosperm evolution by duplicated and neofunctionalized type I MADS-box transcription factors. Molecular Biology and Evolution, 39(1) https://doi.org/10.1093/molbev/msab355

Quint, M., Delker, C., Franklin, K. A., Wigge, P. A., Halliday, K. J., & van Zanten, M. (2016). Molecular and genetic control of plant thermomorphogenesis. Nature Plants, 2(1), 15190.

Radchuk, V., & Borisjuk, L. (2014). Physical, metabolic and developmental functions of the seed coat. Frontiers in Plant Science, 5. https://www.frontiersin.org/articles/10.3389/fpls.2014.00510

Raissig, M. T., Baroux, C., & Grossniklaus, U. (2011). Regulation and flexibility of genomic imprinting during seed development. The Plant Cell, 23(1), 16–26.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The arctic has warmed nearly four times faster than the globe since 1979. Communications Earth & Environment, 3(1), 168.

Raunsgard, A., Opedal, Ø. H., Ekrem, R. K., Wright, J., Bolstad, G. H., Armbruster, W. S., & Pélabon, C. (2018). Intersexual conflict over seed size is stronger in more outcrossed populations of a mixed-mating plant. Proceedings of the National Academy of Sciences of the United States of America, 115(45), 11561–11566.

Raza, A., Hafeez, M. B., Zahra, N., Shaukat, K., Umbreen, S., Tabassum, J., Charagh, S., Khan, R. S. A., & Hasanuzzaman, M. (2020). The plant family Brassicaceae: Introduction, biology, and importance. I M. Hasanuzzaman (Red.), The plant Family brassicaceae: Biology and physiological responses to environmental stresses (s. 1–43). Springer Singapore.

Rebernig, C. A., Lafon-Placette, C., Hatorangan, M. R., Slotte, T., & Köhler, C. (2015). Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. PLoS Genetics, 11(6), e1005295.

Rodrigues, J. A., & Zilberman, D. (2015). Evolution and function of genomic imprinting in plants. Genes & Development, 29(24), 2517–2531.

Rong, H., Yang, W., Zhu, H., Jiang, B., Jiang, J., & Wang, Y. (2021). Genomic imprinted genes in reciprocal hybrid endosperm of *Brassica napus*. BMC Plant Biology, 21(1), 140.

Roth, M., Florez-Rueda, A. M., & Städler, T. (2019). Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. Genetics, 212(1), 141–152.

Rudall, P. J. (2021). Evolution and patterning of the ovule in seed plants. Biological Reviews of the Cambridge Philosophical Society, 96(3), 943–960.

Schmickl, R., & Koch, M. A. (2011). *Arabidopsis* hybrid speciation processes. Proceedings of the National Academy of Sciences, 108(34), 14192–14197.

Scott, R. J., Spielman, M., Bailey, J., & Dickinson, H. G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development, 125(17), 3329 LP – 3341.

Sekine, D., Ohnishi, T., Furuumi, H., Ono, A., Yamada, T., Kurata, N., & Kinoshita, T. (2013). Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. The Plant Journal: for Cell and Molecular Biology, 76(5), 792–799.

Shimizu-Inatsugi, R., Lihová, J., Iwanaga, H., Kudoh, H., Marhold, K., Savolainen, O., Watanabe, K., Yakubov, V. V., & Shimizu, K. K. (2009). The allopolyploid *Arabidopsis kamchatica* originated from multiple individuals of *Arabidopsis lyrata* and *Arabidopsis halleri*. Molecular Ecology, 18(19), 4024–4048.

Shirzadi, R., Andersen, E. D., Bjerkan, K. N., Gloeckle, B. M., Heese, M., Ungru, A., Winge, P., Koncz, C., Aalen, R. B., Schnittger, A., & Grini, P. E. (2011). Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin–dependent regulation of *AGAMOUS-LIKE36*. PLoS Genetics, 7(2), e1001303.

Sitch, L. A., & Snape, J. W. (1987). Factors affecting haploid production in wheat using the *Hordeum bulbosum* system. 1. Genotypic and environmental effects on pollen grain germination, pollen tube growth and the frequency of fertilization. Euphytica/ Netherlands Journal of Plant Breeding, 36(2), 483–496.

Skrede, I., Borgen, L., & Brochmann, C. (2009). Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic–alpine distributions. Heredity, 102(3), 293–302.

Skrede, Inger, Brochmann, C., Borgen, L., & Rieseberg, L. H. (2008). Genetics of intrinsic postzygotic isolation in a circumpolar plant species, *Draba nivalis* (Brassicaceae). Evolution; International Journal of Organic Evolution, 62(8), 1840–1851.

Song, J., Xie, X., Chen, C., Shu, J., Thapa, R. K., Nguyen, V., Bian, S., Kohalmi, S. E., Marsolais, F., Zou, J., & Cui, Y. (2021). *LEAFY COTYLEDON1* expression in the endosperm enables embryo maturation in *Arabidopsis*. Nature Communications, 12(1), 3963.

Stebbins, G. L. (1959). The role of hybridization in evolution. Proceedings of the American Philosophical Society, 103(2), 231–251.

Strother, P. K., & Foster, C. (2021). A fossil record of land plant origins from charophyte algae. Science, 373(6556), 792–796.

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). Molecular Ecology, 25(11), 2427–2442.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408(6814), 796–815.

The Arabidopsis Information Resource (TAIR). (2023). The Arabidopsis Information Resource (TAIR). www.arabidopsis.org. Retrieved 23. November 2023, from: https://www.arabidopsis.org/servlets/Search?action=new_search&type=ecotype

Thörn, E. C. (1992). The influence of genotype and environment on seed and embryo development in barley (*Hordeum vulgare* L.) after crossing with *Hordeum bulbosum* L. Euphytica/ Netherlands Journal of Plant Breeding, 59(2), 109–118.

Tiara, T., & Larter, E. N. (1978). Factors influencing development of wheat-rye hybrid embryos in vitro. Crop Science, 18(2), cropsci1978.0011183X001800020042x.

Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., Spielman, M., Guzmán, P., Oakey, R. J., Kinoshita, T., & Scott, R. J. (2008). MATERNALLY EXPRESSED *PAB C-TERMINAL*, a novel imprinted gene in *Arabidopsis*, encodes the conserved C-terminal domain of polyadenylate binding proteins. The Plant Cell, 20(9), 2387–2398.

Tonosaki, K., Sekine, D., Ohnishi, T., Ono, A., Furuumi, H., Kurata, N., & Kinoshita, T. (2018). Overcoming the species hybridization barrier by ploidy manipulation in the genus *Oryza*. The Plant Journal: for Cell and Molecular Biology, 93(3), 534–544.

Trigg, S. A., Garza, R. M., MacWilliams, A., Nery, J. R., Bartlett, A., Castanon, R., Goubil, A., Feeney, J., O'Malley, R., Huang, S.-S. C., Zhang, Z. Z., Galli, M., & Ecker, J. R. (2017). CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. Nature Methods, 14(8), 819–825.

Vallejo-Marín, M., Buggs, R. J. A., Cooley, A. M., & Puzey, J. R. (2015). Speciation by genome duplication: Repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. Evolution; International Journal of Organic Evolution, 69(6), 1487–1500.

van Ekelenburg, Y. S., Hornslien, K. S., Van Hautegem, T., Fendrych, M., Van Isterdael, G., Bjerkan, K. N., Miller, J. R., Nowack, M. K., & Grini, P. E. (2023). Spatial and temporal regulation of parent-of-origin allelic expression in the endosperm. Plant Physiology, 191(2), 986–1001.

Veselá, A., Hadincová, V., Vandvik, V., & Münzbergová, Z. (2021). Maternal effects strengthen interactions of temperature and precipitation, determining seed germination of dominant alpine grass species. American Journal of Botany, 108(5), 798–810.

Vielle-Calzada, J. P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M. A., & Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotic DDM1 activity. Genes & Development, 13(22), 2971–2982.

Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J., & Comai, L. (2009). Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. Current Biology: CB, 19(13), 1128–1132.

Wang, B., Yang, X., Jia, Y., Xu, Y., Jia, P., Dang, N., Wang, S., Xu, T., Zhao, X., Gao, S., Dong, Q., & Ye, K. (2022). High-quality *Arabidopsis thaliana* genome assembly with Nanopore and HiFi Long Reads. Genomics, Proteomics & Bioinformatics, 20(1), 4–13.

Wang, L., Yuan, J., Ma, Y., Jiao, W., Ye, W., Yang, D.-L., Yi, C., & Chen, Z. J. (2018). Rice interploidy crosses disrupt epigenetic regulation, gene expression, and seed development. Molecular Plant, 11(2), 300–314.

Warwick, S. I., Francis, A., & Al-Shehbaz, I. A. (2006). Brassicaceae: Species checklist and database on CD-Rom. Plant Systematics and Evolution, 259(2), 249–258.

Warwick, Suzanne I. (2011). Brassicaceae in agriculture. I R. Schmidt & I. Bancroft (Red.), Genetics and Genomics of the Brassicaceae (s. 33–65). Springer New York.

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., & Köhler, C. (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genetics, 6(10), e1001152.

Wieczorek, K., Golecki, B., Gerdes, L., Heinen, P., Szakasits, D., Durachko, D. M., Cosgrove, D. J., Kreil, D. P., Puzio, P. S., Bohlmann, H., & Grundler, F. M. W. (2006). Expansins are involved in the formation of nematode-induced syncytia in roots of *Arabidopsis thaliana*. The Plant Journal: for Cell and Molecular Biology, 48(1), 98–112.

Wolf, J. B., & Hager, R. (2006). A maternal–offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biology, 4(12), e380.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T. A., Spillane, C., Nordborg, M., Rehmsmeier, M., & Köhler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. PLoS Genetics, 7(6), e1002126.

Woodward, A. W., & Bartel, B. (2018). Biology in bloom: A primer on the *Arabidopsis thaliana* model system. Genetics, 208(4), 1337–1349.

Wullschleger, S. D., Breen, A. L., Iversen, C. M., Olson, M. S., Näsholm, T., Ganeteg, U., Wallenstein, M. D., & Weston, D. J. (2015). Genomics in a changing arctic: critical questions await the molecular ecologist. Molecular Ecology, 24(10), 2301–2309.

Xiong, H., Wang, W., & Sun, M.-X. (2021). Endosperm development is an autonomously programmed process independent of embryogenesis. The Plant Cell, 33(4), 1151–1160.

Xu, W., Dai, M., Li, F., & Liu, A. (2014). Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. Nucleic Acids Research, 42(11), 6987–6998.

Xu, W., Sato, H., Bente, H., Santos-González, J., & Köhler, C. (2023). Endosperm cellularization failure induces a dehydration-stress response leading to embryo arrest. The Plant Cell, 35(2), 874–888.

Yadegari, R., & Drews, G. N. (2004). Female gametophyte development. The Plant Cell, 16, S133–S141.

Yakimowski, S. B., & Rieseberg, L. H. (2014). The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. American Journal of Botany, 101(8), 1247–1258.

Yant, L., & Bomblies, K. (2017). Genomic studies of adaptive evolution in outcrossing *Arabidopsis* species. Current Opinion in Plant Biology, 36, 9–14.

Yoshida, T., & Kawabe, A. (2013). Importance of gene duplication in the evolution of genomic imprinting revealed by molecular evolutionary analysis of the type I MADS-box gene family in *Arabidopsis* species. PloS One, 8(9), e73588.

Yoshida, T., Kawanabe, T., Bo, Y., Fujimoto, R., & Kawabe, A. (2018). Genome-wide analysis of parent-of-origin allelic expression in endosperms of Brassicaceae species, *Brassica rapa*. Plant & Cell Physiology, 59(12), 2590–2601.

Zhang, H.-Y., Luo, M., Johnson, S. D., Zhu, X.-W., Liu, L., Huang, F., Liu, Y.-T., Xu, P.-Z., & Wu, X.-J. (2016). Parental genome imbalance causes post-zygotic seed lethality and deregulates imprinting in rice. Rice , 9(1), 1–12.

Zhang, Mei, Xie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., Li, H., Yang, W., Zhao, H., Wang, G., Chen, Z., Sun, S., Hauck, A., Jin, W., & Lai, J. (2014). Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Research, 24(1), 167–176.

Zhang, Mei, Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Zhao, H., Guan, H., Hu, X., Jiao, Y., Song, W., & Lai, J. (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proceedings of the National Academy of Sciences of the United States of America, 108(50), 20042–20047.

Zhang, Meishan, Li, N., He, W., Zhang, H., Yang, W., & Liu, B. (2016). Genome-wide screen of genes imprinted in *Sorghum* endosperm, and the roles of allelic differential cytosine methylation. The Plant Journal: For Cell and Molecular Biology, 85(3), 424–436.

Zhang, S., Wang, D., Zhang, H., Skaggs, M. I., Lloyd, A., Ran, D., An, L., Schumaker, K. S., Drews, G. N., & Yadegari, R. (2018). FERTILIZATION-INDEPENDENT SEED-Polycomb Repressive Complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. Plant Physiology, 177(1), 285–299.

Papers I - III

I

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) **116,** 942–961 doi: 10.1111/tpj.16401

RESOURCE

Structural evidence for MADS-box type I family expansion seen in new assemblies of *Arabidopsis arenosa* and *A. lyrata*

Jonathan Bramsiepe[1,2,†], Anders K. Krabberød[1,†], Katrine N. Bjerkan[1,2], Renate M. Alling[1,2], Ida M. Johannessen[1], Karina S. Hornslien[1], Jason R. Miller[3], Anne K. Brysting[1,2] and Paul E. Grini[1,*]

[1]*Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, 0316 Oslo, Norway,*
[2]*CEES, Department of Biosciences, University of Oslo, 0316 Oslo, Norway, and*
[3]*College of STEM, Shepherd University, Shepherdstown, West Virginia 25443-5000, USA*

Received 22 January 2023; revised 24 May 2023; accepted 13 July 2023; published online 30 July 2023.
*For correspondence (e-mail paul.grini@ibv.uio.no).
†Joint 1st authors.

SUMMARY

***Arabidopsis thaliana* diverged from *A. arenosa* and *A. lyrata* at least 6 million years ago. The three species differ by genome-wide polymorphisms and morphological traits. The species are to a high degree reproductively isolated, but hybridization barriers are incomplete. A special type of hybridization barrier is based on the triploid endosperm of the seed, where embryo lethality is caused by endosperm failure to support the developing embryo. The MADS-box type I family of transcription factors is specifically expressed in the endosperm and has been proposed to play a role in endosperm-based hybridization barriers. The gene family is well known for its high evolutionary duplication rate, as well as being regulated by genomic imprinting. Here we address MADS-box type I gene family evolution and the role of type I genes in the context of hybridization. Using two *de-novo* assembled and annotated chromosome-level genomes of *A. arenosa* and *A. lyrata* ssp. *petraea* we analyzed the MADS-box type I gene family in Arabidopsis to predict orthologs, copy number, and structural genomic variation related to the type I loci. Our findings were compared to gene expression profiles sampled before and after the transition to endosperm cellularization in order to investigate the involvement of MADS-box type I loci in endosperm-based hybridization barriers. We observed substantial differences in type-I expression in the endosperm of *A. arenosa* and *A. lyrata* ssp. *petraea*, suggesting a genetic cause for the endosperm-based hybridization barrier between *A. arenosa* and *A. lyrata* ssp. *petraea*.**

Keywords: MADS-box, Arabidopsis genome assembly, *A. lyrata*, *A. arenosa*, endosperm.

INTRODUCTION

Arabidopsis thaliana diverged from its closest relative, *A. arenosa* and *A. lyrata*, at least 6 million years ago (mya) (Hohmann et al., 2015), corresponding with the basal chromosome number reduction from eight to five in *A. thaliana* (Lysak et al., 2006). The species are identified by genome-wide polymorphisms or morphological traits, but only the monophyly of *A. thaliana* is convincingly described as being supported 100% at the level of individual gene trees (Novikova et al., 2016). The species are to a high degree reproductively isolated, but natural field studies (Marburger et al., 2019; Schmickl & Koch, 2011; Schmickl & Yant, 2021) and interspecific hybridization in controlled conditions (Bjerkan et al., 2020; Burkart-Waco et al., 2012; Burkart-Waco et al., 2013; Burkart-Waco et al., 2015; Chen et al., 1998; Comai et al., 2000; Josefsson et al., 2006; Nasrallah et al., 2000; Walia et al., 2009) show that hybridization barriers are incomplete.

A special type of hybridization barrier is based on the triploid endosperm of the seed, where a syncytial growth phase accumulating storage components followed by cellularization switches the endosperm from the nutrition sink to the primary nutrition source of the embryo (Hehenberger et al., 2012). Embryo lethality is frequently observed in

hybrid seeds and contrasted by embryo survival when cultivated *in vitro* after microdissection, suggesting that the endosperm fails to support and transfer nutrition to the embryo (Florez-Rueda et al., 2016; Lafon-Placette et al., 2017; Rebernig et al., 2015; Tonosaki et al., 2017).

The MADS-box type I family of transcription factors is specifically expressed in the endosperm at the time of transition to cellularization (Bemer et al., 2010; Bjerkan et al., 2020; Masiero et al., 2011; Shirzadi et al., 2011; Zhang et al., 2018), and several lines of evidence have suggested a role of type I genes in endosperm based hybridization barriers (Bjerkan et al., 2020; Burkart-Waco et al., 2015; Josefsson et al., 2006; Walia et al., 2009). The gene family is well known for its high duplication rate, and is contrasted by its sister lineage, the MADS-box type II family, which consists of highly conserved single-copy genes (Gramzow & Theissen, 2010; Parenicová et al., 2003; Qiu & Köhler, 2022). Members of the MADS-box type I family are also well known for being regulated by genomic imprinting, parent of origin-dependent allelic expression in the endosperm (Bjerkan et al., 2020; Köhler et al., 2003; Masiero et al., 2011; Shirzadi et al., 2011; Zhang et al., 2018).

Whether gene duplication in the MADS-box type I family is favored or sensed by genomic imprinting is disputed (Erilova et al., 2009; Nam et al., 2004; Yoshida & Kawabe, 2013). One of the first discovered imprinted genes, *MEDEA* (*MEA*), is crucial for endosperm cellularization (Grossniklaus et al., 1998; Kinoshita et al., 1999), and a primary target of MEA is *PHERES1* (*PHE1*), a paternally expressed MADS-box type I transcription factor. A reduction of *PHE1* expression rescues the *mea* mutant phenotype in seeds, and in hybrid crosses, parental imprinting of *PHE1* is altered to the opposite parent (Josefsson et al., 2006; Köhler et al., 2003; Walia et al., 2009). Recently, a propagation of PHERES1 target sites by transposon events was described and indicated a powerful and recent evolutionary effect (Batista et al., 2019; Qiu & Köhler, 2020). Addressing the involvement of MADS-box type I genes and genomic imprinting in the context of hybridization, and especially in crosses between outcrossing species, is important since the occurrence of the phenomenon in selfing species may be considered an evolutionary remnant from outcrossing species.

Such analyses require well-assembled genomes to allow for correct ortholog prediction and identification of paralogs, and the ability to address structural variation in the DNA sequences. The *A. thaliana* genome caused an explosion in plant genome research, supporting questions in many fields (The Arabidopsis Genome Initiative, 2000). The small genome size and the strong homozygosity of *A. thaliana* made this effort successful but is not archetypical for the majority of plants, not even the genus Arabidopsis. The complexity of whole-genome duplications, ancient and recent polyploidization events, and also high transposon activity creating abundant pseudogenes, make the assembly of plant genomes especially challenging (Kress et al., 2022). Thus, for many sequenced plant genomes, the assembly of repeats such as centromeres, telomeres, and ribosomal DNA has become feasible only after considerable methodological improvements implemented by long-read sequencing technology and Hi-C scaffolding (Kovaka et al., 2023; Naish et al., 2021).

Using assembled Arabidopsis genomes and new assemblies of the genomes of the European *A. arenosa* and *A. lyrata* ssp. *petraea*, we have analyzed the MADS-box type I gene family in Arabidopsis to predict orthologs, copy number, and structural genomic variation related to the type I loci. The analysis is compared to gene expression profiles sampled before and after the transition to endosperm cellularization in order to investigate the involvement of MADS-box type I loci in endosperm-based hybridization barriers.

RESULTS AND DISCUSSION

Assembly and annotation of *A. arenosa* and *A. lyrata* ssp. *petraea* genomes

To identify MADS-box type I family orthologs between species, allowing interspecies gene expression studies of orthologs between *Arabidopsis* species, and comparative analysis of orthologous and paralogous genes including their genomic neighborhood, we sequenced and assembled the genomes of one individual of *A. arenosa* from Pusté Pole in Slovakia and one individual of *A. lyrata* ssp. *petraea* from Pernitz in Austria (see Figure S1 and Data S1 for a summary of assembly methods and statistics). The genome sizes were estimated as 201 Mbp for *A. lyrata* ssp. *petraea* and 179 Mbp for *A. arenosa* Pusté Pole (see K-mer analysis in Figure S2), which were confirmed by flow-cytometric data (Data S2) and previous reports (Dart et al., 2004; Johnston et al., 2005; Lysak et al., 2009), and by the assembled sequence lengths of 188 Mbp and 153 Mbp. Whole-genome alignments between scaffolded assemblies were analyzed for breaks and inconsistencies (Figures S3–S6) (Dudchenko et al., 2018; Durand et al., 2016; Robinson et al., 2011), revealing chromosome-length scaffolds comparable in quality to other published genomes of Arabidopsis (Data S1).

Localization of rDNA, telomeric and centromeric repeats indicates complete assemblies and the remaining gaps clustered close to the centromeres (Figure 1a,b). Repeats and transposable elements (TE) were classified and compared to their occurrence in *A. lyrata* ssp. *lyrata* and *A. thaliana* genomes, revealing a higher percentage of repeat sequences in *A. arenosa* and *A. lyrata* compared to *A. thaliana* (Figure 1c). Comparison of the predicted proteomes of *A. arenosa* Pusté Pole, *A. lyrata* ssp. *petraea* and *A. thaliana* detected more than 21000 shared orthogroups and a few (<2%) species-specific orthogroups (Figure 1d). Orthogroups shared by *A. arenosa* and *A. lyrata* ssp.

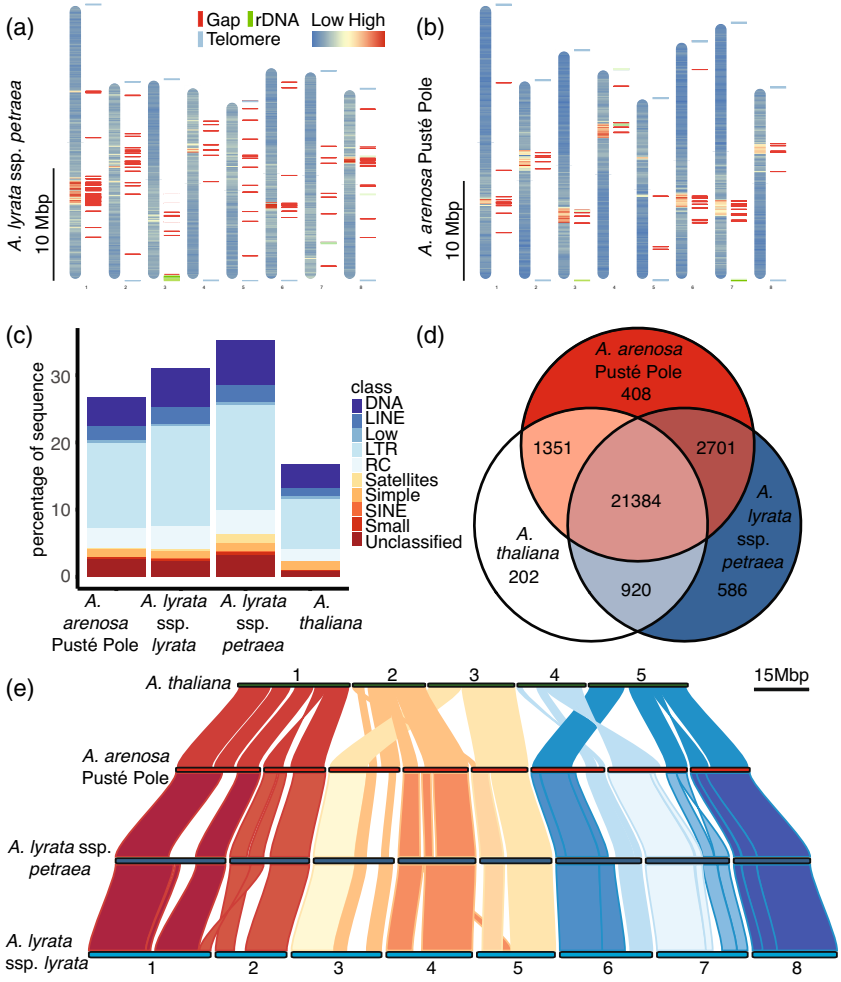


Figure 1. Comparison of Arabidopsis genome assemblies. Repeats and genes were predicted using RepeatModeler and Braker.
(a,b) Idiograms of *A. lyrata* ssp. *petraea* (a) and *A. arenosa* Pusté Pole (b). The repeat density is shown as a heat map. Gaps (red), rDNA (green), and telomeric sequences (blue) are indicated as bars next to the chromosomes. Hi-C scaffolding gives a continuous genome assembly, and the few remaining gaps cluster in centromeric regions.
(c) Abundance of repeat classes in *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole compared to *A. thaliana* and *A. lyrata* ssp. *lyrata*. The *A. thaliana* genome has the lowest number of repeats relative to genome size. The increased number of repetitive elements in *A. lyrata* ssp. *petraea* compared to *A. lyrata* ssp. *lyrata* is probably due to the technical advantage of using long-read sequencing technology (Pucker et al., 2022).
(d) Venn diagram showing overlap of predicted orthologs between *A. lyrata* ssp. *petraea, A. arenosa* Pusté Pole, and *A. thaliana*. Orthogroups were predicted using OrthoFinder. A majority of orthologs are common for all three species, while few species-specific genes were identified.
(e) Synteny of *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole assemblies to *A. lyrata* ssp. *lyrata* and *A. thaliana* references. Known rearrangements can be found in the comparison to the n = 5-based karyotype in *A. thaliana*. The karyotypes between *A. arenosa* and *A. lyrata* overlap, with one large inversion on the end of scaffold 7. The *A. lyrata* ssp. *petraea* assembly was corrected for a previously reported misassembly on scaffold 1 (Slotte et al., 2013).

petraea only were two to three times more frequent, reflecting the recent split of the two species. Analysis of local structural variants of *A. arenosa* Pusté Pole and *A. lyrata* ssp. *petraea* to *A. thaliana* and *A. lyrata* ssp. *lyrata* revealed a high syntenic relationship and limited chromosomal rearrangement (Figure 1e, Figure S7).

MADS-box gene family characterization identifies diagnostic motifs

To identify MADS-box type I family orthologs between species, gene sequences from *A. thaliana* were used to identify the MADS-box genes in four currently available Arabidopsis genomes of sufficient quality (*A. thaliana, A. lyrata* ssp. *lyrata*, *A. halleri*, and *A. arenosa* Strecno) in addition to the two newly assembled genomes (see Data S3.1 for all sequences). In a phylogenetic tree using two *Capsella* species as outgroups (*C. rubella* and *C. grandiflora*), the MADS-box genes were categorized into types and subgroups according to phylogenetic placement (Figure 2, Figures S8 and S9). As expected, the resulting tree separates type I from type II, as well as the three subgroups of type I (Mα, Mβ, and Mγ) and the two subgroups

of type II (MIKC and MIKC* [also referred to as Mδ]) (Arora et al., 2007; Gramzow & Theißen, 2013; Henschel et al., 2002; Parenicová et al., 2003; Qiu & Köhler, 2022; Thangavel & Nayar, 2018). Monophyletic groups of Arabidopsis MADS-box genes, sharing a common ancestor with genes from the outgroup *Capsella*, are numbered starting from the top of the tree. Additional gene duplications that occurred in the common ancestor of Arabidopsis, but after Arabidopsis and *Capsella* separated, are marked with an additional letter. For instance, "Mγ-1a" and "Mγ-1b" indicate that these two clades share a last common ancestor with *C. rubella* (i.e. "Mγ-1"), but have a gene duplication specific for Arabidopsis that occurred after the separation of Arabidopsis from the rest of the Brassicales (Figure 2; for a fully expanded tree with names of all groups and sequences, see Figure S8, Data S3).

A principal component analysis of MEME motifs showed that the distribution of sequence motifs follows the same patterns as the phylogenetic tree (Figure 2, Figure S10). This allowed the identification of motifs that are diagnostic for clades or even larger groups, as exemplified by Mγ motif 3 or Mα motifs 7 and 8 (Figure 2). The protein motifs in MADS-box genes were generally found to be more similar inside clades than between clades. Some motifs were distributed in all clades except one, as exemplified by motif 46 (not in MIKC) and motif 17 (not in Mα). Motifs that classified MADS-box type I gene sequences into one of the Mα, Mβ, and Mγ clades are shown in Figure 2, Figures S8 and S9.

Variance in MADS-box type I gene copy number

The gene copy number was determined for each *Arabidopsis* species in each clade of the phylogenetic tree (Figure 2). The gene copy number varies for the MADS-box type I gene clades. The number of copies is low for all clades in type II MIKC (Figure S9) except the type II clade *FLOWERING LOCUS C* (*FLC*), which spans MIKC28 to MIKC32, including *MAF1-MAF5* and *FLC* (Figure S9), and which is arranged as a tandem array on chromosome 5 in *A. thaliana* (Ratcliffe et al., 2003). The *FLC* duplication in *A. lyrata* ssp. *petraea* and its effect on flowering were previously studied (Kemi et al., 2013).

Members of the Mβ group are highly duplicated. The Mβ-1, Mβ-3, and Mβ-5 clades have three to four copies in *A. arenosa* and *A. lyrata*. *A. thaliana* has seven copies in the Mβ-1 and Mβ-3 clades and two copies in the Mβ-5 clade (Figure 2). The Mα group is separated into two monophyletic groups (Parenicová et al., 2003). In order to investigate the biological roles of the two groups, we compared the newly assembled versions of those genes to previously published interaction maps between MADS-box proteins (de Folter et al., 2005; Qiu & Köhler, 2022). The first group (Mα-1 to Mα-11) contains clades with proteins for which only interaction with Mγ proteins has been found. Mα-5 and Mα-9 are the exceptions for which protein interaction with the Mβ group was also found in a yeast two-hybrid screen (de Folter et al., 2005; Qiu & Köhler, 2022). In strong contrast, for the second group (Mα-12 to Mα-18), all *A. thaliana* proteins could interact with proteins of the Mβ class, while Mα-13, Mα-14, Mα-17, and Mα-19 could also interact with Mγ in the same yeast two-hybrid studies (de Folter et al., 2005; Qiu & Köhler, 2022).

The groups are also distinguished by gene copy number, with the second group having higher numbers comparable to the Mβ class. The Mα-1 (*AGL23/28*) and Mα-4 (*DIANA*) clades, which are involved in female gametophyte development in *A. thaliana* (Bemer et al., 2008; Colombo et al., 2008; Steffen et al., 2008), are each duplicated once in *A. lyrata* and *A. arenosa* Pusté Pole. The Mα-2 (*AGL40*) and the Mα-3 clade containing *AGL62* in *A. thaliana* are single-copy genes in *A. thaliana*, *A. lyrata*, and *A. arenosa*. A possible explanation for the lack of expansion in Mα-2 and Mα-3 is their functional requirement in endosperm development. Mutations in *AGL62* cause early endosperm cellularization in *A. thaliana*, and it is suggested that *AGL40* may have functions in the same pathway since *agl40* mutant plants produce smaller seeds (Kang et al., 2008; Kirkbride et al., 2019; Roszak & Köhler, 2011). Like Mβ-1 and Mβ-3, the Mα-7a clade is notable for its high duplication rate in *A. thaliana*: four duplicate genes (*AGL58*, *AGL59*, *AGL64*, and *AGL85*) were found in the Mα-7a clade, while only single orthologs were detected in *A. lyrata* and *A. arenosa*. The opposite is found for the Mα-9 clade, which contains three loci in *A. lyrata* and *A. arenosa* but only one single gene in *A. thaliana* (*AGL102*). In the second Mβ-binding group, the Mα-12 stands out with high duplication rates in *A. lyrata* and *A. arenosa*, represented by *AGL73*, *AGL83*, and *AGL84* in *A. thaliana*. The other Mβ-binding Mα clades have, except for Mα-15, single-gene duplications or losses depending on the *Arabidopsis* species.

In the Mγ class, the Mγ-1 and Mγ-3 clades have variable MADS-box type I gene copy numbers. Mγ-1a, known as *PHERES1* (*PHE1*) in *A. thaliana*, shows six copies in *A. lyrata* ssp. *petraea* but only three in *A. lyrata* ssp. *lyrata*. The homolog *PHE2* in *A. thaliana* is grouped in the distinct Mγ-1b clade and duplicated in both *A. lyrata* subspecies. The Mγ-3 clade has three copies, *AGL34*, *AGL36*, and *AGL90* in *A. thaliana*, and a triplication can also be found in *A. arenosa* Pusté Pole contrasted by single-copy loci in the other species studied (Figure 2).

Distribution and clustering of MADS-box type I genes differ in Arabidopsis

We compared gene family distributions on chromosomes between *A. thaliana* and other species in the genus. Whereas *A. thaliana*, MADS-box type II genes are highly conserved but distributed across all chromosomes

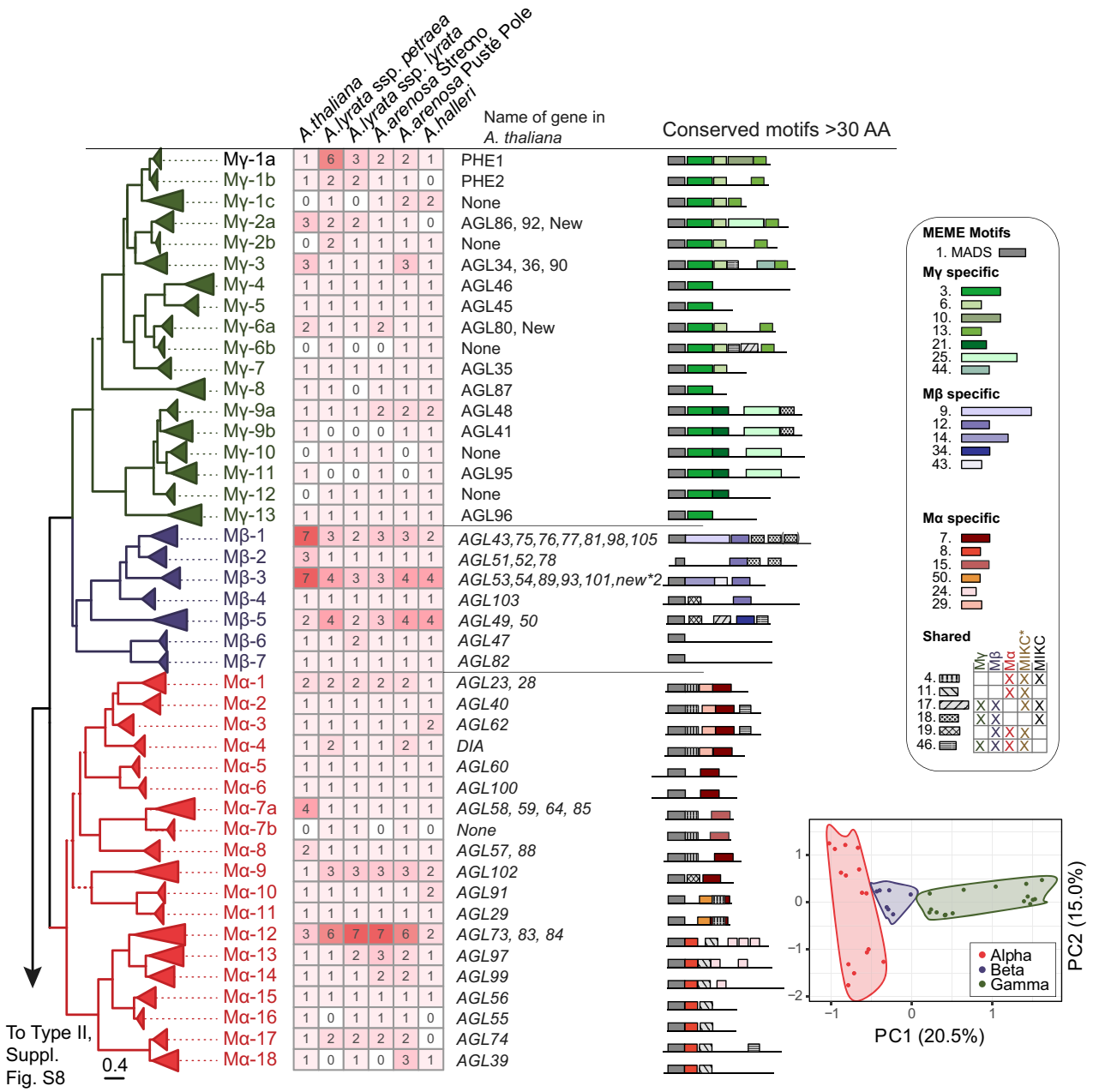


Figure 2. Phylogenetic analysis of MADS-box type I genes in Arabidopsis.
The tree was derived by a maximum likelihood analysis of 362 identified MADS-box type I sequences from *A. thaliana, A. lyrata* ssp. *petraea, A. arenosa* Strecno, *A. arenosa* Pusté Pole, and *A. halleri* with two species of *Capsella* (100 sequences; not shown in the figure) used as outgroup. Solid branches represent bootstrap support >85%, while branches with support values <85% are dashed. The root of the tree is placed between type I and type II genes (the corresponding tree and heatmap for type II genes can be found in Figure S9). Triangles represent clades where branches are collapsed at the most recent gene duplication event in the last common ancestor of the genus Arabidopsis. The length of the triangles corresponds to the overall branch length of the collapsed clade. The groups are colored with Mα in red, Mβ in blue, and Mγ in green (see the main text for naming schemes of clades). The heatmap shows the number of gene copies for each clade in the five genomes. The column next to the heatmap indicates canonical MADS-box names of the genes in *A. thaliana* found in the corresponding clade; "none" means that a gene representing the clade is not found in the *A. thaliana* genome, while "new" indicates that the gene does not have a given MADS-box name in *A. thaliana*. The corresponding *A. thaliana* gene locus code (AGI) and commonly used synonymous names for MADS-box genes can be found in Data S3. The last column shows a simplified representation of the MEME motifs. The insert in the lower right corner shows PCA on the occurrence of MEME motifs (Full PCA in Figure S10). A fully expanded phylogenetic tree with individual tip labels, support values for all branches, and the outgroup *Capsella* can be found in Figure S8. Results from the full MEME analysis can be found in Figure S8.

(Gramzow & Theissen, 2010; Parenicová et al., 2003; Qiu & Köhler, 2022), the type I genes are primarily localized on chromosomes 1 and 5, hypothesized to be a result of recent and local duplications (Parenicová et al., 2003). In *A. lyrata* and *A. arenosa*, however, we find that the MADS-box type I genes are distributed across all eight chromosomes (Figure 3). The highly variable type I clades cluster in the vicinity of centromeric repeats.

Chromosomal localization may explain the high duplication rate of type I clades in *A. lyrata* and *A. arenosa* compared to *A. thaliana*. For example, the variable Mγ-2 group is found close to centromeric repeats in *A. thaliana*, *A. lyrata*, and *A. arenosa*. There are three Mγ-2 in *A. thaliana*, and four in *A. lyrata* ssp. *petraea*, and two in *A. arenosa* Pusté Pole - all on chromosome 1 (Figure 3). On the other hand, the Mγ-1 clades with *PHE1* and *PHE2* are located distantly from the centromere on the lower arm of chromosome 1 in *A. thaliana* and contain only these two genes. In *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole, the local duplications in these Mγ-1 clades are clustered close to the centromere of chromosome 2.

A similar pattern can be seen in regard to *A. thaliana* chromosome 5. The highly variable Mβ-3 cluster is located close to the centromeric repeats on chromosome 5 of *A. thaliana* and on chromosome 6 of *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole. In contrast, the Mα-12 loci are centrally located on the lower arm of chromosome 5 of *A. thaliana*, distant from the centromere, yet have a syntenic relationship to a region close to the centromeric repeats on chromosome 8 of *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole (Figure 3).

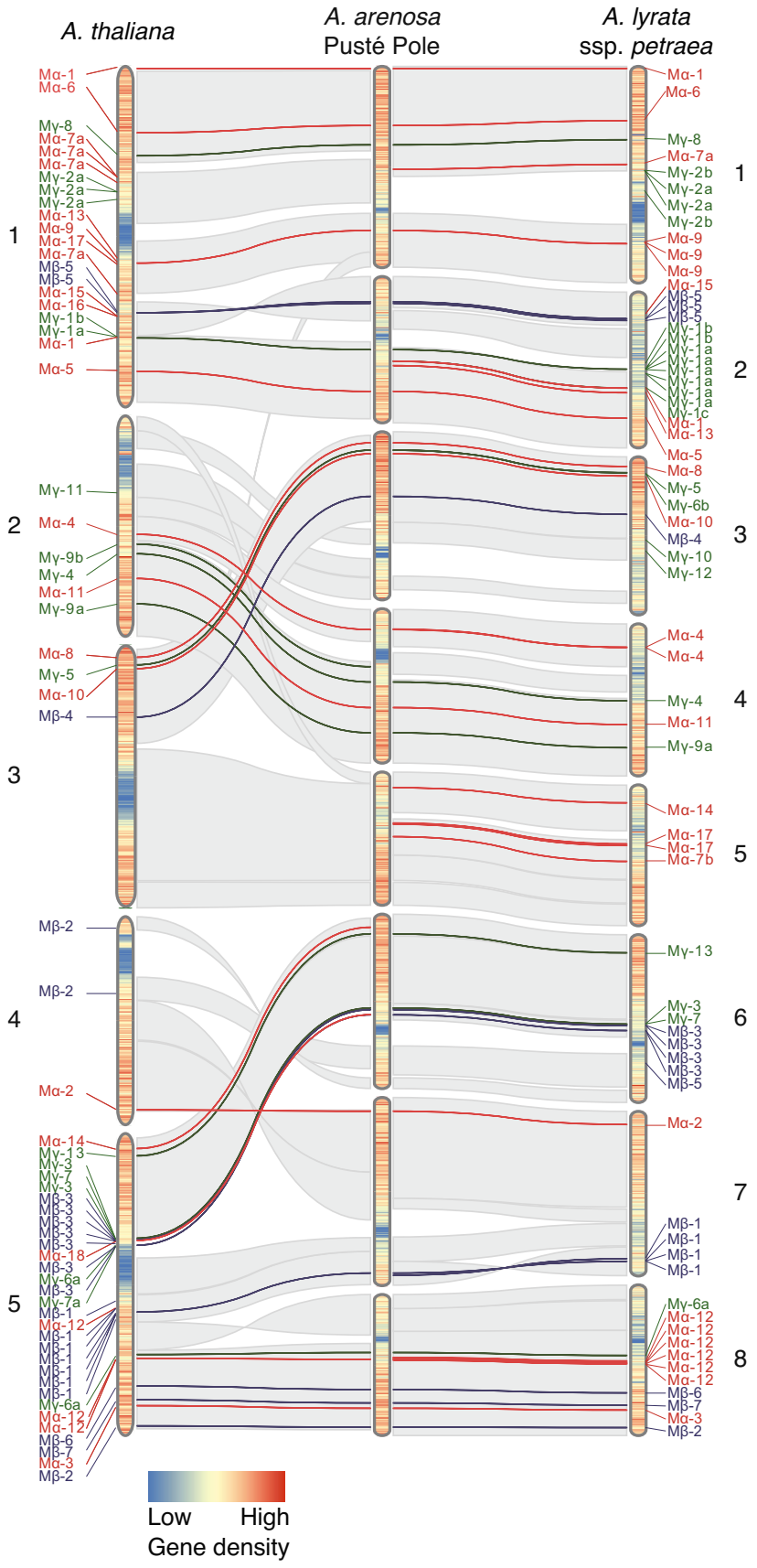
AGL23 and *AGL28* of the Mα-1 clade are located on alternate chromosome arms of chromosome 1 of *A. thaliana*. The distant localization is maintained in syntenic regions in *A. lyrata* and *A. arenosa*, transferred to chromosomes 1 and 2 (Figure 3). The separate localization of these genes is consistent with their distinct expression (Figure 4). Likewise, *DIANA* (*AGL61*/Mα-4) is a single-copy gene on chromosome 2 of *A. thaliana*. However, the Mα-4 clade has a local duplication on chromosome 4 of *A. lyrata* and *A. arenosa* (Figure 3). Also, these genes show distinct expression during seed development (Figure 4).

Differential expression of MADS-box orthologs in Arabidopsis seeds

In order to further investigate the relationship between MADS-box type I expansion, their genomic neighborhood, and gene expression in the seed, we generated seed transcript profiles of *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole. Several MADS-box type I genes regulate early seed development in *A. thaliana* and are expressed in an imprinted or biparental manner at the transition to endosperm cellularization (Bemer et al., 2010; Bjerkan et al., 2020; Masiero et al., 2011; Shirzadi et al., 2011; Zhang et al., 2018). In *A. thaliana*, a steep decline in MADS-box type I gene expression can be seen between 6 and 9 DAP (days after pollination), coinciding with developmental stages before and after cellularization (Bjerkan et al., 2020) (Figure 4). Seed development in *A. lyrata* and *A. arenosa* progresses slower, and the comparable developmental phenotypes can be staged to 9 and 15 DAP (Lafon-Placette et al., 2017). We, therefore, compared MADS-box type I RNA-Seq profiles from 9 and 15 DAP *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole seeds to *A. thaliana* 6 to 9 DAP seed transcriptomes. While individual MADS-box type II genes were either never expressed or constantly expressed at both sampled stages in *A. thaliana* (Figure S11), their expression levels differed in *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole (Figure 4). In all species, high expression of the majority of Mγ genes was observed before cellularization, whereas Mβ expression is low or absent, and undergoes only minor changes during later-stage endosperm development. The Mα group expression pattern reflects its phylogeny and splits into two groups (Figure 2), as only the Mα-1 to Mα-11 clades are significantly expressed during early seed development. The weakly expressed Mα genes and the whole group of Mβ were characterized by high duplication levels. Although not significant when comparing all MADS-box type I genes ($R = -0.15$ $P = 0.077$), we suggest that recent gene duplication in MADS-box type I genes correlates negatively with the observed seed-specific expression of these loci.

The Mα and Mγ expression patterns also reveal differences between *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole. In *A. arenosa*, the expression patterns are similar to *A. thaliana* with an up and down-regulation before and after endosperm cellularization, respectively. Surprisingly this pattern can not be found in *A. lyrata* ssp. *petraea*, where the expression of most MADS-box type I genes continues to increase at the later time point, sampled after endosperm cellularization. Mα-3 (*AGL62*), of which mutation leads to precocious cellularization in *A. thaliana* (Kang et al., 2008), is down-regulated at cellularization in *A. thaliana* and *A. arenosa* but its expression remains unchanged

Figure 3. Genome-wide distribution of MADS-box type I genes in *Arabidopsis arenosa* and *A. lyrata* ssp. *petraea*.
Localization of MADS-box type I Mα (red), Mβ (blue), and Mγ (green) are shown on top of the predicted syntenic sequence context. The naming corresponds to clades in the MADS-box phylogeny in Figure 2. While the MADS-box type I genes are mainly located on chromosomes 1 and 5 in *A. thaliana*, they are broadly distributed across all chromosomes of *A. arenosa* Pusté Pole and *A. lyrata* ssp. *petraea*. Local duplication and clusters of paralogs can be found in all three species. The heatmap covering the idiogram indicates gene density. Scaffolds are not in scale comparing species but show the proportion of the individual genomes (for scale, see Figure 1e and Figure S7).



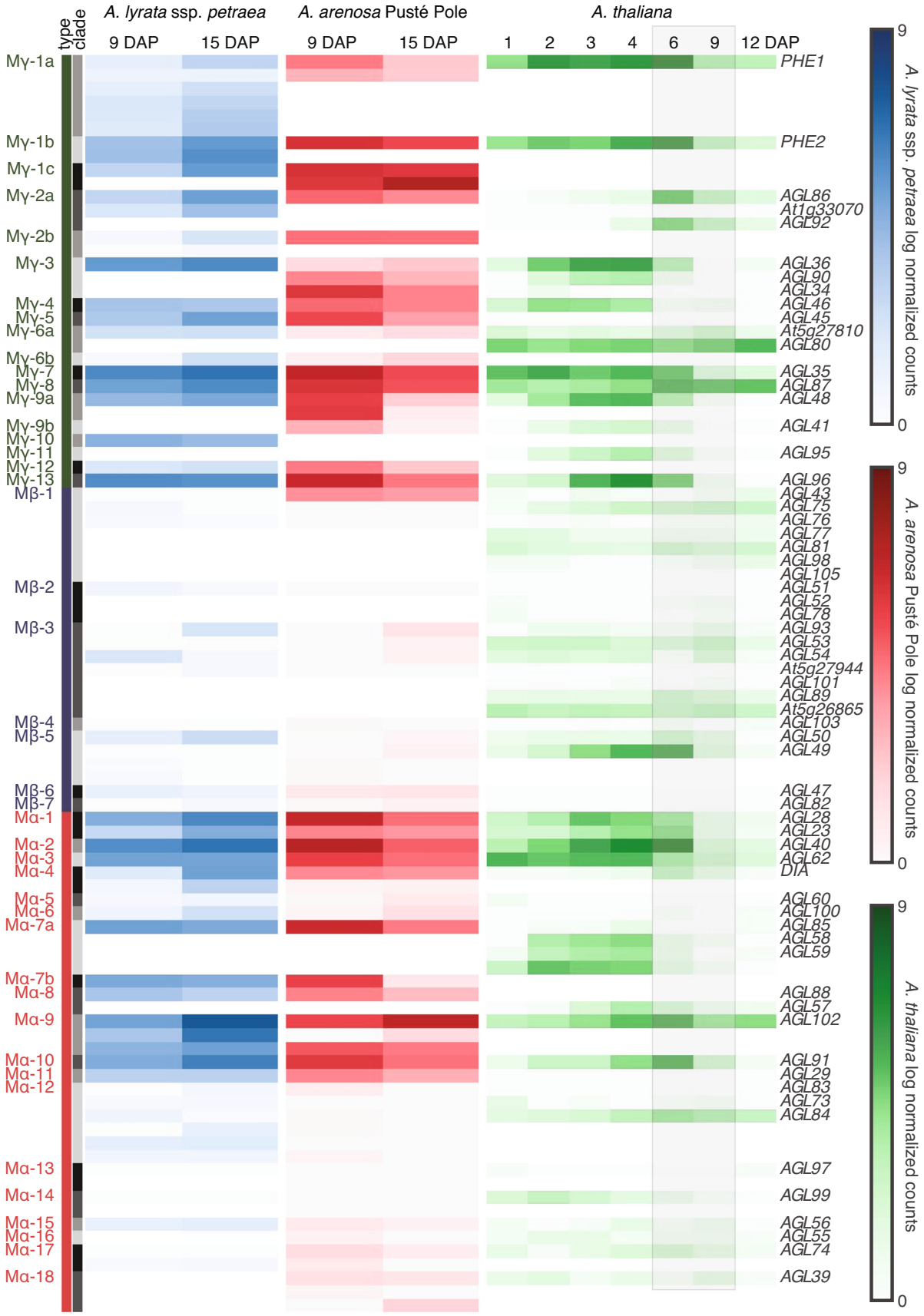


Figure 4. MADS-box type I expression during Arabidopsis seed development.
Gene expression profiles are displayed for all identified MADS-box type I genes and compared between *A. lyrata* ssp. *petraea* (left column), *A. arenosa* Pusté Pole (middle column), and *A. thaliana* (right column). The *A. thaliana* development time series serves as a reference (Bjerkan et al., 2020). The endosperm cellularization in *A. thaliana* occurs between 6 and 9 days after pollination (DAP), indicated by the gray boxed area. To adjust for the relatively slower development in *A. arenosa* and *A. lyrata*, corresponding stages before and after endosperm cellularization were sampled at 9 and 15 DAP, respectively. Orthologous genes are grouped and ordered and follow the MADS-box phylogeny described in Figure 2. Counts normalized for each sample are shown on a base-2 logarithmic scale. Most Mγ and half of the Mα genes show a high expression before cellularization and a substantial decline post-cellularization in *A. thaliana* and *A. arenosa*. In *A. lyrata* ssp. *petraea*, however, this decline of expression cannot be found; instead, most of the expressed Mγ and Mα genes even increase in expression at the later (15 DAP) stage.

in *A. lyrata* ssp. *petraea*. Likewise, *Mγ-1a* (*PHE1*) expression declines in *A. thaliana* with the onset of endosperm cellularization, and *Mγ-1a* is also higher expressed during early seed development in *A. arenosa* and declines at cellularization. In *A. lyrata* ssp. *petraea*, however, the *Mγ-1a* duplications show a higher level of expression at the 15 DAP time point compared to 9 DAP (Figure 4). In summary, the observed down-regulation of type I class Mα and Mγ in *A. thaliana* and *A. arenosa* is absent, weakened, or strongly delayed in *A. lyrata* ssp. *petraea*, or endosperm cellularization occurs independently of MADS-box type I expression.

Recent duplication in Mγ 1 (*PHERES1*)

The complete *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole genomes allow interspecies comparisons of gene family expansion. We investigated the observed lack of *Mγ-1a* down-regulation and extensive expansion of this clade in *A. lyrata* ssp. *petraea*. In *A. thaliana*, *Mγ-1a* (*PHE1*) is paternally expressed while the maternal allele is repressed by the POLYCOMB REPRESSIVE COMPLEX 2 (PRC2). It is proposed that PRC2 repression is circumvented in the paternal allele of *PHE1* by DNA-methylation of a repeat-rich region in the three-prime regulatory region of *PHE1* (Makarevich et al., 2008). We, therefore, used the whole-genome assemblies of *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole to address and compare the Mγ-1 duplications in their genomic landscape (Figure 5).

To exclude artifacts, we compared the *A. lyrata* ssp. *petraea Mγ-1a* genomic loci with *A. lyrata* ssp. *lyrata* and our primary and scaffolded assemblies supported this area well. Both genome assemblies aligned well through this region. We addressed this area's syntenic relation, the

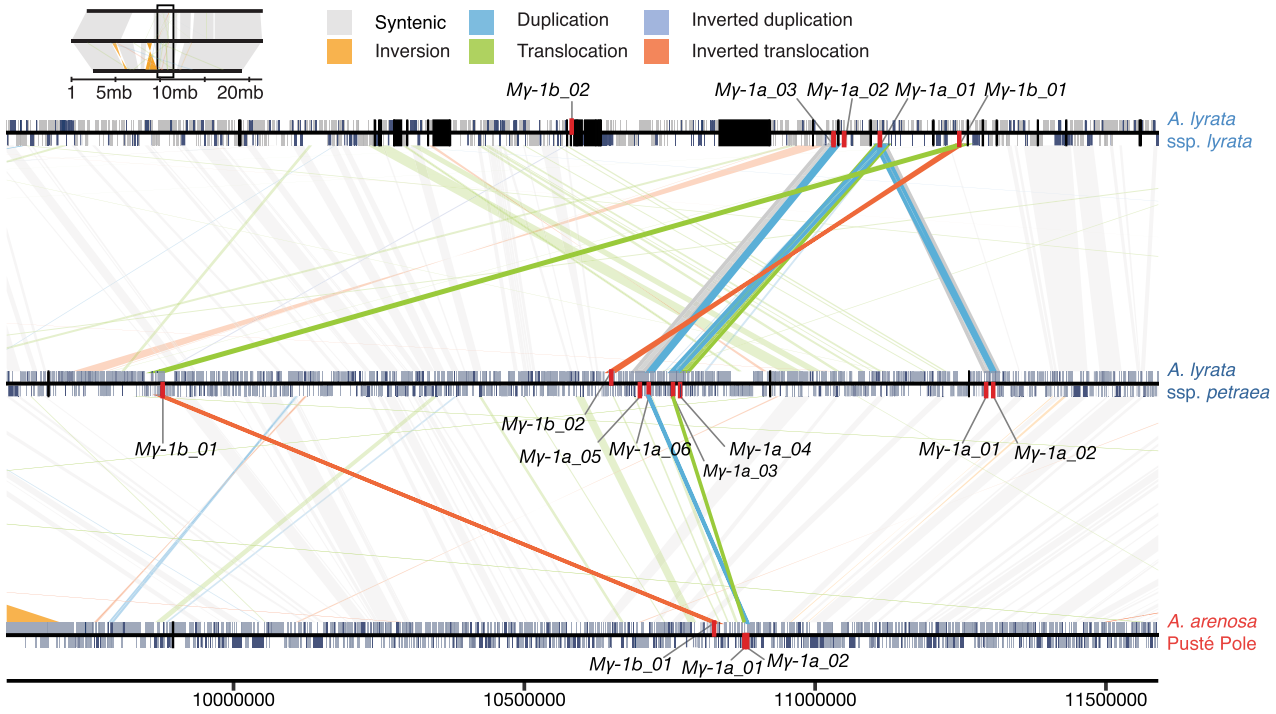


Figure 5. *Mγ-1* duplications in *Arabidopsis lyrata* ssp. *lyrata*, *A. lyrata* ssp. *petraea*, and *A. arenosa*.
Localization and transpositions of *Mγ-1* genes (red) on segments of chromosome 2 in *A. lyrata* ssp. *lyrata*, *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole, with genes marked in blue, repeat in gray and assembly gaps in black. Syntenic sequences between assemblies are indicated by light gray connections. Inversions (orange), duplications (blue), translocations (green), inverted duplications (light purple), and inverted translocations (orange-red) were classified by SyRI. Scale bar and nucleotide positions refer to the *A. lyrata* ssp. *petraea* assembly.

origin, and the underlying nature of these gene duplications. *Mγ-1a* is represented by 3, 6, and 2 orthologs in *A. lyrata* ssp. *lyrata, A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole, respectively (Figure 5).

We used SyRI synteny analysis to investigate the structural variants and the genomic context surrounding these orthologs close to the centromeric region of chromosome 2. The analysis shows that the *A. lyrata* ssp. *lyrata Mγ-1a*_01 is quadrupled in *A. lyrata* ssp. *petraea*, creating *A. lyrata* ssp. *petraea Mγ-1a_01, Mγ-1a_02, Mγ-1a_03* and *Mγ-1a_04*. The *A. lyrata* ssp. *lyrata Mγ-1a*_03 is duplicated generating *Mγ-1a_05* and *Mγ-1a_06* in *A. lyrata* ssp. *petraea*. The *A. lyrata* ssp. *lyrata Mγ-1a_02* is structurally related to its paralog *Mγ-1a_01* in *A. lyrata* ssp. *lyrata*, but exhibits no syntenic relation to *A. lyrata* ssp. *petraea*. The two *Mγ-1a* paralogs in *A. arenosa* Pusté Pole, *Mγ-1a_01* and *Mγ-1a_02* have a syntenic relation to *Mγ-1a_03* and *Mγ-1a_06* of *A. lyrata* ssp. *petraea*, respectively (Figure 5). Alignments of genomic sequences surrounding *Mγ-1* loci were generated using MAFFT, showing that the downstream repeat-rich area responsible for imprinting of *A. thaliana Mγ-1a* (Makarevich et al., 2008) can be found in one *A. arenosa* Pusté Pole (*Mγ-1a_01*) and four *A. lyrata* ssp. *petraea* loci (*Mγ-1a_01, Mγ-1a_02, Mγ-1a_04, and Mγ-1a_06*) (Figure S12). The presence of the repeat-rich area, and thereby duplication of the regulatory region, indicates that these genes may be imprinted in a similar manner as *PHE1* (Makarevich et al., 2008). However, imprinting studies in *A. lyrata* ssp. *petraea* is required to demonstrate this, since a repeat-rich region is also detected in one *A. lyrata* ssp. *lyrata* paralog, but no *Mγ-1a* imprinting was identified in a previous study searching for imprinting in *A. lyrata* ssp. *lyrata* (Klosinska et al., 2016).

To further analyze the functional regulation of the *Mγ-1* group, we used *A. arenosa* and *A. lyrata* ssp. *petraea* genome resources to identify the orthologous genes of previously described *Mγ-1a (PHE1)* transcription factor targets in *A. thaliana* (Batista et al., 2019). These *PHE1* targets were previously clustered into three groups in *A. thaliana* depending on their expression in the seed. Characteristic of cluster 1 is the down-regulation in *A. thaliana* during endosperm cellularization, whereas the two other clusters show only minor differential expression changes (Batista et al., 2019). Our differential expression analysis of *A. arenosa* and *A. lyrata* ssp. *petraea* orthologous *Mγ-1a (PHE1)* targets before and after endosperm cellularization demonstrated a strong down-regulation of cluster 1 *Mγ-1a* targets in *A. arenosa* Pusté Pole seeds (Figure 6a). In contrast, this drop in expression could not be detected in *A. lyrata* ssp. *petraea* cluster 1 *Mγ-1a* targets, where the expression level was relatively constant (Figure 6a). In summary, cluster 1

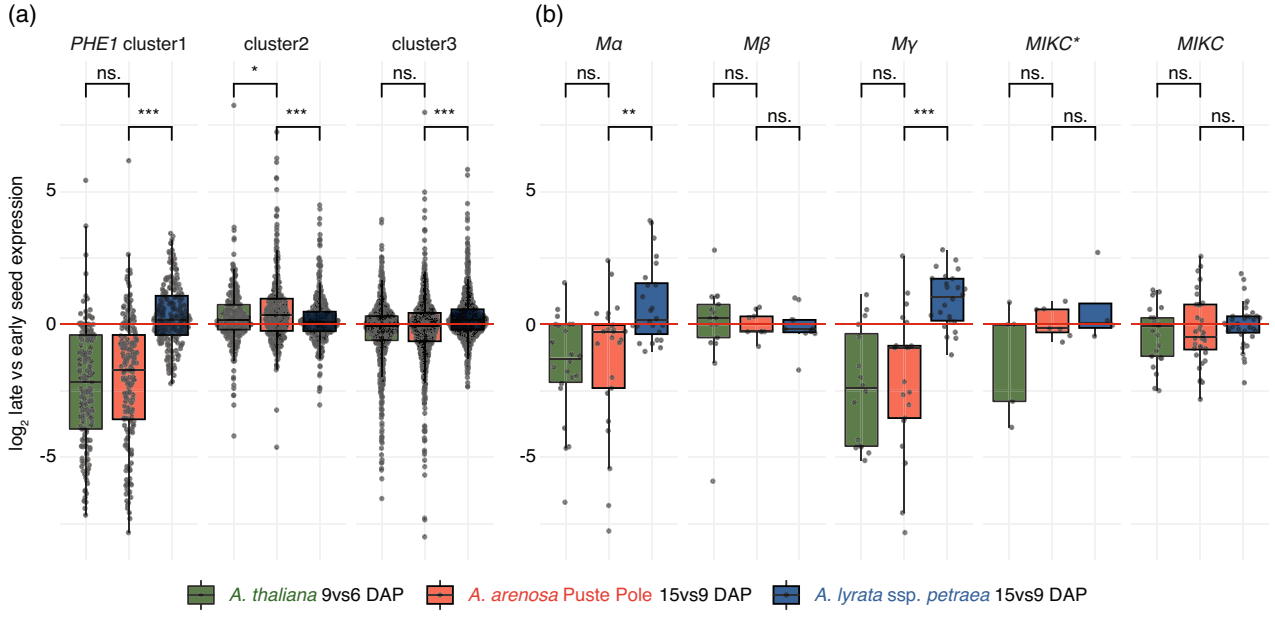


Figure 6. Differential expression of orthologs of *Mγ-1a PHE1*-targets and MADS-box genes before and after endosperm cellularization.
(a) *Mγ-1a PHE1*-targets in *A. thaliana* were clustered depending on their expression during seed development (Batista et al., 2019). A characteristic of cluster 1 is the down-regulation in *A. thaliana* during endosperm cellularization (green). Differential expression of orthologs of *Mγ-1a PHE1*-targets in *A. arenosa* Pusté Pole and *A. lyrata* ssp. *petraea* are plotted in red and blue, respectively. Note that repression in cluster 1 is also seen in *A. arenosa* (red) but not in *A. lyrata* ssp. *petraea* (blue).
(b) Differential gene expression of MADS-box genes before and after endosperm cellularization in *A. thaliana, A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole. Similar to potential orthologs of *Mγ-1a PHE1*-targets, *Mα* and *Mγ* MADS-box type I gene expression declines around endosperm cellularization in *A. thaliana* (green) and *A. arenosa* (red) while increases in *A. lyrata* ssp. *petraea* (blue). DAP, days after pollination. Significance, ** $P < 0.01$;*** $P < 0.001$; n.s. not significant (Wilcoxon signed-rank test).

Mγ-1a targets are differentially regulated in *A. thaliana* and *A. arenosa* versus *A. lyrata* ssp. *petraea* during endosperm cellularization.

Focusing on the MADS-box type I genes, Mα and Mγ genes are down-regulated in *A. arenosa* Pusté Pole and *A. thaliana*, whereas in *A. lyrata* ssp. *petraea* seeds the orthologous Mα and Mγ genes are upregulated (Figure 6b). The Mβ genes do not show significant differences. The substantial differences in Mα and Mγ expression between *A. arenosa* and *A. lyrata* ssp. *petraea* before and after cellularization indicate that these species differ in the regulation of the developmental timing of endosperm development. *A. thaliana* Mα and Mγ expression is regulated similarly to *A. arenosa*. In hybrid seeds between the species, endosperm developmental failure plays an important role in the establishment of species barriers (Bjerkan et al., 2020; Lafon-Placette et al., 2017), and the observed expression difference between the species may be part of the genetic mechanism leading to the species barrier in hybrid seeds.

CONCLUSION

Here we have investigated and compared MADS-box type I gene evolution in the genus Arabidopsis to gene expression in seed stages before and after endosperm cellularization. To do so, we generated chromosome-scale assemblies of *A. lyrata* ssp. *petraea* and *A. arenosa*.

The quality of the reference genomes is essential for knowledge transfer from model species such as *A. thaliana* to their relatives, and research in the genus Arabidopsis has been supported by recent genome sequencing (Hu et al., 2011; Rawat et al., 2015) and improved annotation of the North American *A. lyrata* ssp. *lyrata*. This genome has served as reference for functional, ecological, and evolutionary experiments in *A. lyrata* and also *A. arenosa* (Arnold et al., 2016; Bjerkan et al., 2020; Klosinska et al., 2016; Lafon-Placette et al., 2017; Yant et al., 2013). *A. lyrata* has also been used as the reference for Arabidopsis reference-guided assemblies (Burns et al., 2021; Jaegle et al., 2023; Kolesnikova et al., 2023; Paape et al., 2018). The genus Arabidopsis has been extensively sequenced, e.g. *A. lyrata* ssp. *petraea* (Akama et al., 2014; Paape et al., 2018), *A. halleri* (Briskine et al., 2016; Legrand et al., 2019), *A. kamchatica* (Paape et al., 2018), *A. suecica* (Burns et al., 2021; Jiang et al., 2021; Novikova et al., 2017), and *A. arenosa* (Barragan et al., 2021; Bohutínská et al., 2021; Burns et al., 2021; Liu et al., 2020). The assembly presented here is the only fully *de novo* assembly of *A. arenosa* long reads, and only the fourth chromosome-level genome of any *Arabidopsis* species, to be published and released.

The genome assemblies allowed interspecies comparisons of MADS-box type I orthologs, gene family expansion analysis, and gene expression studies comparing orthologs. We demonstrate using MEME analysis identified motifs that are diagnostic for MADS-box type I clades, an additional guide for MADS-box type I classification but also allowing for future studies of the roles of these motifs. Our analysis suggests that chromosomal localization may explain the high duplication rate of MADS-box type I clades in *A. lyrata* and *A. arenosa* compared to *A. thaliana*. Gene expansion is observed in orthologs located in centromeric regions whereas the same orthologs in a different species do not expand when located on chromosome arms. A negative correlation between gene duplication in MADS-box type I genes and seed-specific expression of these loci was observed, i.e. highly duplicated genes were associated with low expression values. Furthermore, the duplication rate was linked to previously identified MADS-box interaction (Parenicová et al., 2003; Qiu & Köhler, 2022), where highly duplicated genes were associated with less interactions. We also observed substantial differences in type I Mα and Mγ expression between *A. arenosa* and *A. lyrata* ssp. *petraea* before and after endosperm cellularization indicating major differences in developmental regulation in the endosperm between the two species, suggesting a genetic cause for the endosperm-based hybridization barrier developmental failure in *A. arenosa* and *A. lyrata* ssp. *petraea* hybrid seeds.

EXPERIMENTAL PROCEDURES

Plant lines and growth conditions

Arabidopsis thaliana accessions were obtained from the Nottingham Arabidopsis Stock Centre (NASC). The *A. arenosa* population MJ09-4 originates from Nízke Tatry Mts.; Pusté Pole (N 48.8855, E 20.2408) and *A. lyrata* ssp. *petraea* population MJ09-11 originates from lower Austria; street from Pernitz to Pottenstein (N 47.9190, E 15.9755) (Bjerkan et al., 2020; Jørgensen et al., 2011; Lafon-Placette et al., 2017). Seed sterilization was performed by washing in three steps with 70% ethanol, bleach solution (20% Klorin (Lilleborg Industrier, Oslo, Norway), 0.1% Tween20), and wash solution (0.001% Tween20) or by over-night chlorine gas sterilization (Lindsey et al., 2017). Sterile seeds were planted on 0.5 Murashige and Skoog (MS) plates (Murashige & Skoog, 1962) supplemented with 2% sucrose. Seeds were stratified for 1–3 weeks at 4°C before transfer to growth chambers at 18°C long-day conditions (16 h light, 160 Imol/m^2/sec, relative humidity 60–65%). Seedlings were transferred to soil for vegetative growth and vernalized at 9°C for 3 weeks to induce flowering.

DNA isolation

DNA was isolated from one individual of *A. arenosa* Pusté Pole and one individual of *A. lyrata* ssp. *petraea*. All available flowers were harvested as input tissue, which resulted in approximately 1–1.5 g of tissue for each plant. Two different protocols were used: one protocol tailored for Illumina short reads, the other for PacBio Sequel I long reads. In both cases, the extraction protocols were the same for both species. For Illumina sequencing, DNA was isolated with the Ezna Plant DNA isolation kit (Omega Biotek Inc, Norcross, GE, USA). To extract high-molecular-weight (HMW) genomic DNA suitable for long-read sequencing, we first performed nuclei isolation (Rachael Workman et al., 2021), followed by the Nanobind Plant Nuclei Big DNA kit from Circulomics to isolate high molecular-weight DNA. For the isolation of nuclei, the

centrifuge speed used was 3800 ***g*** and the DNA isolation was performed using twice the amount of proteinase K and PL1 buffer, two nanodisks for DNA binding, and three washing steps using PW1. DNA purity and concentrations were checked with a NanoDrop ND-1000 and a Qubit 3 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Fragment length was verified by running 2–5 μl of the isolate on a 0.5% agarose gel overnight at 30 V. Four samples with 2 g of plant tissue including stem, flowers, and leaves were flash-frozen for Hi-C sequencing.

Library preparation and sequencing

For both species, 5 μg of HMW DNA was used to generate a 20 kb library according to the manufacturer's instructions (Pacific Biosciences, Menlo Park, CA, USA). The libraries were sequenced on PacBio Sequel I using Sequel Polymerase v3.0 and Sequencing Chemistry v3.0. Two SMRT cells were sequenced for *A. arenosa* Pusté Pole and three SMRT cells for *A. lyrata* ssp. *petraea*. Loading was performed by diffusion and the movie time was 600 min. Sequencing yielded 1 236 862 reads for *A. arenosa* (estimated coverage 86×), and 1 468 776 reads for *A. lyrata* (estimated coverage 72×). Sequencing on Illumina HiSeq 4000 resulted in 101 609 762 pair-end reads for *A. arenosa* (estimated coverage 169×) and 80 006 646 pair-end reads for *A. lyrata* (estimated coverage 120×). Crosslinked Hi-C DNA sequenced on Illumina HiSeq 4000 resulted in 308 885 555 paired reads (estimated coverage 515×) (Data S1.1). All sequencing was performed at the Norwegian Sequencing Center (NSC; https://www.sequencing.uio.no/).

Genome size estimation

Mercury v1.0 (Rhie et al., 2020) was used to create a k-mer frequency spectrum of the Illumina reads with a k-mer length of 19. High-coverage (≥1000) k-mers were observed to be enriched in ribosomal, mitochondrial, and chloroplast DNA. GenomeScope 2.0 (Ranallo-Benavidez et al., 2020) was used to predict genome size and heterozygosity based on the k-mer spectra. Following the vertebrate genome project (Rhie et al., 2021), all k-mers were included. Relative fluorescence intensities were estimated by flow cytometry (FCM) using fresh plant tissue (Dolezel et al., 2007). *Solanum pseudocapsicum L.* (2C = 2.59 pg) was used as an internal standard. Relative fluorescence intensity of at least 3000 nuclei was recorded using a Partec Space flow cytometer (Partec GmbH, Münster, Germany) equipped with the UV-LED chip (365 nm). FCM results were expressed as fluorescence intensities relative to the unit fluorescence intensity of the internal reference. The estimated haploid genome sizes are comparable to our flow cytometric measurements (Data S2) and estimates in previous studies (Pellicer & Leitch, 2020). The genome size estimates were used to estimate sequencing read coverage.

De novo genome assembly

Six strategies were used to generate a variety of assemblies. One assembly per species was selected for further analysis. Long reads were repeatedly assembled with either Canu v2.1 (Koren et al., 2017), FALCON-unzip v1.3.7 (Chin et al., 2016), or Flye v2.6 (Kolmogorov et al., 2019). Long and short reads were assembled with MaSuRCA v3.3.5 (Zimin et al., 2013). Long reads were also polished with short reads using *LoRDEC* (Salmela & Rivals, 2014), then repeatedly assembled with Canu or Flye. All assemblies were computed on the Saga supercomputer (https://documentation.sigma2.no/hpc_machines/saga.html). Assemblies were compared by total length, N50-type statistics, longest contig, and gene content (BUSCO versions 3.0.2 and 4.14 (Manni et al., 2021; Simão et al., 2015; Waterhouse et al., 2018) using embryophyta_odb9 with 1440 genes from OrthoDB (https://www.orthodb.org/)). For both species, the Canu long-read contig assembly was selected for further processing.

Polishing and haplotig phasing

The contig assemblies were polished using Illumina short reads from the same individual plants with 9–12 iterations of Pilon v1.23 (Walker et al., 2014). BUSCO (Manni et al., 2021; Simão et al., 2015; Waterhouse et al., 2018) found high inclusion rates but also high duplication rates for genes expected to be single copy (Figures S1 and S3), indicating that the assembly might have separated both haplotypes of portions of the diploid genomes. Following the documentation on https://github.com/broadinstitute/pilon/wiki, the Purge Haplotigs v1.1.1 (Roach et al., 2018) and Purge_Dups v1.2.3 (Guan et al., 2020) pipelines were run to remove alternate haplotigs. For quality assessment, see Figures S1–S5 and Data S1.

Reference-guided scaffolding

The *A. lyrata* ssp. *petraea* assembly was scaffolded using RaGOO v1.11 (Alonge et al., 2019) guided by the published *A. lyrata* ssp. *lyrata* assembly (Hu et al., 2011). Manual curation of the computational results revealed the perpetuation of a chloroplast insertion on scaffold 2 and a misassembly on scaffold 1 of the reference (Burns et al., 2021; Henry et al., 2014; Slotte et al., 2013). The computation had inserted a chloroplast-like contig in the first case, and broken several contigs in the second. The contigs and scaffolds were repaired manually in both cases.

Hi-C scaffolding

Hi-C data for *A. arenosa* were generated with Arima kits and Illumina sequencing. The Hi-C reads were mapped to the draft assemblies with BWA-mem v0.7.17 (Li, 2013; Li & Durbin, 2009) and filtered with matlock (https://github.com/phasegenomics/matlock, commit 9fe3fdd). Scaffold candidates were generated by three algorithms (Data S1): SALSA2 (Ghurye et al., 2019), https://github.com/marbl/SALSA, commit ed76685; FALCON-Phase vBeta Update 1 (Kronenberg et al., 2021); and ALLHiC (Zhang et al., 2019), https://github.com/tangerzhang/ALLHiC, commit ffaa10e. The FALCON-Phase scaffolding did not increase genome contiguity significantly and was excluded from further analysis (see Data S1). Scaffolded assemblies were compared to each other by constructing whole genome alignments between the newly scaffolded genomes, as well as to the *A. lyrata* ssp. *lyrata* reference genome (Figures S4 and S5). Scaffolds created by SALSA2 and ALLHiC were visualized in Juicebox (Durand et al., 2016) and curated with the Juicebox Assembly Tools, JBAT (Dudchenko et al., 2018), https://github.com/aidenlab/Juicebox, commit 12bc674 (Figure S6). The JBAT version of the ALLHiC scaffolds was selected for further processing. The scaffolds representing both species were processed in the PBJelly gap-filling pipeline v15.8.24 (English et al., 2012).

Repeat classification and masking

One of the largest influences on the quality of the gene prediction and the number of predicted genes was the initial RepeatModeler repeat masking step, which was carefully tuned to known repeat and genome models in Arabidopsis. The RepeatModeler v2.0.1 pipeline was run for *de novo* identification of transposable elements (TEs) (Flynn et al., 2020). This included TRF, RepeatScout, RECON TE detection, LTRharvest, and LTR_retriever. Results were merged, clustered, and deduplicated. Repeats were classified by comparison to Dfam v3.1 (Storer et al., 2021). False positives were

removed based on sequence homology (blastn ≤1e-10) to an *A. thaliana* cds database (Araport11_cds_20160703) (Cheng et al., 2017), which was beforehand cleaned for sequences with high sequence homology to known TEs in Brassicales (Dfam_3.1 database, (Storer et al., 2021)). The *de novo* predicted TE families were combined with the Dfam_3.1 and RepBase-20 170 127 databases and used in RepeatMasker v4.0.9 to annotate and soft mask the assemblies before gene annotation (Smit et al., 2015). Centromeric, ribosomal, and telomeric repeats were identified and labeled using previously described sequences (Jin et al., 2020; Kawabe & Charlesworth, 2006; Maheshwari et al., 2017; Rhie et al., 2021).

Genome annotation

The gene prediction was based on the BRAKER v2.1.5 pipeline, which trains GeneMark-EX and AUGUSTUS, including extrinsic evidence from RNA-seq and protein homology (Brůna et al., 2021; Hoff et al., 2016; Hoff et al., 2019). We used available RNA-seq for *A. arenosa* Pusté Pole and *A. lyrata* ssp. *petraea* from the Sequence Read Archive (Leinonen et al., 2011) and complemented this with seed transcriptomes (see below). In addition, the *A. thaliana* proteome was aligned by GenomeThreader (Gremme et al., 2005) to improve our species-specific training further and guide the gene prediction.

Genome quality assessment

After each assembly processing step, quality control was used to assess our results and change or choose software and parameters. Primary genome statistics were determined by QUAST v5.0.2 (Gurevich et al., 2013). Feature Response Curve (FRC) was used to compare sequence quality (https://github.com/vezzi/FRC_align). Merqury v1.3 gave a k-mer-based approach to sequence quality, sequence completes, and duplication rate (Rhie et al., 2020). BUSCO v4.14 (embryophyta_odb9) was used to monitor gene completeness and duplication rate (Manni et al., 2021). The structural variation was addressed using minimap2 (Li, 2018) alignments between assemblies and also the *A. lyrata* ssp. *lyrata* reference genome (Li, 2018). The alignments were further inspected with Assemblytics (Nattestad & Schatz, 2016) and dotPlotly (https://github.com/tpoorten/dotPlotly, commit 1 174 484) for visualization.

Orthology prediction and synteny analysis

Orthogroups were inferred between *A. arenosa* Pusté Pole, *A. lyrata* ssp. *petraea* and *A. thaliana* following the OrthoFinder v2.5.2 tutorials (Emms & Kelly, 2019). For synteny and collinearity assignment of homologous regions, MCScanX (Wang et al., 2012) was used with default parameters. Syntenic relations were further plotted with the RIdeogram package v0.2.2 (Hao et al., 2020). In addition, sequence differences were compared by whole-genome comparisons and structural rearrangements were classified by SyRI vV1.5 (Goel et al., 2019).

Identification of MADS-box genes

In addition to the two genomes produced in this study, we downloaded publicly available high-quality genome assemblies from the genus Arabidopsis from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) and Phytozome 13 (https://phytozome-next.jgi.doe.gov/). The minimum requirement for the assemblies was a BUSCO completeness score above 95% and high contig continuity (i.e. near chromosome length assemblies). As an outgroup to Arabidopsis, we added genomes of *Capsella rubella* and *C. grandiflora* (Data S1) with a similar requirement for BUSCO-score and contig continuity. For genome assemblies without previously predicted genes, we used AUGUSTUS with the -arabidopsis gene model.

We then constructed a database consisting of the known AGAMOUS-like genes (AGLs) from *A. thaliana* (108 in total) retrieved from the Arabidopsis Information Resource (TAIR, https://www.arabidopsis.org/). AGL26 (At5g26880) was excluded from the database since this is a gene coding for an RNA methyltransferase and does not have a MADS box. In addition, we identified five MADS-box genes from the PlantTFDB website (http://planttfdb.cbi.pku.edu.cn/), which do not currently have an AGL designation in *A. thaliana*. For the list of the genes with their corresponding AGL names as well as commonly used names and abbreviations in *A. thaliana*, see Data 4.1. Previously identified MADS-box genes from *A. thaliana* were added as a basic local alignment search tool (BLAST) database in Geneious 2020.2.4 (https://www.geneious.com).

The database of MADS-box genes from *A. thaliana* was used in blastp searches [BLAST+ v2.12.0 (Altschul et al., 1990)] against the predicted genes of *A. lyrata* ssp. *petraea*, *A. arenosa* Pusté Pole, and three published Arabidopsis genomes that met our criteria: *A. lyrata* ssp. *lyrata* (https://phytozome-next.jgi.doe.gov/info/Alyrata_v2_1), *A. arenosa* Strecno (https://www.ncbi.nlm.nih.gov/assembly/GCA_902996965.1), and *A. halleri* (https://www.ncbi.nlm.nih.gov/assembly/GCA_900078215.1) (Data S1.5). In addition, we blasted the predicted genes of two species of *Capsella* against the same database (*Capsella grandiflora*, https://phytozome-next.jgi.doe.gov/info/Cgrandiflora_v1_1 and *Capsella rubella*, https://www.ncbi.nlm.nih.gov/assembly/GCF_000375325.1/). The top five hits for each known AGL were kept for each of the species, with any duplicates removed. All sequences were then annotated with InterProscan (http://www.ebi.ac.uk/interpro implemented in Geneious Prime 2021) to verify the presence of the MADS-box region and sequences with no MADS-box domain were removed. The results from the InterProscan v5.47 analysis were also used to identify potential mistakes in the predicted gene sequences which were manually corrected (Data S3.2).

Phylogenetic analysis

A multiple sequence alignment was constructed from the blast results together with the canonical A. *thaliana* MADS-box genes from TAIR (https://www.arabidopsis.org) and PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) with MUSCLE version 3.8.425 (Edgar, 2004). A phylogenetic tree was constructed based on the alignment with FastTree2 v2.1.11 (Price et al., 2010) and visualized in Geneious Prime 2021. The resulting tree was manually inspected for any outliers and long branches. Long-branched clades and sequences without any clear homologs to known MADS-box genes were pruned from the tree and removed from the alignment, the remaining sequences were realigned, and a new tree was constructed with FastTree2. We repeated the process of tree building, pruning, and realignment until no more spurious clades remained in the tree. The resulting alignment with 841 amino acid sequences was refined with MUSCLE, trimmed with trimAl, and a final phylogenetic tree was constructed with IQ-tree v1.6.12 with ultrafast bootstrap approximation and SH-like approximate likelihood ratio test (Minh et al., 2020). The trees in Figure 2 and Figure S8 were visualized in FigTree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Motif analysis with MEME

To identify conserved motifs between the MADS-box genes of the Arabidopsis genus, MEME version 5.1.1 (Multiple Expectation

Maximization for Motif Elicitation, http://meme-suite.org/) was used. We scanned for the 50 most common motifs with lengths above 20 AA; a graphical representation of the relative position of the motifs on the sequences can be found in Figure S8. The full results, motif sequences, and a html version can be found at https://github.com/PaulGrini/Arabidopsis_assemblies. The full-length sequences of the proteins were grouped based on the phylogenetic analysis and commonly occurring MEME motifs into Mα, Mβ, Mγ, MIKC*, and MIKC. A principal component analysis was performed on the occurrence of MEME motifs using the PCAtools package in R (Blighe & Lun, 2022).

Analysis of syntenic regions surrounding Mγ-1

Genomic sequences of *Mγ-1* and their surrounding upstream and downstream regions from *A. thaliana*, *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole were aligned with MAFFT v7.470 (Katoh & Standley, 2013) to identify syntenic regions and localize transpositions. SyRI (Synteny and Rearrangement Identifier, https://github.com/swbak/SyRI commit 29a9272 (Goel et al., 2019)) was used to detect and classify structural differences between genomes, that is, inversions, duplications, translocations, inverted duplications, and inverted translocations.

Expression analysis of MADS-box genes in *A. arenosa* and *A. lyrata* compared to *A. thaliana*

Seeds of *A. lyrata* ssp. *petraea* and *A. arenosa* Pusté Pole were dissected from siliques at 9 and 15 DAP and shock frozen in liquid nitrogen. Around 20 seeds were pooled into one tube, and RNA was extracted from three to four replicates per plant and stage using the Spectrum Plant Total RNA kit (SIGMA). MagNA Lyser Green Beads (Roche, Basel, Switzerland) was used for the initial lysis step as described in Shirzadi et al. (Shirzadi et al., 2011). RNA concentration and quality were measured with a NanoDrop ND-1000, Bioanalyzer 2100 Agilent RNA 6000 Nano kit, and a Qubit 3 fluorometer (Thermo Fisher Scientific) using the Qubit RNA BR Assay kit (Invitrogen, Waltham, MA, USA). All kits were used according to the manufacturer's instructions. Whole cDNA libraries were prepared by the NSC from total RNA using the Illumina TruSeq Standard mRNA Library Prep kit. The samples were sequenced on an Illumina (HiSeq 4000) sequencer creating 150 bp pair-end reads with 350 bp inserts. TrimGalore v0.4.4, a wrapper script around Cutadapt and FastQC (Krueger, 2016), was used for adapter and quality trimming. The reads were mapped using HISAT2 v2.2.1 (Kim et al., 2019) and default parameters and quantified by featureCounts (Liao et al., 2014) and DESeq2 library normalization (Love et al., 2014). MADS-box type I genes were selected, grouped, and ordered following the MADS-box phylogeny (Figure 2). The log2-transformed reads were displayed in a heat map drawn with the iheatmapr v0.5.2.9000 package (in Figure 4 and Figure S11) (Schep & Kummerfeld, 2017) including as a reference previously published *A. thaliana* seed expression samples (Bjerkan et al., 2020). The reads were reanalyzed using the same (described above) processing steps.

ACKNOWLEDGEMENTS

The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities. We thank the Laboratory of Flow Cytometry, the Institute of Botany, Academy of Sciences (Czech Republic) for their assistance.

AUTHOR CONTRIBUTIONS

PEG, JB, AKK, KNB & AKB designed the research; JB, AKK, KNB, IMJ, RMA & KSH performed the experiments; JB, AKK, KNB, RMA, KSH, JRM, AKB & PEG analyzed and discussed the data; JB, AKK, KNB, AKB & PEG wrote the article; All authors revised and approved the article.

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Pipeline for genome assembly. *A. lyrata* ssp. *petraea* (blue) and *A. arenosa* Pusté Pole (red). *A. arenosa* was assembled *de novo* from PacBio long-read data and scaffolded using Hi-C data. For *A. lyrata* ssp. *petraea* a draft assembly was made de novo from PacBio long-reads, which were subsequently scaffolded against the published genome of *A. lyrata* ssp. *lyrata* (Hu et al., 2011). (a) Pipeline indicating each step in the assembly process. For each step additional programs or methods were tested; details of these analyses can be found in Data S1. (b) Statistics used in quality control and software decisions. Sequence contiguity is shown in NG50, using the calculated genome size of 201 144 702 bp for *A. lyrata* ssp. *petraea*, and 179 232 250 bp for *A. arenosa* Pusté Pole. (c) BUSCO scores using the gene set for Embryophyta odb9 with 1440 single-copy genes. For the assembly steps, completeness and duplication can be seen in light and dark blue, respectively. The combination of PacBio long reads with Canu created near-complete assemblies. Allelic duplications were identified and removed using Purge_Dups. (d) Phred Quality Score (QV) was calculated with Merqury (Rhie et al., 2020). A QV of 30 corresponds to 99.9% accuracy and QV of 40 to 99.99%. The quality of both long read assemblies was improved by the Pilon polishing step using Illumina reads. While the *A. lyrata* ssp. *petraea* assembly shows a consistently lower error rate, both assemblies have high quality during all assembly steps. For more assemblies and detailed statistics see Data S1.

Figure S2. GenomeScope profiles. Mercury produced the k-mer frequency spectra of Illumina reads (blue) for a k-mer length of 19 nt. To this spectrum, the GenomeScope model (black line) was fitted, and genome length, repetitiveness, and heterozygosity rates were predicted. (a) K-mer spectrum for *A. lyrata* ssp. *petraea*. GenomeScope v2.0 infers a genome length of 201 144 702 bp with a heterozygosity rate of 1.46%. The heterozygous and homozygous peaks were detected with an approximal coverage of 50 and 100, respectively. K-mers with coverage above 1000 were enriched for ribosomal, mitochondrial, and chloroplast DNA. (b) K-mer spectrum for *A. arenosa* Pusté Pole. The heterozygous and homozygous peaks were detected with an approximal coverage of 50 and 100, respectively. GenomeScope infers a genome length of 179 232 250 bp with a heterozygosity rate of 1.61%.

Figure S3. Copy number spectrum plots. For quality control, the k-mer frequencies from every assembly were compared to the k-mer frequencies of the corresponding Illumina reads. The k-mer spectrum shows how many unique 19 k-mer exist (Count) with a specific coverage (k-mer multiplicity). Each Illumina read k-mer is further grouped and colored depending on how often it is found in the assemblies. The first peak at half coverage is expected to

contain k-mers found only on one haplotype, while the second peak should include k-mer from both haplotypes. (a) Copy number spectrum plot (spectra-cn) for the primary *A. lyrata* ssp. *petraea* assembly. The assembly was created with the Canu assembler using PacBio long reads. The spectrum indicates a high-quality assembly where nearly all k-mers found in the Illumina reads (not used for the assembly) were also detected in the assembly with their expected frequencies. K-mers found only in the long-read assembly are displayed as a red/blue bar to the left of the spectrum plot. (b) Spectra-cn plot for the primary *A. arenosa* Pusté Pole assembly. The Canu assembler using uncorrected PacBio long-reads created the most complete draft assembly. Slightly reduced haploid resolution compared to the *A. lyrata* ssp. *petraea* assembly. (c) Spectra-cn plot of the *A. lyrata* ssp. *petraea* assembly after the Pilon polishing step. Using Illumina reads for polishing reduced the number of suspected erroneous k-mers found only in the assembly (blue/red bar on the left). (d) Spectra-cn plot of the *A. arenosa* Pusté Pole assembly after the Pilon polishing step. (e) Spectra-cn plot of the *A. lyrata* ssp. *petraea* assembly after the haplotic purging step. Only the primary haplotig is compared to the Illumina read set. (f) Spectra-cn plot of the *A. arenosa* Pusté Pole assembly after the haplotig purging. K-mer frequencies for the primary haplotig are shown. Half of the single-copy k-mers are missing and are found in the alternative haplotig contigs.

Figure S4. *A. lyrata* ssp. *petraea*-ssp. *lyrata* alignments. After each processing step, all *A. lyrata* ssp. *petraea* contigs and scaffolds were aligned to the *A. lyrata* ssp. *lyrata* reference genome as quality control. Larger misassemblies were easily detected and excluded. The reference scaffolds are on the x-axis, while the respective new assemblies are sorted on the y-axis. (a) Alignment of the *A. lyrata* ssp. *petraea* Canu draft assembly. (b) Alignment of the contigs after the Pilon polishing step shows a high duplication rate compared to the haploid reference. (c) Alignment of the contigs after the haplotic purging step. Contiguity remained after the removal of duplicated sequences. (d) Reference alignment after the reference-based scaffolding step using RaGoo. (e) Alignment after the gab closure by PBJelly displays no considerable changes. (f) Alignment of our final curated *A. lyrata* ssp. *petraea* assembly. The rearrangement of the previously reported misassembly for *A. lyrata* ssp. *lyrata* can be seen between scaffolds one and two (sorted by size) in the top left corner.

Figure S5. *A. arenosa* Pusté Pole-*A. lyrata* ssp. *lyrata* alignments. For quality control, all *A. arenosa* Pusté Pole sequences were aligned to the *A. lyrata* ssp. *lyrata* reference genome. Although structural variation is expected between the species, inconsistent rearrangements between assemblies were fast detected and excluded. The y-axis represents the length of sorted scaffolds from the published *A. lyrata* ssp. *lyrata* genome. (a) Contigs of the *A. arenosa* Pusté Pole Canu draft assembly on the y-axis show a high number of overlaps. (b) Alignment of the same contigs after the Pilon polishing step. (c) Alignment of the contigs after the haplotic purging step. Only the primary haplotig is shown. Contiguity remains after the removal of duplicated sequences. (d) The alignment plots were especially helpful to remove erroneous Hi-C scaffolding approaches. The best performance displayed here is our selected approach using ALLHiC. (e) Minor scaffolding mistakes were curated with the Juicebox Assembly Tools (JBAT) based on the Hi-C linkage map. The curated result aligned astoundingly well with the *A. lyrata* ssp. *lyrata* scaffolds. (f) Gap closure with PBJelly did not change the scaffold arrangement. The previously reported misassembly on *A. lyrata* ssp. *lyrata* scaffold can be seen in the top left corner. In addition, one more extensive inversion is displayed on scaffold 7, in the center of the plot.

Figure S6. *A. arenosa* Pusté Pole Hi-C contact maps. Hi-C contact heat maps indicate the number of contacts between any given pair of loci in the assembly (red scale). Scaffolds are indicated by a blue line, and the green line documents the manual separation used for rearrangements. (a) Hi-C assembly heat map for *A. arenosa* Pusté Pole produced by the SALSA scaffolder. Strong signals far away from the diagonal were used for further scaffolding improvements. (b) The resulting contact heat map of the JBAT curated SALSA scaffolds. (c) Hi-C map of the *A. arenosa* Pusté Pole assembly scaffolded by the ALLHiC software. Problematic sequences are indicated by the low number of connections and stronger far-away signals. (d) Contact matrix after JBAT curation of miss joints and inversions. Green triangles indicate scaffold brakes. This assembly was selected and finalized by gap-filling with PBJelly (Figures S5e,f).

Figure S7. Genome alignments indicate genomic rearrangements. SyRI (Synteny and Rearrangement Identifier) was used to detect and classify structural differences between genomes. (a) Comparison of our *A. lyrata* ssp. *petraea* assembly (dark blue) to the *A. lyrata* ssp. *lyrata* reference (light blue). Syntenic regions are shown as gray blocks, while structural differences are grouped in translocations (green), inversions (orange), and duplications (blue). For visualization purposes, the relocations between non-homolog scaffolds have been excluded (e.g., for translocation between scaffolds one and two, see Figure 1e). (b) Synteny alignment between *A. lyrata* ssp. *petraea* (dark blue) and *A. arenosa* Pusté Pole (red). Structural differences are indicated between largely syntenic scaffolds.

Figure S8. Extended MADS-box phylogeny. Phylogeny of all 841 identified MADS-box genes in Arabidopsis and *Capsella*. The groups correspond to previously published results with type I genes divided into three main groups (Mα, Mβ, Mγ), while type II genes fell into two monophyletic groups, MIKC and MIKC* [also referred to as Mδ] (Arora et al., 2007; Gramzow & Theißen, 2013; Henschel et al., 2002; Parenicová et al., 2003; Qiu & Köhler, 2022; Thangavel & Nayar, 2018). The groups are colored with a red box around Mα, blue around Mβ green around Mγ, yellow around MIKC*, and gray around MIKC. Sequences of *A. thaliana* are marked in blue, and the outgroup *Capsella* in orange. Monophyletic groups of Arabidopsis MADS-box genes, sharing a common ancestor with genes from the outgroup *Capsella*, are numbered starting from the top of the tree and delineated with solid lines. Additional gene duplications that occurred in the common ancestor of Arabidopsis, but after Arabidopsis and *Capsella* separated, are delineated with a dashed line and marked with an additional letter. For instance, "Mγ-1a" and "Mγ-1b" indicate that these two clades share a last common ancestor with *C. rubella* (i.e. "Mγ-1"), but have a gene duplication specific for Arabidopsis that occurred after the separation of Arabidopsis and *Capsella*. The right-hand column contains the result of the MEME analysis for each sequence. For the full result of the MEME analysis consult the material on GitHub.

Figure S9. Phylogenetic analysis of MADS-box type II genes in Arabidopsis. The tree was derived by a maximum likelihood analysis of 275 identified MADS-box type I sequences from *A. thaliana, A. lyrata* ssp. *petraea, A. arenosa* Strecno, *A. arenosa* Pusté Pole, and *A. halleri* with two species of *Capsella* (89 sequences; not shown in the figure) used as outgroup. Solid branches represent bootstrap support >85%, while branches with support values <85% are dashed. The root of the tree is placed between type I and type II genes (the corresponding tree and heatmap for type I genes can be found in Figure 2). Triangles represent clades where branches are collapsed at the most recent gene duplication event in the last common ancestor of the genus Arabidopsis. The length

of the triangles corresponds to the overall branch length of the collapsed clade, see the main text for naming schemes of clades. The heatmap shows the number of gene copies for each clade in the genomes of *A. thaliana, A. lyrata* ssp. *petraea, A. arenosa* Strecno, *A. arenosa* Pusté Pole, and *A. halleri*. The column next to the heatmap indicates the canonical AGL names of the genes in *A. thaliana* found in the corresponding clade; "none" means that a gene representing the clade is not found in the *A. thaliana* genome, while "new" indicates that the gene does not have a given AGL name. The last column shows a simplified representation of the MEME motifs. A fully expanded phylogenetic tree with individual tip labels, support values for all branches, and the outgroup *Capsella* can be found in Figure S8. Results from the full MEME analysis can be found in Figure S8.

Figure S10. PCA of the distribution of MEME motifs on MADS-box type I and type II genes. The PCA was constructed from motifs identified by MEME and counted for each clade in the phylogeny. The first principal component axis captures 24.8% of the variation, and the second axis 10.2%. A polygon is drawn around each of the main groups with Mα in red, Mβ in blue, Mγ in green, MIKC* in yellow, and MIKC in gray.

Figure S11. MADS-box type II expression during seed development. Gene expression profiles are displayed for all identified MADS-box type II genes and compared between *A. thaliana* (right column), *A. arenosa* Pusté Pole (middle column), and *A. lyrata* ssp. *petraea* (left column). The *A. thaliana* development time series (Bjerkan et al., 2020) serves as a reference. The endosperm cellularization in *A. thaliana* occurs between the 6 and 9 days after pollination (DAP). To adjust for the relatively slower development in *A. arenosa* Pusté Pole and *A. lyrata* ssp. *petraea*, corresponding stages before and after endosperm cellularization were sampled at 9 and 15 DAP, respectively. Ortholog genes are grouped and ordered and follow our MADS-box phylogeny (Figure 2). Sample normalized counts are shown with a base-2 logarithmic scale. The MADS-box type II seed expressions remain rather constant compared to the strong decline of type I expression around endosperm cellularization (Figure 4). However, expression differences can be seen between the *Arabidopsis* species. The *SEPALLATA* (*MIKC-1/2/3*) genes are strongly expressed in *A. lyrata* ssp. *petraea* and *A. arenosa* seeds while only weakly or absent in *A. thaliana*. In addition to their largely functional redundant role in flower development and ovary formation (Kaufmann et al., 2009; Pelaz et al., 2001), a crucial role in fruit development and ripening has been reported for *SEPALLATA* orthologs in tomato, strawberry, and apple (Ampomah-Dwamena et al., 2002; Schaffer et al., 2013; Seymour et al., 2011). Furthermore, the orthologs of *MIKC-15* (*AGL19*), *MIKC-24* (*AGL16*), *MIKC-26* (*AGL24*) as well as the FLC clade *MIKC-28* to *MIKC-32* show strong expression differences reflecting the different regulations of flowering and strengthen of vernalization between the perennial *A. lyrata* ssp. *petraea* and *A. arenosa* plants compared to the annual *A. thaliana* (Alexandre & Hennig, 2008; Hu et al., 2014; Kemi et al., 2013; Müller-Xing et al., 2022; Schönrock et al., 2006; Soppe et al., 2021). We detect differing expressions of *MIKC*s* between the species. *MIKC*-2* and *MIKC*-3* (*AGL30* and *AGL65*) are expressed in *A. lyrata* ssp. *petraea* and *A. arenosa* seeds but are not detected in *A. thaliana*. On the other hand, *MIKC*-6* and *MIKC*-7* (*AT4G37435* and *AGL33*) show expression in *A. arenosa* and *A. thaliana* but not in *A. lyrata* ssp. *petraea* seeds. *MIKC*s* are known for their transcription activity in pollen (Verelst et al., 2007) but have also been detected in the endosperm (Zhang et al., 2018). In addition, a high level of redundant heterodimers has been reported and double mutants reduce pollen fertility (Adamczyk & Fernandez, 2009).

Figure S12. MAFFT alignments of genomic sequences containing and surrounding *Mγ-1*. Protein coding sequences (CDS), which are displayed as dark blue arrows of *PHE1* orthologs from *A. thaliana* (green), *A. arenosa* (red), *A. lyrata* ssp. *lyrata* (light blue) and *A. lyrata* ssp. *petraea* (dark blue), are placed in the center. *A. thaliana* is highlighted by a light green background. On top, the consensus with identity score indicates a high similarity of the 3 prime regions downstream of the *Mγ-1* loci. Around 2200 bp 3 prime of *PHE1* lies the repeat-rich region crucial for its parental-specific expression (inside the AT1TE79790 RC/Helitron, highlighted by a transparent gray box). Similar locations of Line 1 (light green arrows) and LTR transposons (light blue arrows) can be found next to the *Mγ-1* loci between species. Further repeats are marked as light gray arrows and lncRNA as violet arrows (XR_002328948.1 and XR_002334149.1).

Data S1. An overview of the sequencing effort for *A. arenosa* and *A. lyrata* ssp. *petraea*

Data S2. Flow cytometry and k-mer estimation of genome size

Data S3. Sequence names for all MADS-box genes included in the phylogeny

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available.

DATA AVAILABILITY STATEMENT

All sequences generated in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with project number PRJNA844220. The *A. lyrata* ssp. *petraea* genome is deposited at NCBI with the reference number GCA_026151145.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_026151145.1/), the *A. arenosa* Pusté Pole genome has reference number GCA_026151155.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_026151155.1/). An accompanying GitHub repository can be found at https://github.com/PaulGrini/Arabidopsis_assemblies. The repository contains predicted genes, scripts, and additional information for the analyses in this paper.

REFERENCES

Adamczyk, B.J. & **Fernandez, D.E.** (2009) MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiology*, **149**, 1713–1723.

Akama, S., **Shimizu-Inatsugi, R.**, **Shimizu, K.K.** & **Sese, J.** (2014) Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid Arabidopsis. *Nucleic Acids Research*, **42**, e46.

Alexandre, C.M. & **Hennig, L.** (2008) FLC or not FLC: the other side of vernalization. *Journal of Experimental Botany*, **59**, 1127–1135.

Alonge, M., **Soyk, S.**, **Ramakrishnan, S.**, **Wang, X.**, **Goodwin, S.**, **Sedlazeck, F.J.** ***et al.*** (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, **20**, 224.

Altschul, S.F., **Gish, W.**, **Miller, W.**, **Myers, E.W.** & **Lipman, D.J.** (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**(3), 403–410. Available from: https://doi.org/10.1016/S0022-2836(05)80360-2

Ampomah-Dwamena, C., **Morris, B.A.**, **Sutherland, P.**, **Veit, B.** & **Yao, J.-L.** (2002) Down-regulation of TM29, a tomato SEPALLATA homolog, causes

parthenocarpic fruit development and floral reversion. *Plant Physiology*, **130**, 605–617.

Arnold, B.J., **Lahner, B.**, **DaCosta, J.M.**, **Weisman, C.M.**, **Hollister, J.D.**, **Salt, D.E.** ***et al.*** (2016) Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 8320–8325.

Arora, R., **Agarwal, P.**, **Ray, S.**, **Singh, A.K.**, **Singh, V.P.**, **Tyagi, A.K.** ***et al.*** (2007) MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics*, **8**, 242.

Barragan, A.C., **Collenberg, M.**, **Schwab, R.**, **Kerstens, M.**, **Bezrukov, I.**, **Bemm, F.** ***et al.*** (2021) Homozygosity at its Limit: Inbreeding Depression in Wild Arabidopsis arenosa Populations. *bioRxiv* 2021.01.24.427284. Available from: https://doi.org/10.1101/2021.01.24.427284

Batista, R.A., **Moreno-Romero, J.**, **Qiu, Y.**, **van Boven, J.**, **Santos-González, J.**, **Figueiredo, D.D.** ***et al.*** (2019) The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife*, **8**, e50541.

Bemer, M., **Heijmans, K.**, **Airoldi, C.**, **Davies, B.** & **Angenent, G.C.** (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiology*, **154**, 287–300.

Bemer, M., **Wolters-Arts, M.**, **Grossniklaus, U.** & **Angenent, G.C.** (2008) The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell*, **20**, 2088–2101.

Bjerkan, K.N., **Hornslien, K.S.**, **Johannessen, I.M.**, **Krabberød, A.K.**, **Ekelenburg, Y.S.**, **Kalantarian, M.** ***et al.*** (2020) Genetic variation and temperature affects hybrid barriers during interspecific hybridization. *The Plant Journal*, **101**, 122–140.

Blighe, K. & **Lun, A.** (2022) PCAtools: Everything Principal Components Analysis. R package version 2.10.0. Github. Available from: https://github.com/kevinblighe/PCAtools

Bohutínská, M., **Handrick, V.**, **Yant, L.**, **Schmickl, R.**, **Kolář, F.**, **Bomblies, K.** ***et al.*** (2021) De-novo mutation and rapid protein (co-)evolution during meiotic adaptation in Arabidopsis arenosa. *Molecular Biology and Evolution*, **38**, 1980–1994.

Briskine, R.V., **Paape, T.**, **Shimizu-Inatsugi, R.**, **Nishiyama, T.**, **Akama, S.**, **Sese, J.** ***et al.*** (2016) Genome assembly and annotation of Arabidopsis halleri, a model for heavy metal hyperaccumulation and evolutionary ecology. *Molecular Ecology Resources*, **17**, 1025–1036.

Brůna, T., **Hoff, K.J.**, **Lomsadze, A.**, **Stanke, M.** & **Borodovsky, M.** (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, **3**, lqaa108.

Burkart-Waco, D., **Josefsson, C.**, **Dilkes, B.**, **Kozloff, N.**, **Torjek, O.**, **Meyer, R.** ***et al.*** (2012) Hybrid incompatibility in Arabidopsis is determined by a multiple-locus genetic network. *Plant Physiology*, **158**, 801–812.

Burkart-Waco, D., **Ngo, K.**, **Dilkes, B.**, **Josefsson, C.** & **Comai, L.** (2013) Early disruption of maternal-zygotic interaction and activation of defense-like responses in Arabidopsis interspecific crosses. *Plant Cell*, **25**, 2037–2055.

Burkart-Waco, D., **Ngo, K.**, **Lieberman, M.** & **Comai, L.** (2015) Perturbation of parentally biased gene expression during interspecific hybridization. *PLoS One*, **10**, e0117293.

Burns, R., **Mandáková, T.**, **Gunis, J.**, **Soto-Jiménez, L.M.**, **Liu, C.**, **Lysak, M.A.** ***et al.*** (2021) Gradual evolution of allopolyploidy in Arabidopsis suecica. *Nature Ecology & Evolution*, **5**, 1367–1381.

Chen, Z.J., **Comai, L.** & **Pikaard, C.S.** (1998) Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in Arabidopsis allopolyploids. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14891–14896.

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

Chin, C.-S., **Peluso, P.**, **Sedlazeck, F.J.**, **Nattestad, M.**, **Concepcion, G.T.**, **Clum, A.** ***et al.*** (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, **13**, 1050–1054.

Colombo, M., **Masiero, S.**, **Vanzulli, S.**, **Lardelli, P.**, **Kater, M.M.** & **Colombo, L.** (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *The Plant Journal*, **54**, 1037–1048.

Comai, L., **Tyagi, A.P.**, **Winter, K.**, **Holmes-Davis, R.**, **Reynolds, S.H.**, **Stevens, Y.** ***et al.*** (2000) Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. *Plant Cell*, **12**, 1551–1568.

Dart, S., **Kron, P.** & **Mable, B.K.** (2004) Characterizing polyploidy inArabidopsis lyratausing chromosome counts and flow cytometry. *Canadian Journal of Botany*, **82**, 185–197.

de Folter, S., **Immink, R.G.H.**, **Kieffer, M.**, **Pařenicová, L.**, **Henz, S.R.**, **Weigel, D.** ***et al.*** (2005) Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell*, **17**, 1424–1433.

Dolezel, J., **Greilhuber, J.** & **Suda, J.** (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, **2**, 2233–2244.

Dudchenko, O., **Shamim, M.S.**, **Batra, S.**, **Durand, N.C.**, **Musial, T.N.**, **Mostofa, R.** ***et al.*** (2018) The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv*, 254797. Available from: https://www.biorxiv.org/content/10.1101/254797v1 [Accessed 10th May 2023].

Durand, N.C., **Robinson, J.T.**, **Shamim, M.S.**, **Machol, I.**, **Mesirov, J.P.**, **Lander, E.S.** ***et al.*** (2016) Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Systems*, **3**, 99–101.

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

Emms, D.M. & **Kelly, S.** (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238.

English, A.C., **Richards, S.**, **Han, Y.**, **Wang, M.**, **Vee, V.**, **Qu, J.** ***et al.*** (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*, **7**, e47768.

Erilova, A., **Brownfield, L.**, **Exner, V.**, **Rosa, M.**, **Twell, D.**, **Scheid, O.M.** ***et al.*** (2009) Imprinting of the Polycomb group Gene MEDEA serves as a ploidy sensor in Arabidopsis. *PLoS Genetics*, **5**, e1000663.

Florez-Rueda, A.M., **Paris, M.**, **Schmidt, A.**, **Widmer, A.**, **Grossniklaus, U.** & **Städler, T.** (2016) Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Molecular Biology and Evolution*, **33**, 2935–2946.

Flynn, J.M., **Hubley, R.**, **Goubert, C.**, **Rosen, J.**, **Clark, A.G.**, **Feschotte, C.** ***et al.*** (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, **117**, 9451–9457.

Ghurye, J., **Rhie, A.**, **Walenz, B.P.**, **Schmitt, A.**, **Selvaraj, S.**, **Pop, M.** ***et al.*** (2019) Integrating hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computational Biology*, **15**, e1007273.

Goel, M., **Sun, H.**, **Jiao, W.-B.** & **Schneeberger, K.** (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, **20**, 277.

Gramzow, L. & **Theissen, G.** (2010) A hitchhiker's guide to the MADS world of plants. *Genome Biology*, **11**, 214.

Gramzow, L. & **Theißen, G.** (2013) Phylogenomics of MADS-box genes in plants – two opposing life styles in one gene family. *Biology*, **2**, 1150–1164.

Gremme, G., **Brendel, V.**, **Sparks, M.E.** & **Kurtz, S.** (2005) Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, **47**, 965–978.

Grossniklaus, U., **Vielle-Calzada, J.P.**, **Hoeppner, M.A.** & **Gagliano, W.B.** (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science*, **280**, 446–450.

Guan, D., **McCarthy, S.A.**, **Wood, J.**, **Howe, K.**, **Wang, Y.** & **Durbin, R.** (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, **36**, 2896–2898. Available from: https://doi.org/10.1093/bioinformatics/btaa025

Gurevich, A., **Saveliev, V.**, **Vyahhi, N.** & **Tesler, G.** (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072–1075.

Hao, Z., **Lv, D.**, **Ge, Y.**, **Shi, J.**, **Weijers, D.**, **Yu, G.** ***et al.*** (2020) RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science*, **6**, e251.

Hehenberger, E., **Kradolfer, D.** & **Köhler, C.** (2012) Endosperm cellularization defines an important developmental transition for embryo development. *Development*, **139**, 2031–2039.

Henry, I.M., **Dilkes, B.P.**, **Tyagi, A.**, **Gao, J.**, **Christensen, B.** & **Comai, L.** (2014) The BOY NAMED SUE quantitative trait locus confers increased

meiotic stability to an adapted natural allopolyploid of Arabidopsis. *Plant Cell*, **26**, 181–194.

Henschel, K., **Kofuji, R.**, **Hasebe, M.**, **Saedler, H.**, **Münster, T.** & **Theissen, G.** (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. *Molecular Biology and Evolution*, **19**, 801–814.

Hoff, K.J., **Lange, S.**, **Lomsadze, A.**, **Borodovsky, M.** & **Stanke, M.** (2016) BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, **32**, 767–769.

Hoff, K.J., **Lomsadze, A.**, **Borodovsky, M.** & **Stanke, M.** (2019) Whole-Genome Annotation with BRAKER. In: Kollmar, M. (Ed.) *Gene Prediction: Methods and Protocols*. Springer New York: New York, NY, pp. 65–95.

Hohmann, N., **Wolf, E.M.**, **Lysak, M.A.** & **Koch, M.A.** (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*, **27**, 2770–2784.

Hu, J.-Y., **Zhou, Y.**, **He, F.**, **Dong, X.**, **Liu, L.-Y.**, **Coupland, G.** ***et al.*** (2014) miR824-regulated AGAMOUS-LIKE16 contributes to flowering time repression in Arabidopsis. *Plant Cell*, **26**, 2024–2037.

Hu, T.T., **Pattyn, P.**, **Bakker, E.G.**, **Cao, J.**, **Cheng, J.F.**, **Clark, R.M.** ***et al.*** (2011) The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nature Genetics*, **43**, 476–481.

Jaegle, B., **Pisupati, R.**, **Soto-Jiménez, L.M.**, **Burns, R.**, **Rabanal, F.A.** & **Nordborg, M.** (2023) Extensive sequence duplication in Arabidopsis revealed by pseudo-heterozygosity. *Genome Biology*, **24**, 44.

Jiang, X., **Song, Q.**, **Ye, W.** & **Chen, Z.J.** (2021) Concerted genomic and epigenomic changes accompany stabilization of Arabidopsis allopolyploids. *Nature Ecology & Evolution*, **5**, 1–12.

Jin, J.-J., **Yu, W.-B.**, **Yang, J.-B.**, **Song, Y.**, **DePamphilis, C.W.**, **Yi, T.-S.** ***et al.*** (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, **21**, 241.

Johnston, J.S., **Pepper, A.E.**, **Hall, A.E.**, **Chen, Z.J.**, **Hodnett, G.**, **Drabek, J.** ***et al.*** (2005) Evolution of genome size in Brassicaceae. *Annals of Botany*, **95**, 229–235.

Jørgensen, M.H., **Ehrich, D.**, **Schmickl, R.**, **Koch, M.A.** & **Brysting, A.K.** (2011) Interspecific and interploidal gene flow in central European Arabidopsis (Brassicaceae). *BMC Evolutionary Biology*, **11**, 346.

Josefsson, C., **Dilkes, B.** & **Comai, L.** (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology*, **16**, 1322–1328.

Kang, I.-H., **Steffen, J.G.**, **Portereiko, M.F.**, **Lloyd, A.** & **Drews, G.N.** (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *Plant Cell*, **20**, 635–647.

Katoh, K. & **Standley, D.M.** (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kaufmann, K., **Muiño, J.M.**, **Jauregui, R.**, **Airoldi, C.A.**, **Smaczniak, C.**, **Krajewski, P.** ***et al.*** (2009) Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biology*, **7**, e1000090.

Kawabe, A. & **Charlesworth, D.** (2006) Patterns of DNA variation among three centromere satellite families in Arabidopsis halleri and A. lyrata. *Journal of Molecular Evolution*, **64**, 237–247.

Kemi, U., **Niittyvuopio, A.**, **Toivainen, T.**, **Pasanen, A.**, **Quilot-Turion, B.**, **Holm, K.** ***et al.*** (2013) Role of vernalization and of duplicated FLOWERING LOCUS C in the perennial Arabidopsis lyrata. *New Phytologist*, **197**, 323–335. Available from: https://doi.org/10.1111/j.1469-8137.2012.04378.x

Kim, D., **Paggi, J.M.**, **Park, C.**, **Bennett, C.** & **Salzberg, S.L.** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**, 907–915.

Kinoshita, T., **Yadegari, R.**, **Harada, J.J.**, **Goldberg, R.B.** & **Fischer, R.L.** (1999) Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. *Plant Cell*, **11**, 1945–1952.

Kirkbride, R.C., **Lu, J.**, **Zhang, C.**, **Mosher, R.A.**, **Baulcombe, D.C.** & **Chen, Z.J.** (2019) Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in Arabidopsis. *Proceedings of the National Academy of Sciences, United States of America*, **116**(7), 2761–2766.

Klosinska, M., **Picard, C.L.** & **Gehring, M.** (2016) Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. *Nature Plants*, **2**, 16145.

Köhler, C., **Hennig, L.**, **Spillane, C.**, **Pien, S.**, **Gruissem, W.** & **Grossniklaus, U.** (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes & Development*, **17**, 1540–1553.

Kolesnikova, U.K., **Scott, A.D.**, **Van de Velde, J.D.**, **Burns, R.**, **Tikhomirov, N.P.**, **Pfordt, U.** ***et al.*** (2023) Transition to self-compatibility associated with dominant S-allele in a diploid Siberian progenitor of allotetraploid Arabidopsis kamchatica revealed by Arabidopsis lyrata genomes. *bioRxiv*, 2022.06.24.497443. Available from: https://doi.org/10.1101/2022.06.24.497443v2 [Accessed 18th May 2023].

Kolmogorov, M., **Yuan, J.**, **Lin, Y.** & **Pevzner, P.A.** (2019) Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, **37**, 540–546.

Koren, S., **Walenz, B.P.**, **Berlin, K.**, **Miller, J.R.**, **Bergman, N.H.** & **Phillippy, A.M.** (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, **27**, 722–736.

Kovaka, S., **Ou, S.**, **Jenike, K.M.** & **Schatz, M.C.** (2023) Approaching complete genomes, transcriptomes and epi-omes with accurate long-read sequencing. *Nature Methods*, **20**, 12–16.

Kress, W.J., **Soltis, D.E.**, **Kersey, P.J.**, **Wegrzyn, J.L.**, **Leebens-Mack, J.H.**, **Gostel, M.R.** ***et al.*** (2022) Green plant genomes: what we know in an era of rapidly expanding opportunities. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(4), e2115640118. Available from: https://doi.org/10.1073/pnas.2115640118

Kronenberg, Z.N., **Rhie, A.**, **Koren, S.**, **Concepcion, G.T.**, **Peluso, P.**, **Munson, K.M.** ***et al.*** (2021) Extended haplotype-phasing of long-read de novo genome assemblies using hi-C. *Nature Communications*, **12**, 1935.

Krueger, F. (2016) *TrimGalore: a wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data*. Available from: https://github.com/FelixKrueger/TrimGalore [Accessed 18th May 2023].

Lafon-Placette, C., **Johannessen, I.M.**, **Hornslien, K.S.**, **Ali, M.F.**, **Bjerkan, K.N.**, **Bramsiepe, J.** ***et al.*** (2017) Endosperm-based hybridization barriers explain the pattern of gene flow between Arabidopsis lyrata and Arabidopsis arenosa in Central Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E1027–E1035.

Legrand, S., **Caron, T.**, **Maumus, F.**, **Schvartzman, S.**, **Quadrana, L.**, **Durand, E.** ***et al.*** (2019) Differential retention of transposable element-derived sequences in outcrossing Arabidopsis genomes. *Mobile DNA*, **10**, 30.

Leinonen, R., **Sugawara, H.**, **Shumway, M.** & **the International Nucleotide Sequence Database Collaboration**. (2011) The sequence read archive. *Nucleic Acids Research*, **39**, D19–D21 [Accessed January 20, 2023].

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]*. Available from:http://arxiv.org/abs/1303.3997

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.

Li, H. & **Durbin, R.** (2009) Fast and accurate short read alignment with burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Liao, Y., **Smyth, G.K.** & **Shi, W.** (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Lindsey, B.E., 3rd, **Rivero, L.**, **Calhoun, C.S.**, **Grotewold, E.** & **Brkljacic, J.** (2017) Standardized method for high-throughput sterilization of Arabidopsis seeds. *Journal of Visualized Experiments*, **128**, 56587. Available from: https://doi.org/10.3791/56587

Liu, Z., **Cheema, J.**, **Vigouroux, M.**, **Hill, L.**, **Reed, J.**, **Paajanen, P.** ***et al.*** (2020) Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nature Communications*, **11**, 5354.

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

Lysak, M.A., **Berr, A.**, **Pecinka, A.**, **Schmidt, R.**, **McBreen, K.** & **Schubert, I.** (2006) Mechanisms of chromosome number reduction in Arabidopsis thaliana and related Brassicaceae species. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5224–5229.

Lysak, M.A., **Koch, M.A.**, **Beaulieu, J.M.**, **Meister, A.** & **Leitch, I.J.** (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. *Molecular Biology and Evolution*, **26**, 85–98.

Maheshwari, S., **Ishii, T.**, **Brown, C.T.**, **Houben, A.** & **Comai, L.** (2017) Centromere location in Arabidopsis is unaltered by extreme divergence in CENH3 protein sequence. *Genome Research*, **27**, 471–478.

Makarevich, G., **Villar, C.B.R.**, **Erilova, A.** & **Köhler, C.** (2008) Mechanism of PHERES1 imprinting in Arabidopsis. *Journal of Cell Science*, **121**, 906–912.

Manni, M., **Berkeley, M.R.**, **Seppey, M.**, **Simão, F.A.** & **Zdobnov, E.M.** (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**, 4647–4654.

Marburger, S., **Monnahan, P.**, **Seear, P.J.**, **Martin, S.H.**, **Koch, J.**, **Paajanen, P.** ***et al.*** (2019) Interspecific introgression mediates adaptation to whole genome duplication. *Nature Communications*, **10**, 5218.

Masiero, S., **Colombo, L.**, **Grini, P.E.**, **Schnittger, A.** & **Kater, M.M.** (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*, **23**, 865–872.

Minh, B.Q., **Schmidt, H.A.**, **Chernomor, O.**, **Schrempf, D.**, **Woodhams, M.D.**, **von Haeseler, A.** ***et al.*** (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**, 1530–1534. Available from: https://doi.org/10.1093/molbev/msaa015

Müller-Xing, R., **Ardiansyah, R.**, **Xing, Q.**, **Faivre, L.**, **Tian, J.**, **Wang, G.** ***et al.*** (2022) Polycomb proteins control floral determinacy by H3K27me3-mediated repression of pluripotency genes in Arabidopsis thaliana. *Journal of Experimental Botany*, **73**, 2385–2402.

Murashige, T. & **Skoog, F.** (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, **15**, 473–497.

Naish, M., **Alonge, M.**, **Wlodzimierz, P.**, **Tock, A.J.**, **Abramson, B.W.**, **Schmücker, A.** ***et al.*** (2021) The genetic and epigenetic landscape of the Arabidopsis centromeres. *Science*, **374**, eabi7489.

Nam, J., **Kim, J.**, **Lee, S.**, **An, G.**, **Ma, H.** & **Nei, M.** (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 1910–1915.

Nasrallah, M.E., **Yogeeswaran, K.**, **Snyder, S.** & **Nasrallah, J.B.** (2000) Arabidopsis species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiology*, **124**, 1605–1614.

Nattestad, M. & **Schatz, M.C.** (2016) Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics*, **32**, 3021–3023.

Novikova, P.Y., **Hohmann, N.**, **Nizhynska, V.**, **Tsuchimatsu, T.**, **Ali, J.**, **Muir, G.** ***et al.*** (2016) Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Publishing Group*, **48**, 1077–1082. Available from: https://doi.org/10.1038/ng.3617

Novikova, P.Y., **Tsuchimatsu, T.**, **Simon, S.**, **Nizhynska V**, **Voronin V**, **Burns R** ***et al.*** (2017) Genome sequencing reveals the origin of the allotetraploid Arabidopsis suecica. *Molecular Biology and Evolution*, **34**(4), 957–968. Available from: https://doi.org/10.1093/molbev/msw299

Paape, T., **Briskine, R.V.**, **Halstead-Nussloch, G.**, **Lischer, H.E.L.**, **Shimizu-Inatsugi, R.**, **Hatakeyama, M.** ***et al.*** (2018) Patterns of polymorphism and selection in the subgenomes of the allopolyploid Arabidopsis kamchatica. *Nature Communications*, **9**, 3909.

Parenicová, L., **de Folter, S.**, **Kieffer, M.**, **Horner, D.S.**, **Favalli, C.**, **Busscher, J.** ***et al.*** (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*, **15**, 1538–1551.

Pelaz, S., **Tapia-López, R.**, **Alvarez-Buylla, E.R.** & **Yanofsky, M.F.** (2001) Conversion of leaves into petals in Arabidopsis. *Current Biology*, **11**, 182–184.

Pellicer, J. & **Leitch, I.J.** (2020) The plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *The New Phytologist*, **226**, 301–305.

Price, M.N., **Dehal, P.S.** & **Arkin, A.P.** (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.

Pucker, B., **Irisarri, I.**, **de Vries, J.** & **Xu, B.** (2022) Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quantitative Plant Biology*, **3**, e5 [Accessed 11th November 2022].

Qiu, Y. & **Köhler, C.** (2020) Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochemical Society Transactions*, **48**, 1005–1017. Available from: https://doi.org/10.1042/bst20190937

Qiu, Y. & **Köhler, C.** (2022) Endosperm evolution by duplicated and Neofunctionalized type I MADS-box transcription factors. *Molecular Biology and Evolution*, **39**(1), msab355. Available from: https://doi.org/10.1093/molbev/msab355

Ranallo-Benavidez, T.R., **Jaron, K.S.** & **Schatz, M.C.** (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, **11**, 1432.

Ratcliffe, O.J., **Kumimoto, R.W.**, **Wong, B.J.** & **Riechmann, J.L.** (2003) Analysis of the Arabidopsis MADS AFFECTING FLOWERING gene family: MAF2 prevents vernalization by short periods of cold. *The Plant Cell Online*, **15**, 1159–1169.

Rawat, V., **Abdelsamad, A.**, **Pietzenuk, B.**, **Seymour, D.K.**, **Koenig, D.**, **Weigel, D.** ***et al.*** (2015) Improving the annotation of Arabidopsis lyrata using RNA-seq data. *PLoS One*, **10**, e0137391.

Rebernig, C.A., **Lafon-Placette, C.**, **Hatorangan, M.R.**, **Slotte, T.** & **Köhler, C.** (2015) Non-reciprocal interspecies hybridization barriers in the Capsella genus are established in the endosperm. *PLoS Genetics*, **11**, e1005295.

Rhie, A., **McCarthy, S.A.**, **Fedrigo, O.**, **Damas, J.**, **Formenti, G.**, **Koren, S.** ***et al.*** (2021) Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, **592**, 737–746.

Rhie, A., **Walenz, B.P.**, **Koren, S.** & **Phillippy, A.M.** (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, **21**, 1–27.

Roach, M.J., **Schmidt, S.A.** & **Borneman, A.R.** (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, **19**, 460.

Robinson, J.T., **Thorvaldsdóttir, H.**, **Winckler, W.**, **Guttman, M.**, **Lander, E.S.**, **Getz, G.** ***et al.*** (2011) Integrative genomics viewer. *Nature Biotechnology*, **29**, 24–26.

Roszak, P. & **Köhler, C.** (2011) Polycomb group proteins are required to couple seed coat initiation to fertilization. *Proceedings of the National Academy of Sciences*, **108**, 20826–20831.

Salmela, L. & **Rivals, E.** (2014) LoRDEC: accurate and efficient long read error correction. *Bioinformatics*, **30**, 3506–3514.

Schaffer, R.J., **Ireland, H.S.**, **Ross, J.J.**, **Ling, T.J.** & **David, K.M.** (2013) SEPALLATA1/2-suppressed mature apples have low ethylene, high auxin and reduced transcription of ripening-related genes. *AoB Plants*, **5**, ls047.

Schep, A.N. & **Kummerfeld, S.K.** (2017) Iheatmapr: interactive complex heatmaps in R. *Journal of Open Source Software*, **2**, 359.

Schmickl, R. & **Koch, M.A.** (2011) Arabidopsis hybrid speciation processes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 14192–14197.

Schmickl, R. & **Yant, L.** (2021) Adaptive introgression: how polyploidy reshapes gene flow landscapes. *The New Phytologist*, **230**, 457–461.

Schönrock, N., **Bouveret, R.**, **Leroy, O.**, **Borghi, L.**, **Köhler, C.**, **Gruissem, W.** ***et al.*** (2006) Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes & Development*, **20**, 1667–1678.

Seymour, G.B., **Ryder, C.D.**, **Cevik, V.**, **Hammond, J.P.**, **Popovich, A.**, **King, G.J.** ***et al.*** (2011) A SEPALLATA gene is involved in the development and ripening of strawberry (Fragaria x ananassa Duch.) fruit, a non-climacteric tissue. *Journal of Experimental Botany*, **62**, 1179–1188.

Shirzadi, R., **Andersen, E.D.**, **Bjerkan, K.N.**, **Gloeckle, B.M.**, **Heese, M.**, **Ungru, A.** ***et al.*** (2011) Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin–dependent regulation of AGAMOUS-LIKE36. *PLoS Genetics*, **7**, e1001303. Available from: https://doi.org/10.1371/journal.pgen.1001303

Simão, F.A., **Waterhouse, R.M.**, **Ioannidis, P.**, **Kriventseva, E.V.** & **Zdobnov, E.M.** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Slotte, T., **Hazzouri, K.M.**, **Agren, J.A.**, **Koenig, D.**, **Maumus, F.**, **Guo, Y.L.** ***et al.*** (2013) The Capsella rubella genome and the genomic consequences of rapid mating system evolution. *Nature Genetics*, **45**, 831–835.

Smit, A.F.A., **Hubley, R.** & **Green, P.** (2015) RepeatMasker Open-4.0. 2013–2015.

Soppe, W.J.J., **Viñegra de la Torre, N.** & **Albani, M.C.** (2021) The diverse roles of FLOWERING LOCUS C in annual and perennial Brassicaceae species. *Frontiers in Plant Science*, **12**, 627258.

Steffen, J.G., **Kang, I.-H.**, **Portereiko, M.F.**, **Lloyd, A.** & **Drews, G.N.** (2008) AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiology*, **148**, 259–268.

Storer, J., **Hubley, R.**, **Rosen, J.**, **Wheeler, T.J.** & **Smit, A.F.** (2021) The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, **12**, 2.

Thangavel, G. & **Nayar, S.** (2018) A survey of MIKC type MADS-box genes in non-seed plants: algae, bryophytes, lycophytes and ferns. *Frontiers in Plant Science*, **9**, 510.

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**, 796–815. Available from: https://doi.org/10.1038/35048692

Tonosaki, K., **Sekine, D.**, **Ohnishi, T.**, **Ono, A.**, **Furuumi, H.**, **Kurata, N.** ***et al.*** (2017) Overcoming the species hybridization barrier by ploidy manipulation in the genus Oryza. *The Plant Journal*, **93**, 534–544. Available from: https://doi.org/10.1111/tpj.13803

Verelst, W., **Saedler, H.** & **Münster, T.** (2007) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiology*, **143**, 447–460.

Walia, H., **Josefsson, C.**, **Dilkes, B.**, **Kirkbride, R.**, **Harada, J.** & **Comai, L.** (2009) Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Current Biology*, **19**, 1128–1132.

Walker, B.J., **Abeel, T.**, **Shea, T.**, **Priest, M.**, **Abouelliel, A.**, **Sakthikumar, S.** ***et al.*** (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**, e112963.

Wang, Y., **Tang, H.**, **Debarry, J.D.**, **Tan, X.**, **Li, J.**, **Wang, X.** ***et al.*** (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.

Waterhouse, R.M., **Seppey, M.**, **Simão, F.A.**, **Manni, M.**, **Ioannidis, P.**, **Klioutchnikov, G.** ***et al.*** (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, **35**, 543–548.

Workman, R., **Timp, W.**, **Fedak, R.**, **Kilburn, D.**, **Hao, S.** & **Liu, K.** (2021) High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing. *Protocol Exchange*. Available from: https://www.protocols.io/view/high-molecular-weight-dna-extraction-from-recalcit-4r3l25jrjl1y/v1 [Accessed 20th January 2023].

Yant, L., **Hollister, J.D.**, **Wright, K.M.**, **Arnold, B.J.**, **Higgins, J.D.**, **Franklin, F.C.H.** ***et al.*** (2013) Meiotic adaptation to genome duplication in Arabidopsis arenosa. *Current Biology*, **23**, 2151–2156.

Yoshida, T. & **Kawabe, A.** (2013) Importance of gene duplication in the evolution of genomic imprinting revealed by molecular evolutionary analysis of the type I MADS-box gene family in Arabidopsis species. *PLoS One*, **8**, e73588.

Zhang, S., **Wang, D.**, **Zhang, H.**, **Skaggs, M.I.**, **Lloyd, A.**, **Ran, D.** ***et al.*** (2018) FERTILIZATION-INDEPENDENT SEED-Polycomb Repressive Complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. *Plant Physiology*, **177**, 285–299.

Zhang, X., **Zhang, S.**, **Zhao, Q.**, **Ming, R.** & **Tang, H.** (2019) Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants*, **5**, 833–845.

Zimin, A.V., **Marçais, G.**, **Puiu, D.**, **Roberts, M.**, **Salzberg, S.L.** & **Yorke, J.A.** (2013) The MaSuRCA genome assembler. *Bioinformatics*, **29**, 2669–2677.

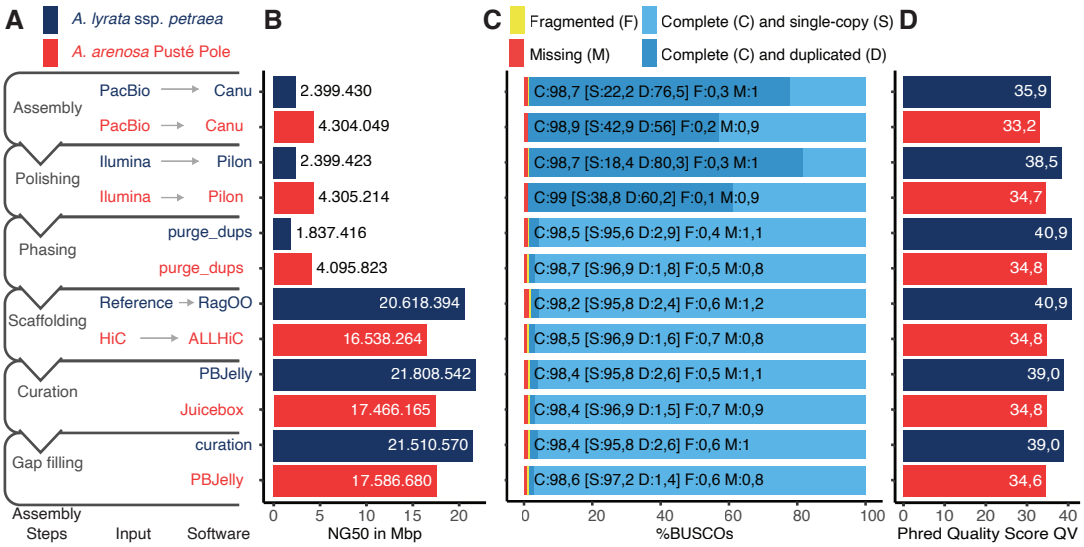


Figure S1. Pipeline for genome assembly. *A. lyrata* ssp. *petraea* (blue) and *A. arenosa* Pusté Pole (red). *A. arenosa* was assembled de novo from PacBio long-read data and scaffolded using Hi-C data. For *A. lyrata* ssp. *petraea* a draft assembly was made de novo from PacBio long-reads, which were subsequently scaffolded against the published genome of *A. lyrata* ssp. *lyrata* (Hu et al., 2011). (a) Pipeline indicating each step in the assembly process. For each step additional programs or methods were tested; details of these analyses can be found in Data S1. (b) Statistics used in quality control and software decisions. Sequence contiguity is shown in NG50, using the calculated genome size of 201 144 702 bp for *A. lyrata* ssp. *petraea*, and 179 232 250 bp for *A. arenosa* Pusté Pole. (c) BUSCO scores using the gene set for Embryophyta odb9 with 1440 single-copy genes. For the assembly steps, completeness and duplication can be seen in light and dark blue, respectively. The combination of PacBio long reads with Canu created near-complete assemblies. Allelic duplications were identified and removed using Purge_Dups. (d) Phred Quality Score (QV) was calculated with Merqury (Rhie et al., 2020). A QV of 30 corresponds to 99.9% accuracy and QV of 40 to 99.99%. The quality of both long read assemblies was improved by the Pilon polishing step using Illumina reads. While the *A. lyrata* ssp. *petraea* assembly shows a consistently lower error rate, both assemblies have high quality during all assembly steps. For more assemblies and detailed statistics see Data S1.

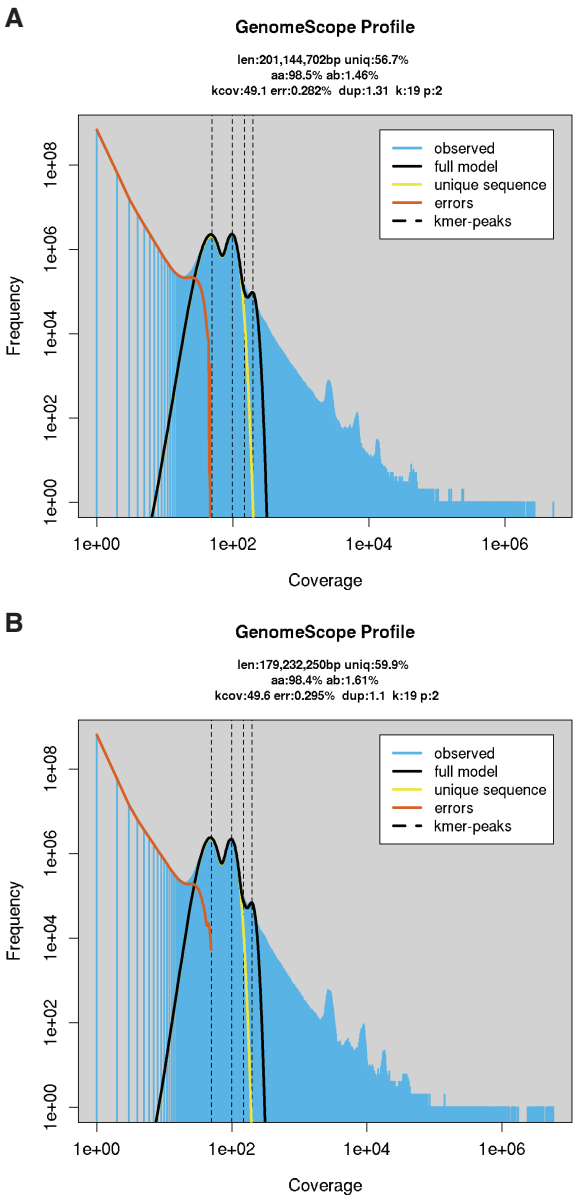


Figure S2. GenomeScope profiles. Mercury produced the k-mer frequency spectra of Illumina reads (blue) for a k-mer length of 19 nt. To this spectrum, the GenomeScope model (black line) was fitted, and genome length, repetitiveness, and heterozygosity rates were predicted. (a) K-mer spectrum for *A. lyrata* ssp. *petraea*. GenomeScope v2.0 infers a genome length of 201 144 702 bp with a heterozygosity rate of 1.46%. The heterozygous and homozygous peaks were detected with an approximal coverage of 50 and 100, respectively. K-mers with coverage above 1000 were enriched for ribosomal, mitochondrial, and chloroplast DNA. (b) K-mer spectrum for *A. arenosa* Pusté Pole. The heterozygous and homozygous peaks were detected with an approximal coverage of 50 and 100, respectively. GenomeScope infers a genome length of 179 232 250 bp with a heterozygosity rate of 1.61%.

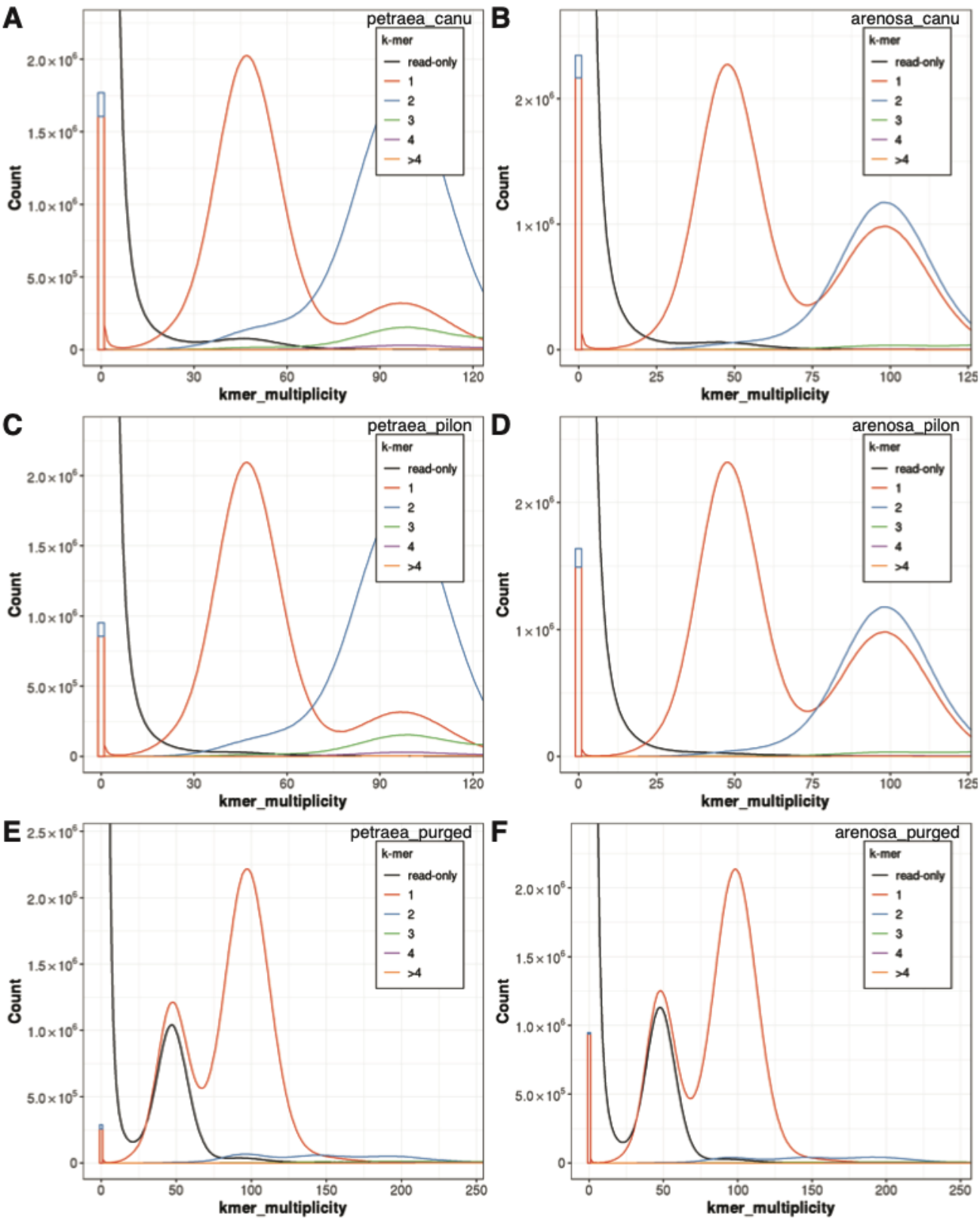


Figure S3. Copy number spectrum plots. For quality control, the k-mer frequencies from every assembly were compared to the k-mer frequencies of the corresponding Illumina reads. The k-mer spectrum shows how many unique 19 k-mer exist (Count) with a specific coverage (k-mer multiplicity). Each Illumina read k-mer is further grouped and colored depending on how often it is found in the assemblies. The first peak at half coverage is expected to contain k-mers found only on one haplotype, while the second peak should include k-mer from both haplotypes. (a) Copy number spectrum plot (spectra-cn) for the primary *A. lyrata* ssp. *petraea* assembly. The assembly was created with the Canu assembler using PacBio long reads. The spectrum indicates a high-quality assembly where nearly all k-mers found in the Illumina reads (not used for the assembly) were also detected in the assembly with their expected frequencies. K-mers found only in the long-read assembly are displayed as a red/blue bar to the left of the spectrum plot. (b) Spectra-cn plot for the primary *A. arenosa* Pusté Pole assembly. The Canu assembler using uncorrected PacBio long-reads created the most complete draft assembly. Slightly reduced haploid resolution compared to the *A. lyrata* ssp. *petraea* assembly. (c) Spectra-cn plot of the *A. lyrata* ssp. *petraea* assembly after the Pilon polishing step. Using Illumina reads for polishing reduced the number of suspected erroneous k-mers found only in the assembly (blue/red bar on the left). (d) Spectra-cn plot of the *A. arenosa* Pusté Pole assembly after the Pilon polishing step. (e) Spectra-cn plot of the *A. lyrata* ssp. *petraea* assembly after the haplotic purging step. Only the primary haplotig is compared to the Illumina read set. (f) Spectra-cn plot of the *A. arenosa* Pusté Pole assembly after the haplotig purging. K-mer frequencies for the primary haplotig are shown. Half of the single-copy k-mers are missing and are found in the alternative haplotig contigs.

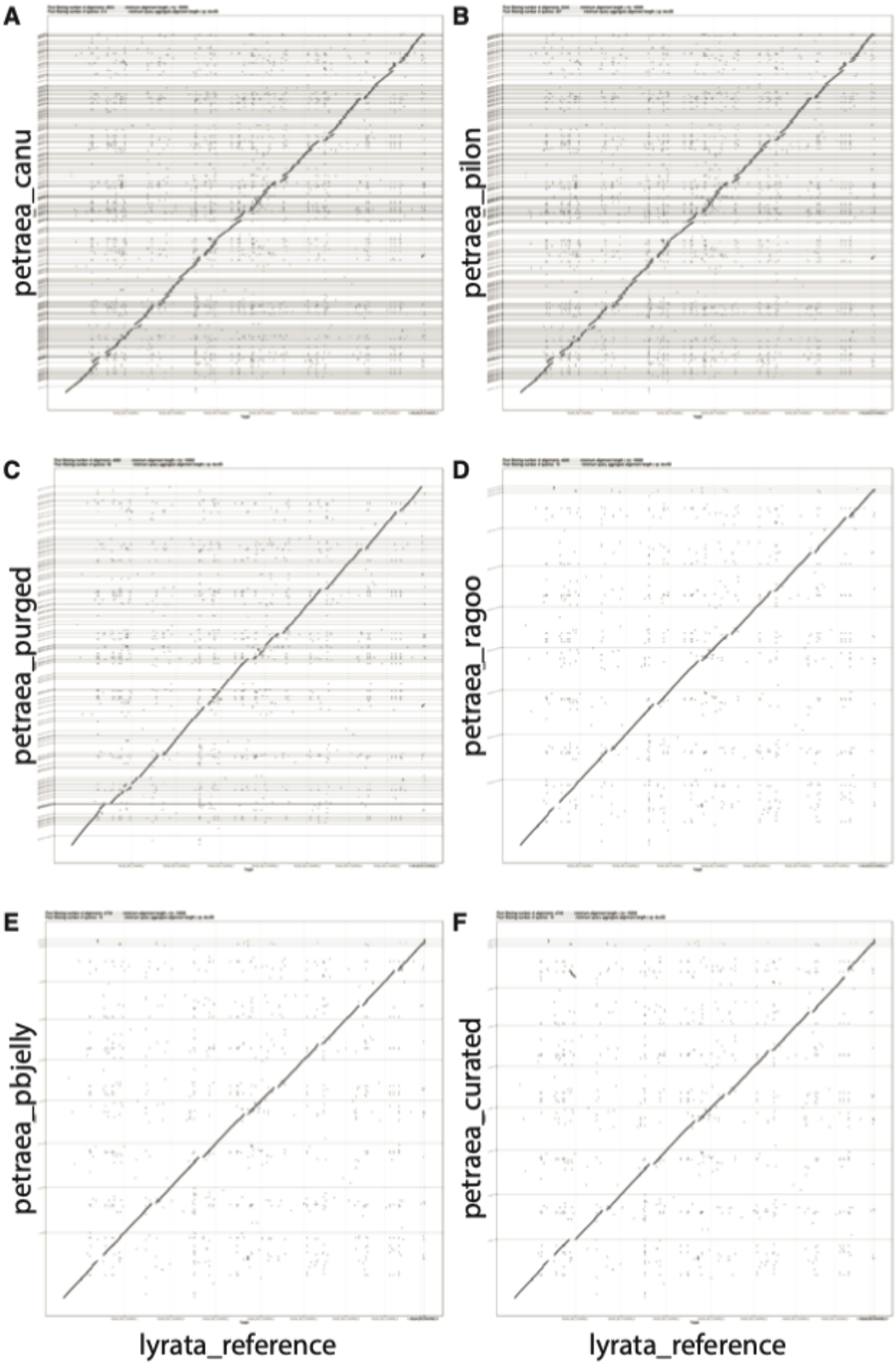


Figure S4. *A. lyrata* ssp. *petraea*-ssp. *lyrata* alignments. After each processing step, all *A. lyrata* ssp. *petraea* contigs and scaffolds were aligned to the *A. lyrata* ssp. *lyrata* reference genome as quality control. Larger misassemblies were easily detected and excluded. The reference scaffolds are on the x-axis, while the respective new assemblies are sorted on the y-axis. (a) Alignment of the *A. lyrata* ssp. *petraea* Canu draft assembly. (b) Alignment of the contigs after the Pilon polishing step shows a high duplication rate compared to the haploid reference. (c) Alignment of the contigs after the haplotic purging step. Contiguity remained after the removal of duplicated sequences. (d) Reference alignment after the reference-based scaffolding step using RaGoo. (e) Alignment after the gab closure by PBJelly displays no considerable changes. (f) Alignment of our final curated *A. lyrata* ssp. *petraea* assembly. The rearrangement of the previously reported misassembly for *A. lyrata* ssp. *lyrata* can be seen between scaffolds one and two (sorted by size) in the top left corner.

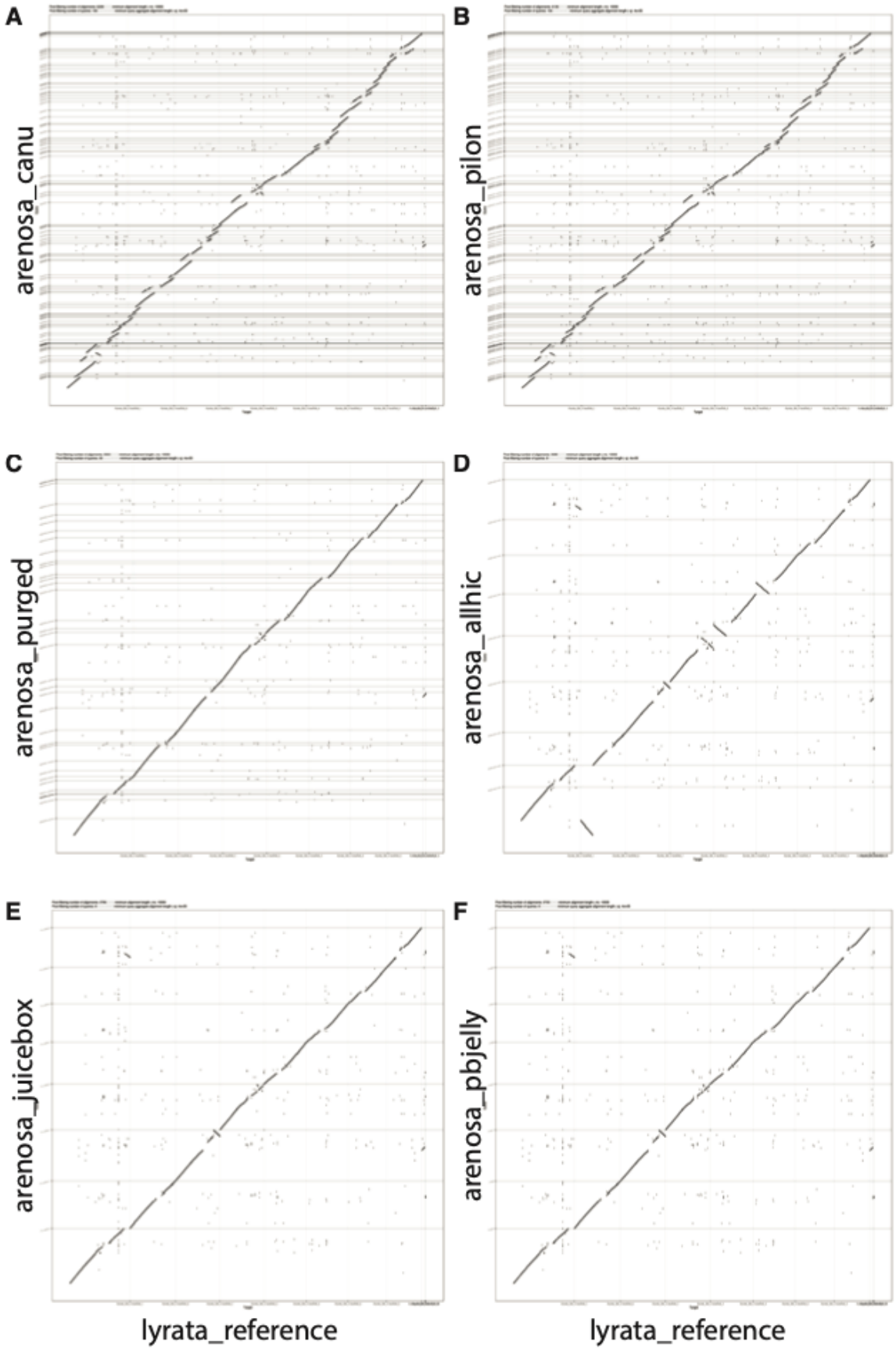


Figure S5. *A. arenosa* Pusté Pole-*A. lyrata* ssp. *lyrata* alignments. For quality control, all *A. arenosa* Pusté Pole sequences were aligned to the *A. lyrata* ssp. *lyrata* reference genome. Although structural variation is expected between the species, inconsistent rearrangements between assemblies were fast detected and excluded. The y-axis represents the length of sorted scaffolds from the published *A. lyrata* ssp. *lyrata* genome. (a) Contigs of the *A. arenosa* Pusté Pole Canu draft assembly on the y-axis show a high number of overlaps. (b) Alignment of the same contigs after the Pilon polishing step. (c) Alignment of the contigs after the haplotic purging step. Only the primary haplotig is shown. Contiguity remains after the removal of duplicated sequences. (d) The alignment plots were especially helpful to remove erroneous Hi-C scaffolding approaches. The best performance displayed here is our selected approach using ALLHiC. (e) Minor scaffolding mistakes were curated with the Juicebox Assembly Tools (JBAT) based on the Hi-C linkage map. The curated result aligned astoundingly well with the *A. lyrata* ssp. *lyrata* scaffolds. (f) Gap closure with PBJelly did not change the scaffold arrangement. The previously reported misassembly on *A. lyrata* ssp. *lyrata* scaffold can be seen in the top left corner. In addition, one more extensive inversion is displayed on scaffold 7, in the center of the plot.

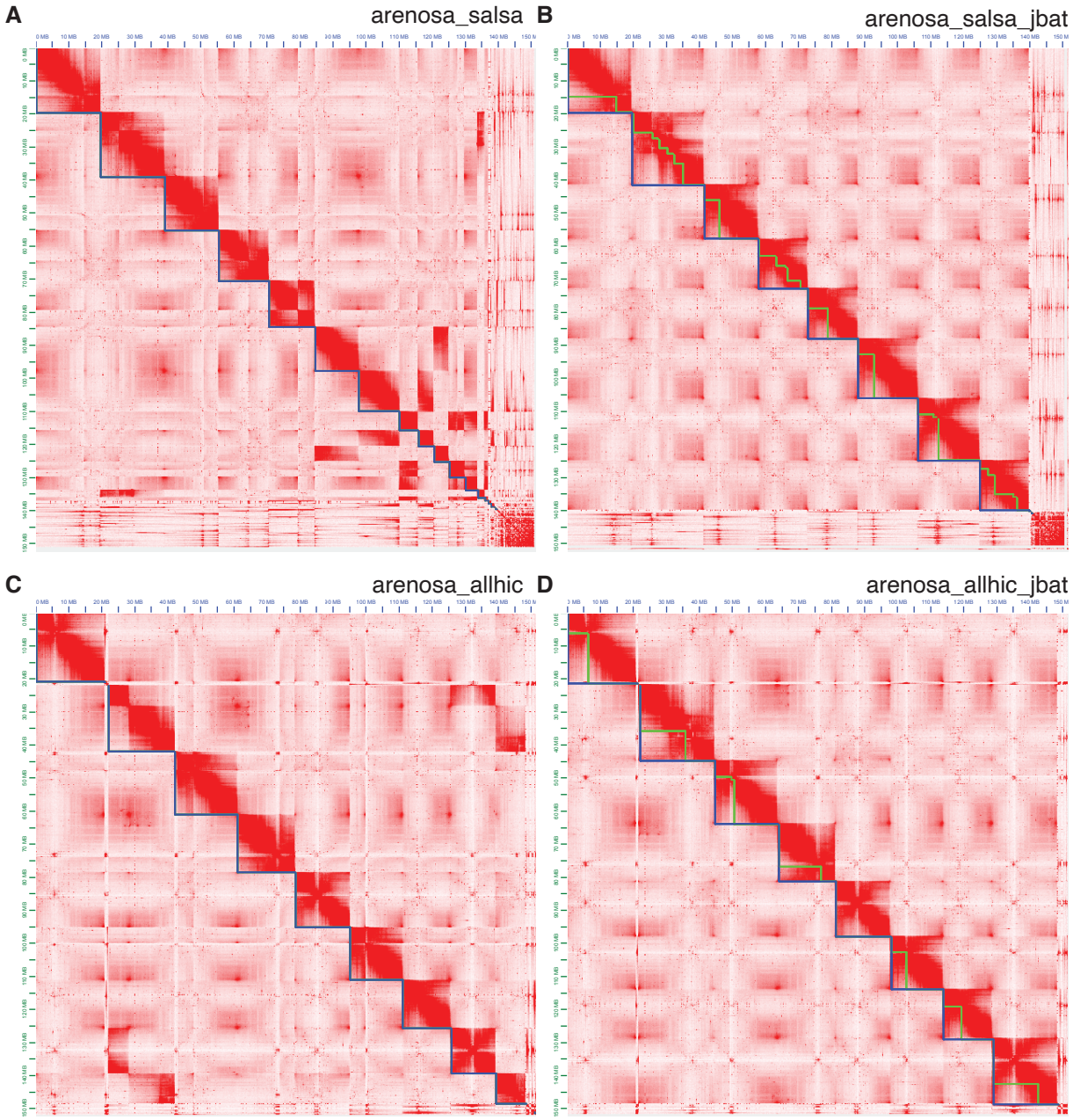


Figure S6. *A. arenosa* Pusté Pole Hi-C contact maps. Hi-C contact heat maps indicate the number of contacts between any given pair of loci in the assembly (red scale). Scaffolds are indicated by a blue line, and the green line documents the manual separation used for rearrangements. (a) Hi-C assembly heat map for *A. arenosa* Pusté Pole produced by the SALSA scaffolder. Strong signals far away from the diagonal were used for further scaffolding improvements. (b) The resulting contact heat map of the JBAT curated SALSA scaffolds. (c) Hi-C map of the *A. arenosa* Pusté Pole assembly scaffolded by the ALLHiC software. Problematic sequences are indicated by the low number of connections and stronger far-away signals. (d) Contact matrix after JBAT curation of miss joints and inversions. Green triangles indicate scaffold brakes. This assembly was selected and finalized by gap-filling with PBJelly (Figures S5e,f).

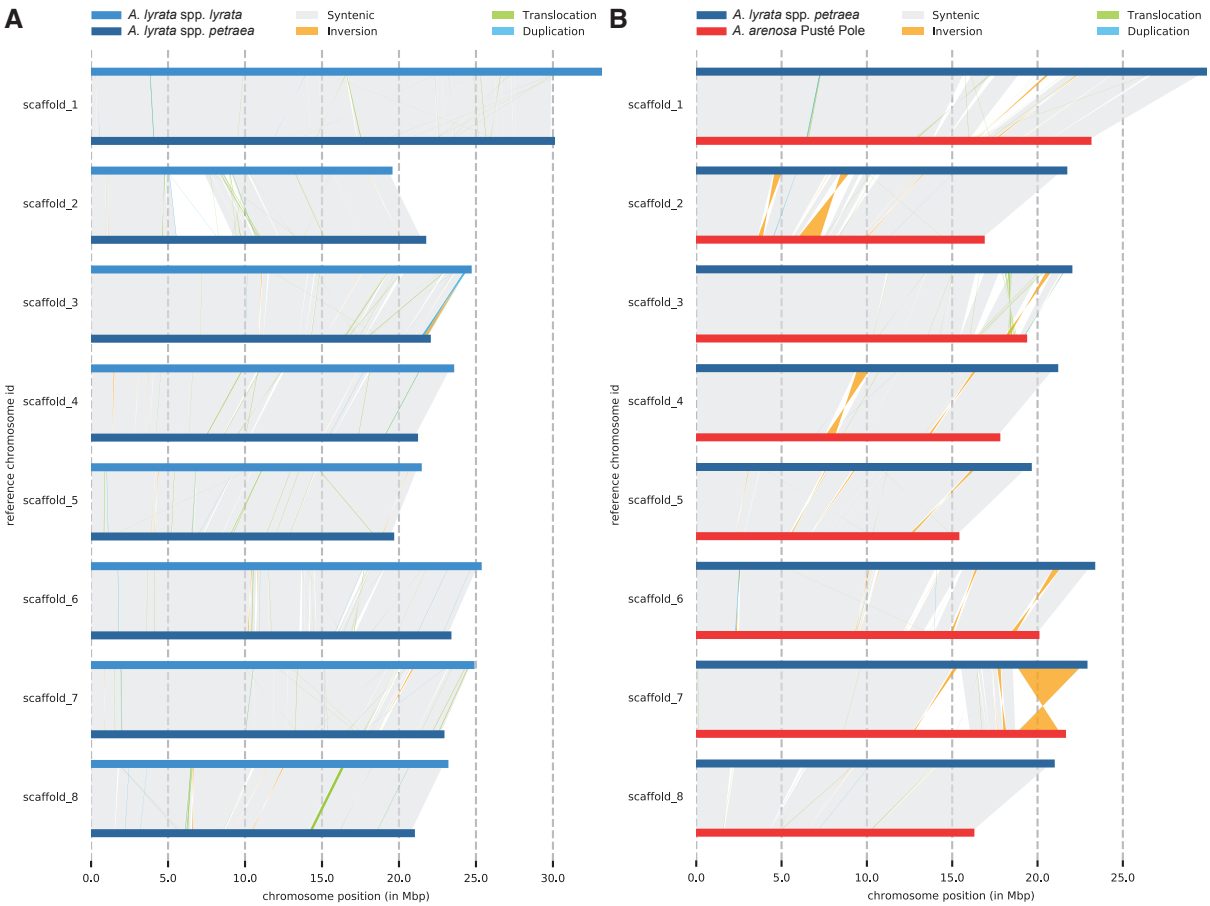
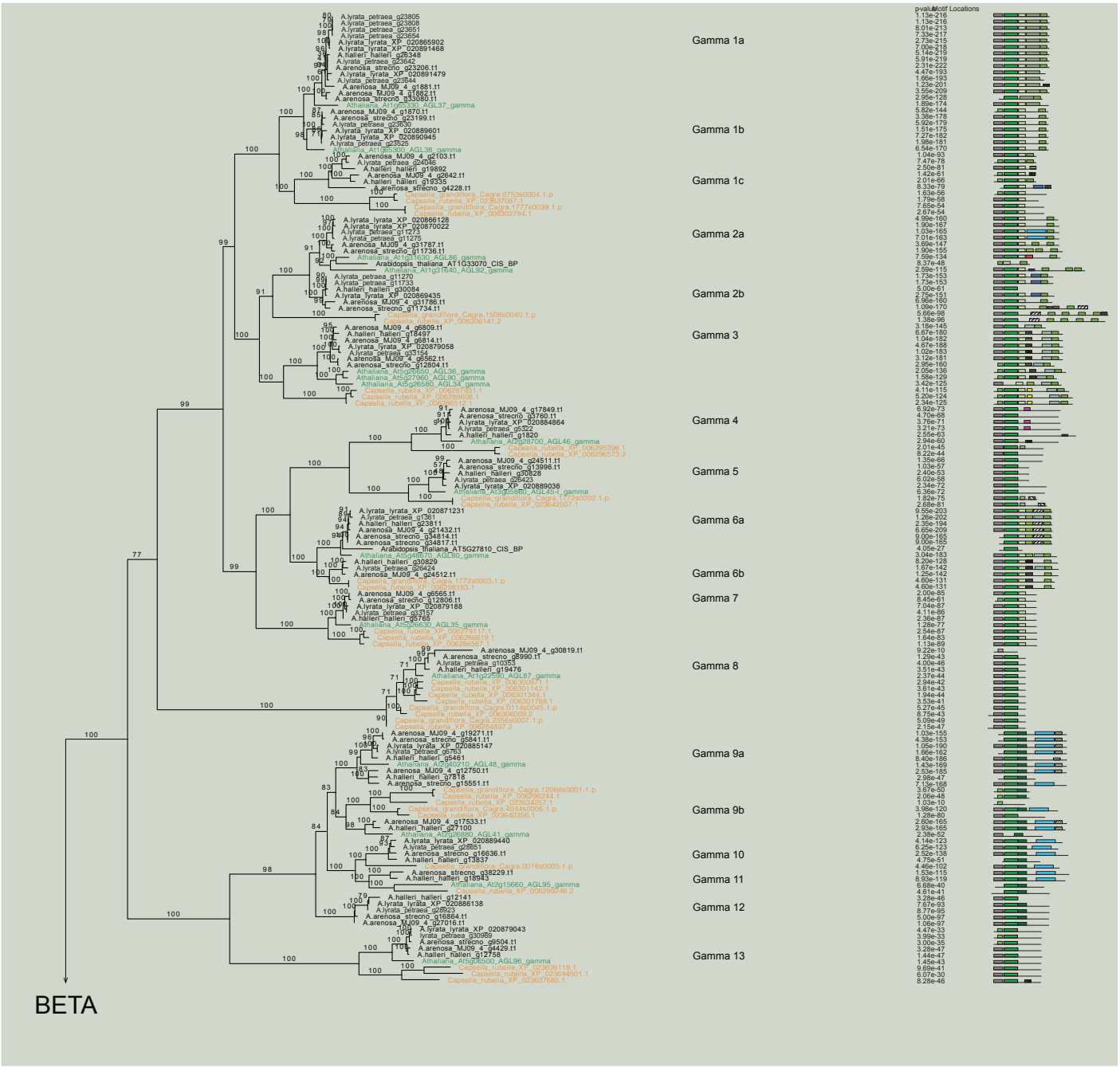


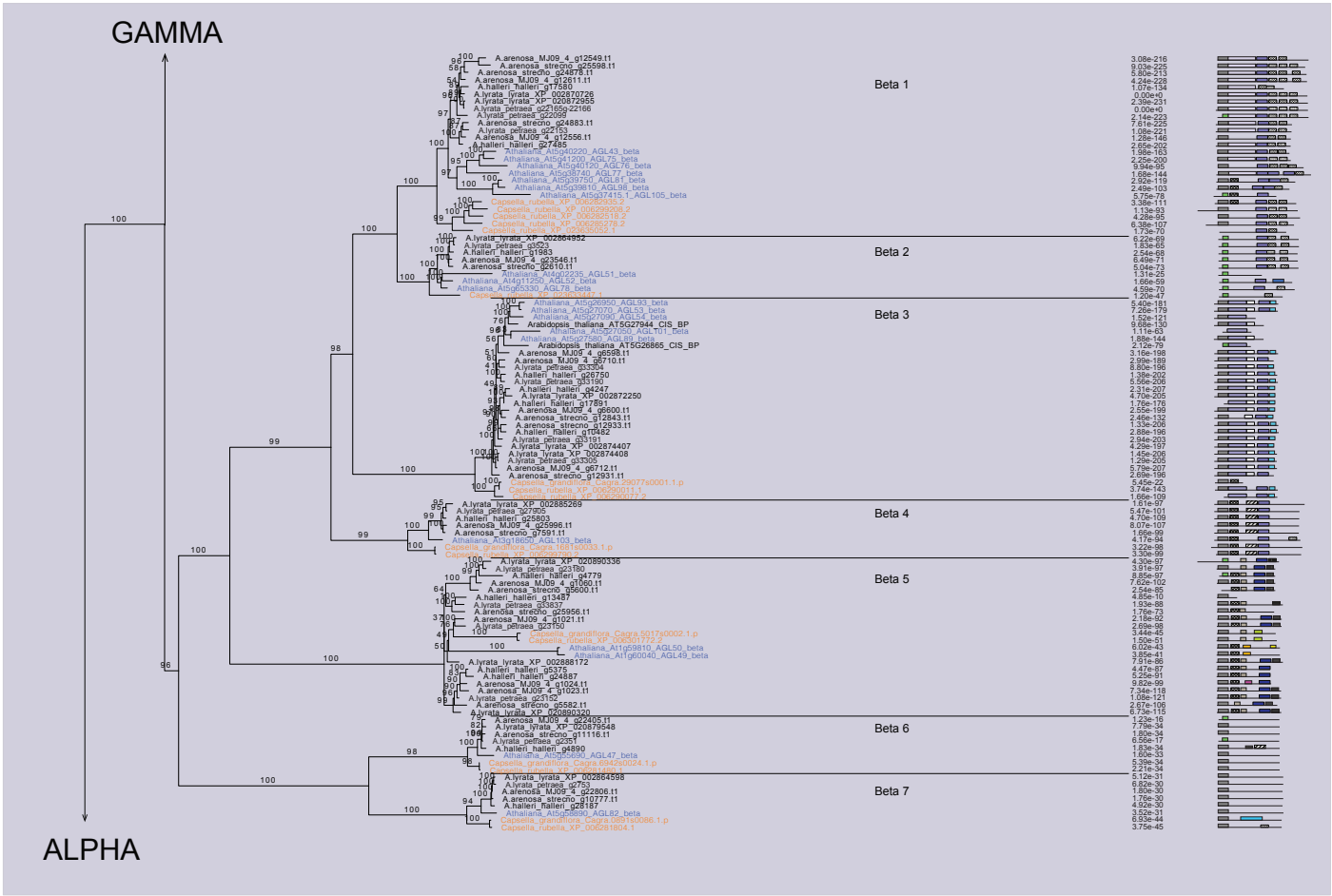
Figure S7. Genome alignments indicate genomic rearrangements. SyRI (Synteny and Rearrangement Identifier) was used to detect and classify structural differences between genomes. (a) Comparison of our *A. lyrata* ssp. *petraea* assembly (dark blue) to the *A. lyrata* ssp. *lyrata* reference (light blue). Syntenic regions are shown as gray blocks, while structural differences are grouped in translocations (green), inversions (orange), and duplications (blue). For visualization purposes, the relocations between non-homolog scaffolds have been excluded (e.g., for translocation between scaffolds one and two, see Figure 1e). (b) Synteny alignment between *A. lyrata* ssp. *petraea* (dark blue) and *A. arenosa* Pusté Pole (red). Structural differences are indicated between largely syntenic scaffolds.

Figure S8: GAMMA



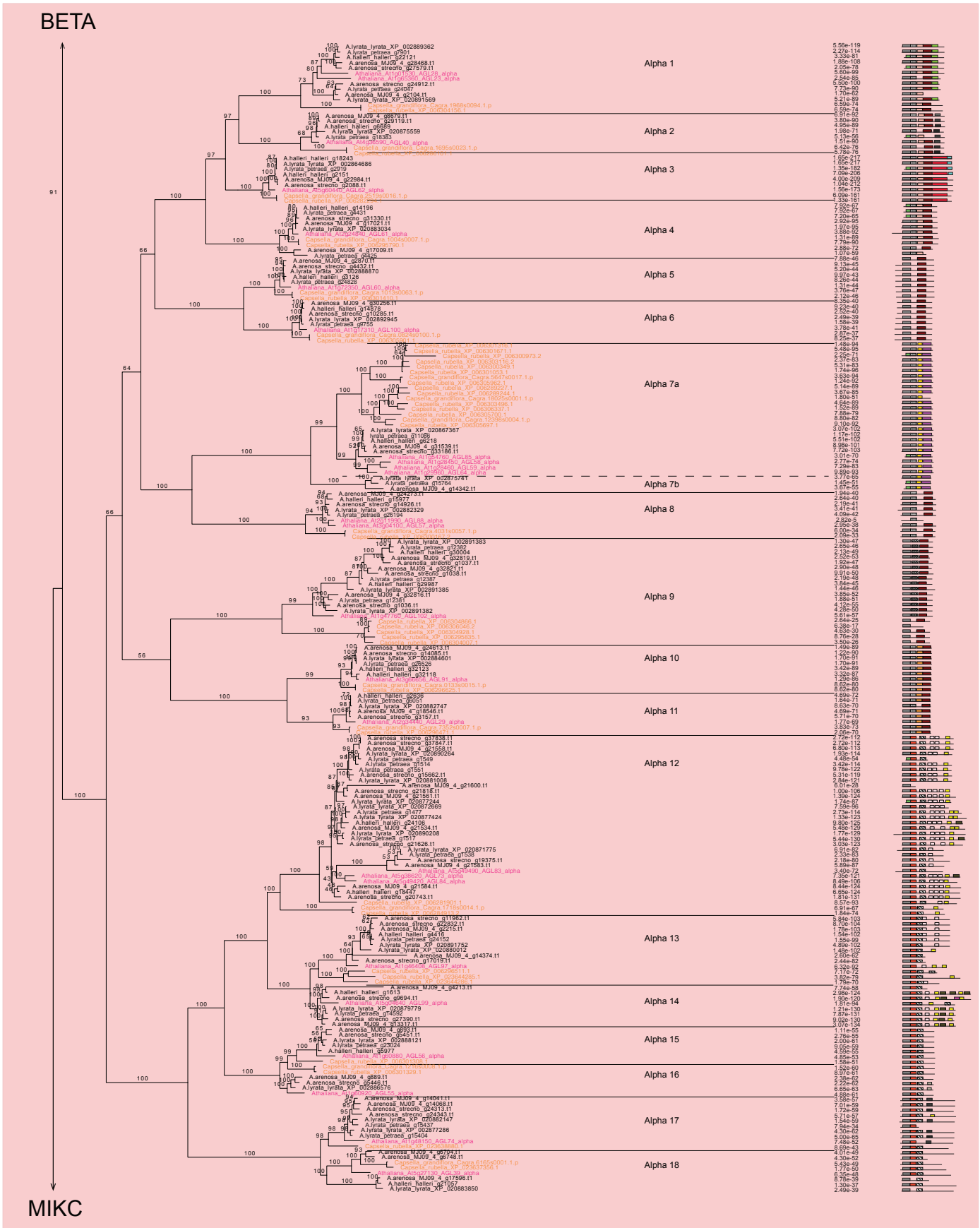
For high resolution view, see: https://github.com/PaulGrini/Arabidopsis_assemblies/tree/main/03_analysis/05_MADS_phylogeny

Figure S8: BETA



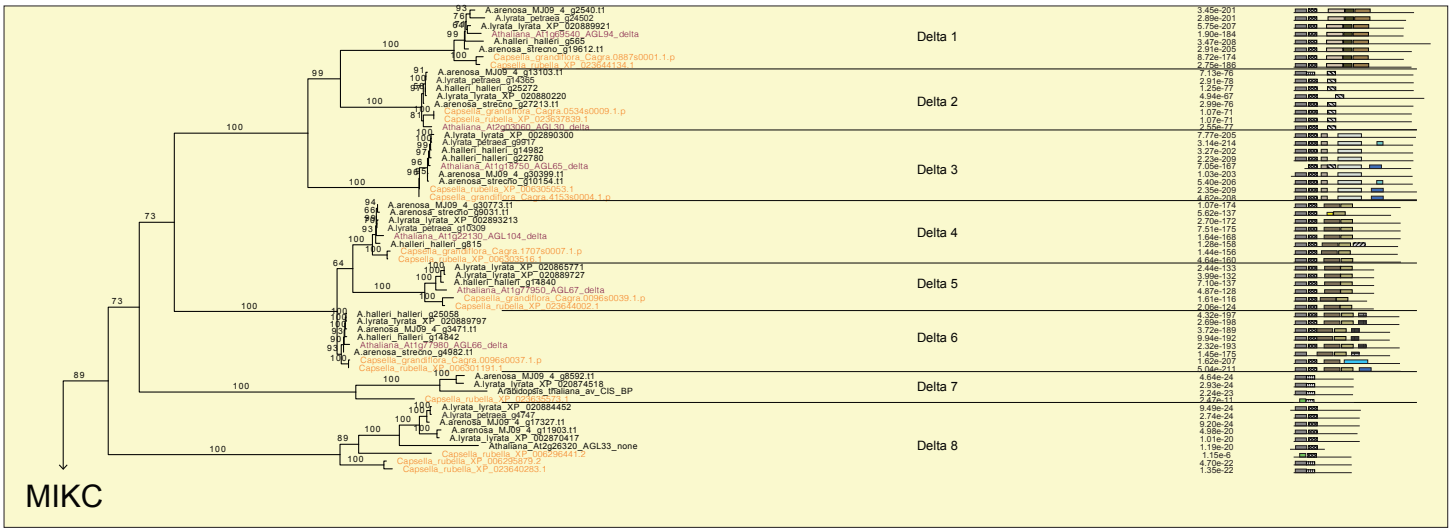
For high resolution view, see: https://github.com/PaulGrini/Arabidopsis_assemblies/tree/main/03_analysis/05_MADS_phylogeny

Figure S8: ALPHA



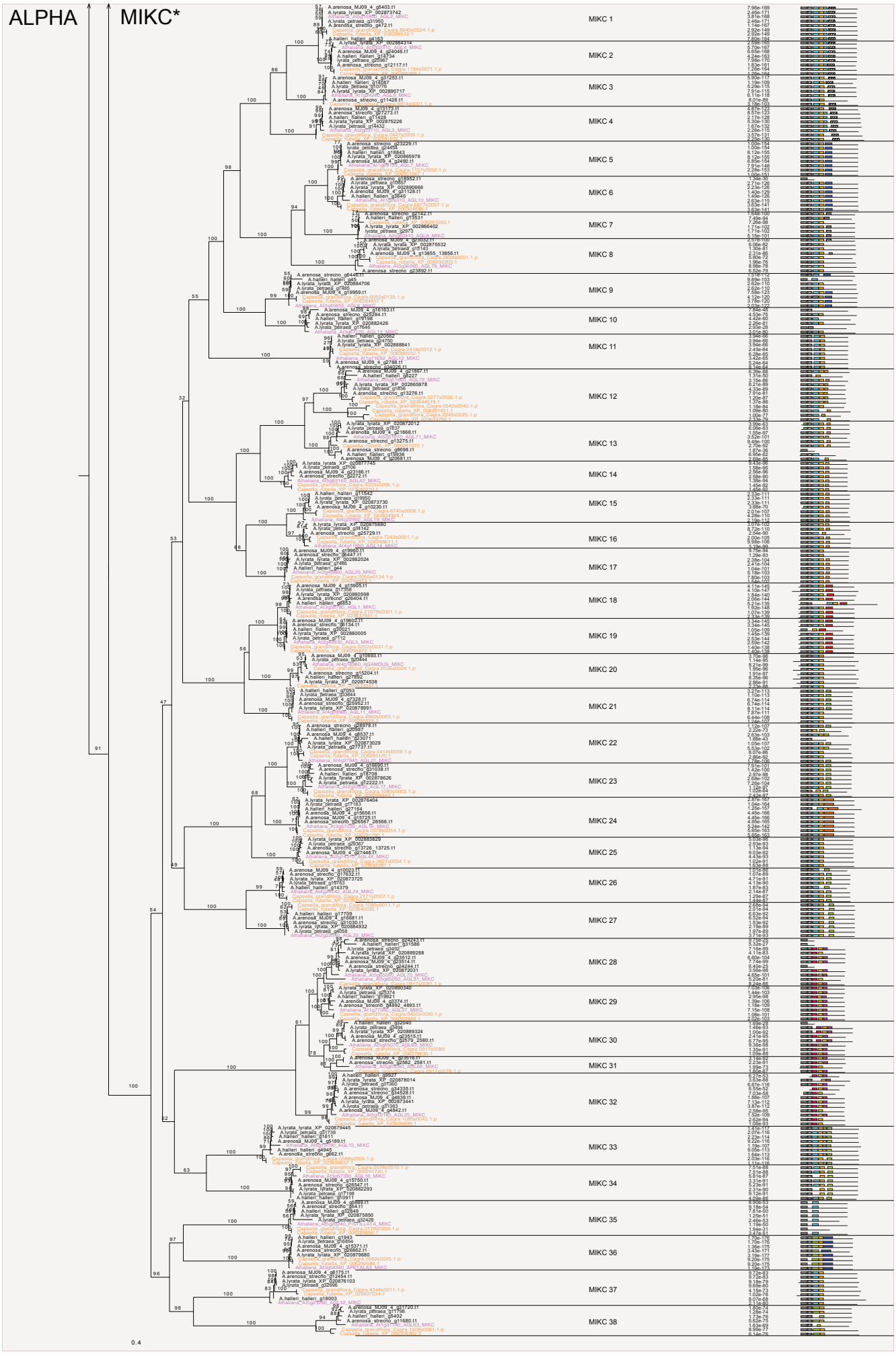
For high resolution view, see: https://github.com/PaulGrini/Arabidopsis_assemblies/tree/main/03_analysis/05_MADS_phylogeny

Figure S8: MIKC*



For high resolution view, see: https://github.com/PaulGrini/Arabidopsis_assemblies/tree/main/03_analysis/05_MADS_phylogeny

Figure S8: MIKC



For high resolution view, see: https://github.com/PaulGrini/Arabidopsis_assemblies/tree/main/03_analysis/05_MADS_phylogeny

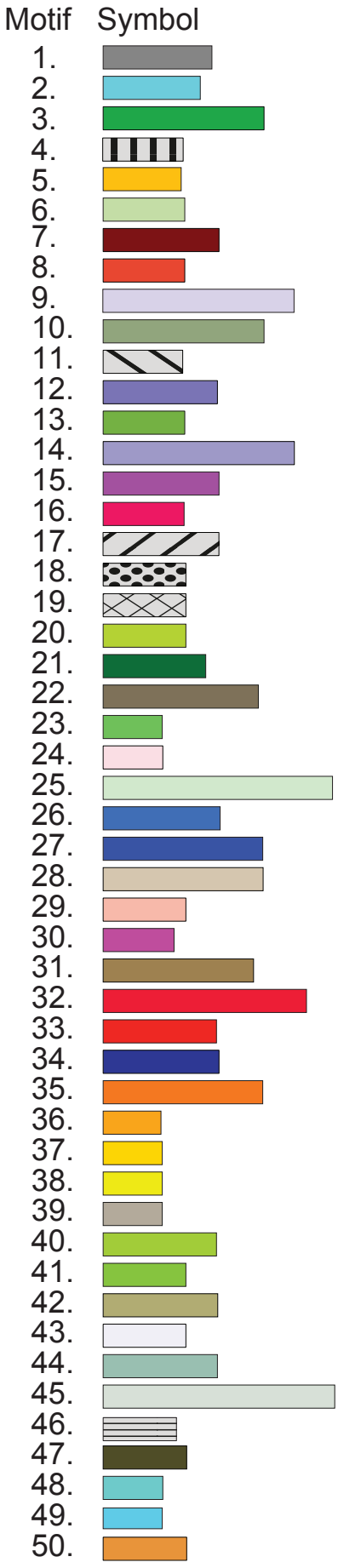


Figure S8. Extended MADS-box phylogeny. Phylogeny of all 841 identified MADS-box genes in *Arabidopsis* and *Capsella*. The groups correspond to previously published results with type I genes divided into three main groups (Mα, Mβ, Mγ), while type II genes fell into two monophyletic groups, MIKC and MIKC* [also referred to as Mδ] (Arora et al., 2007; Gramzow & Theißen, 2013; Henschel et al., 2002; Parenicová et al., 2003; Qiu & Köhler, 2022; Thangavel & Nayar, 2018). The groups are colored with a red box around M α, blue around Mβ green around Mγ, yellow around MIKC*, and gray around MIKC. Sequences of *A. thaliana* are marked in blue, and the outgroup *Capsella* in orange. Monophyletic groups of *Arabidopsis* MADS-box genes, sharing a common ancestor with genes from the outgroup *Capsella*, are numbered starting from the top of the tree and delineated with solid lines. Additional gene duplications that occurred in the common ancestor of *Arabidopsis*, but after *Arabidopsis* and *Capsella* separated, are delineated with a dashed line and marked with an additional letter. For instance, “Mγ-1a” and “Mγ-1b” indicate that these two clades share a last common ancestor with C. rubella (i.e. “Mγ-1”), but have a gene duplication specific for Arabidopsis that occurred after the separation of *Arabidopsis* and *Capsella*. The right-hand column contains the result of the MEME analysis for each sequence. For the full result of the MEME analysis consult the material on GitHub.

For high resolution view, see: https://github.com/PaulGrini/Arabidopsis_assemblies/tree/main/03_analysis/05_MADS_phylogeny

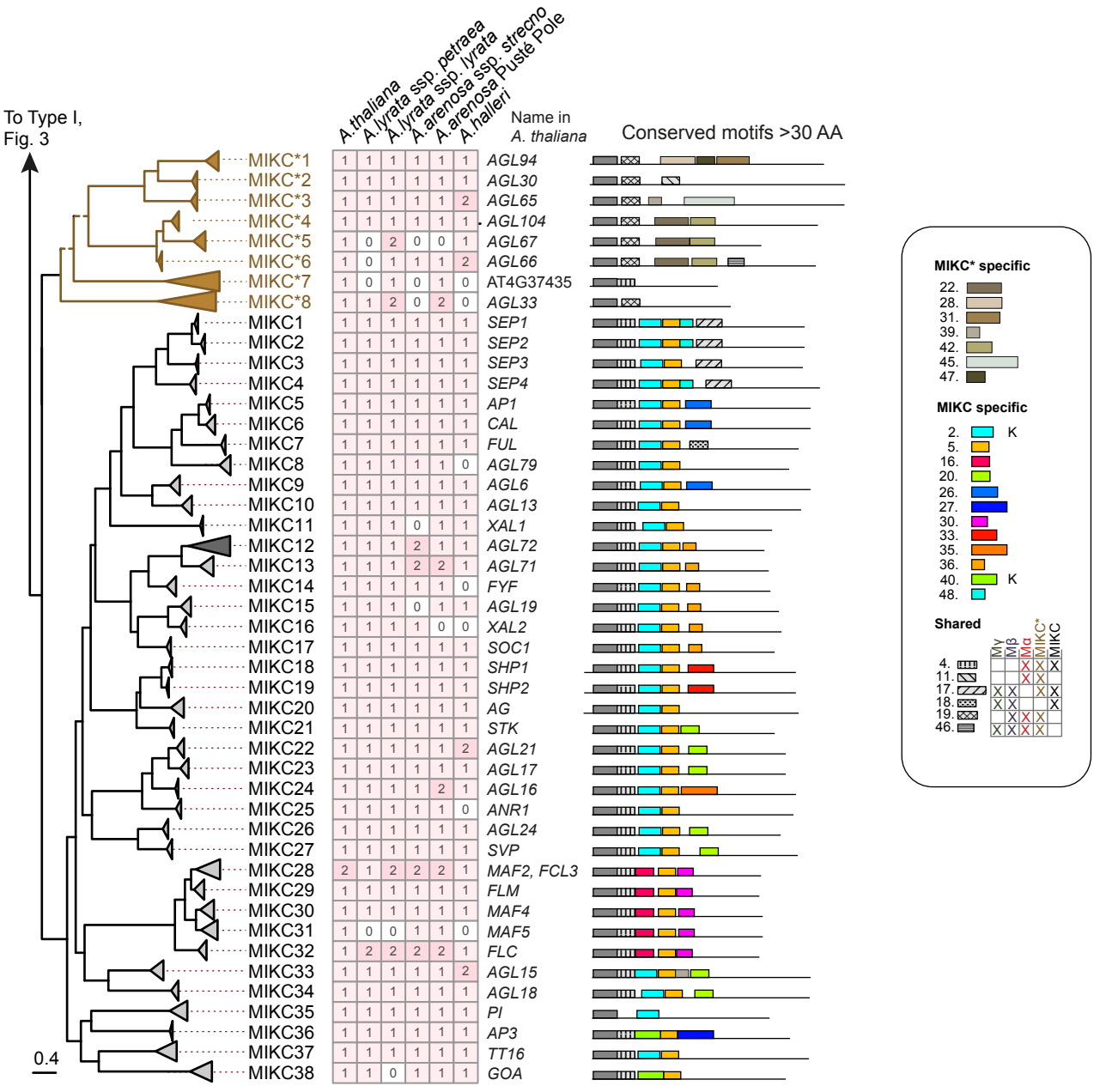


Figure S9. Phylogenetic analysis of MADS-box type II genes in *Arabidopsis*. The tree was derived by a maximum likelihood analysis of 275 identified MADS-box type I sequences from *A. thaliana*, *A. lyrata* ssp. *petraea*, *A. arenosa* Strecno, *A. arenosa* Pusté Pole, and *A. halleri* with two species of *Capsella* (89 sequences; not shown in the figure) used as outgroup. Solid branches represent bootstrap support >85%, while branches with support values <85% are dashed. The root of the tree is placed between type I and type II genes (the corresponding tree and heatmap for type I genes can be found in Figure 2). Triangles represent clades where branches are collapsed at the most recent gene duplication event in the last common ancestor of the genus *Arabidopsis*. The length of the triangles corresponds to the overall branch length of the collapsed clade, see the main text for naming schemes of clades. The heatmap shows the number of gene copies for each clade in the genomes of *A. thaliana*, *A. lyrata* ssp. *petraea*, *A. arenosa* Strecno, *A. arenosa* Pusté Pole, and *A. halleri*. The column next to the heatmap indicates the canonical AGL names of the genes in *A. thaliana* found in the corresponding clade; "none" means that a gene representing the clade is not found in the *A. thaliana* genome, while "new" indicates that the gene does not have a given AGL name. The last column shows a simplified representation of the MEME motifs. A fully expanded phylogenetic tree with individual tip labels, support values for all branches, and the outgroup *Capsella* can be found in Figure S8. Results from the full MEME analysis can be found in Figure S8.

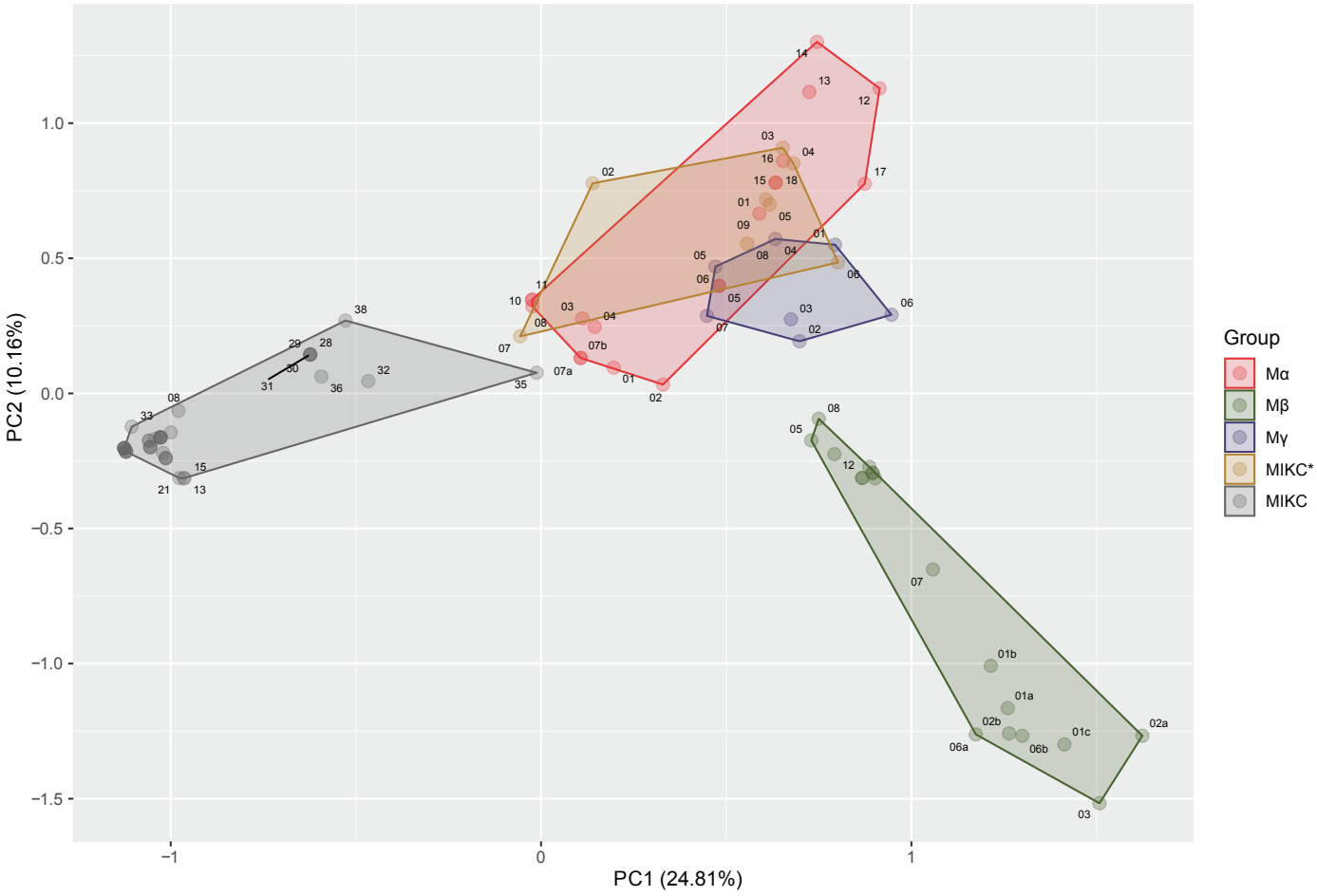


Figure S10. PCA of the distribution of MEME motifs on MADS-box type I and type II genes. The PCA was constructed from motifs identified by MEME and counted for each clade in the phylogeny. The first principal component axis captures 24.8% of the variation, and the second axis 10.2%. A polygon is drawn around each of the main groups with Mα in red, Mβ in blue, Mγ in green, MIKC* in yellow, and MIKC in gray.

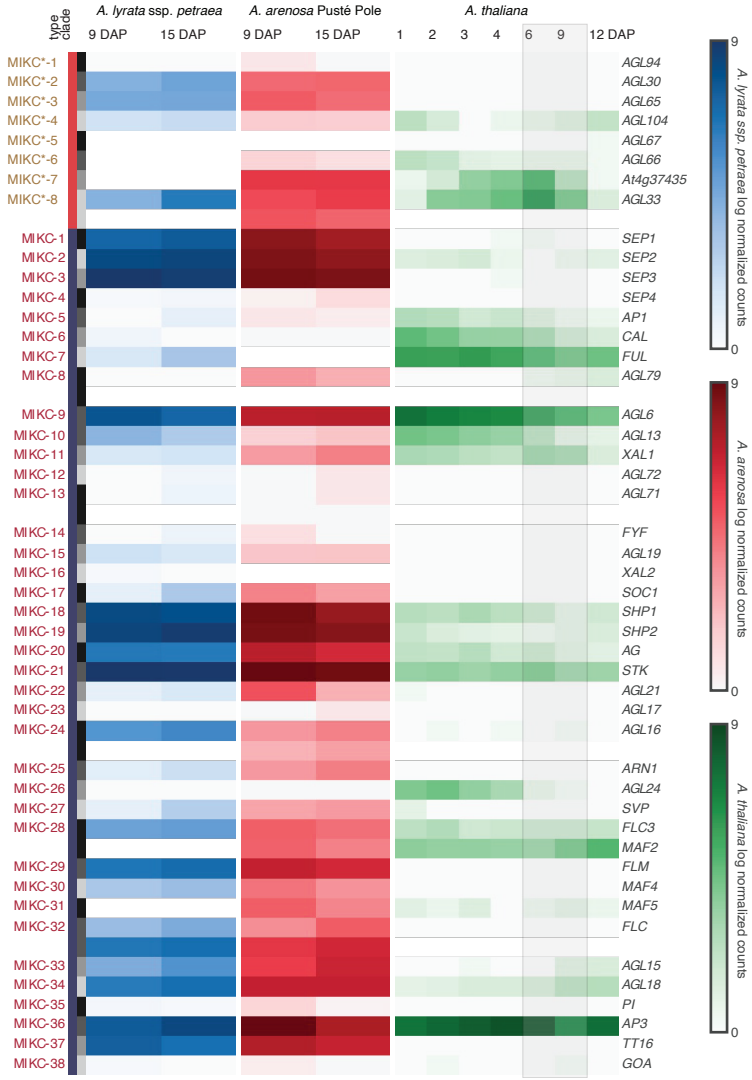


Figure S11. MADS-box type II expression during seed development. Gene expression profiles are displayed for all identified MADS-box type II genes and compared between *A. thaliana* (right column), *A. arenosa* Pusté Pole (middle column), and *A. lyrata* ssp. *petraea* (left column). The *A. thaliana* development time series (Bjerkan et al., 2020) serves as a reference. The endosperm cellularization in *A. thaliana* occurs between the 6 and 9 days after pollination (DAP). To adjust for the relatively slower development in *A. arenosa* Pusté Pole and *A. lyrata* ssp. *petraea*, corresponding stages before and after endosperm cellularization were sampled at 9 and 15 DAP, respectively. Ortholog genes are grouped and ordered and follow our MADS-box phylogeny (Figure 2). Sample normalized counts are shown with a base-2 logarithmic scale. The MADS-box type II seed expressions remain rather constant compared to the strong decline of type I expression around endosperm cellularization (Figure 4). However, expression differences can be seen between the *Arabidopsis* species. The *SEPALLATA* (*MIKC-1/2/3*) genes are strongly expressed in *A. lyrata* ssp. *petraea* and *A. arenosa* seeds while only weakly or absent in *A. thaliana*. In addition to their largely functional redundant role in flower development and ovary formation (Kaufmann et al., 2009; Pelaz et al., 2001), a crucial role in fruit development and ripening has been reported for *SEPALLATA* orthologs in tomato, strawberry, and apple (Ampomah-Dwamena et al., 2002; Schaffer et al., 2013; Seymour et al., 2011). Furthermore, the orthologs of *MIKC-15* (*AGL19*), *MIKC-24* (*AGL16*), *MIKC-26* (*AGL24*) as well as the FLC clade *MIKC-28* to *MIKC-32* show strong expression differences reflecting the different regulations of flowering and strengthen of vernalization between the perennial *A. lyrata* ssp. *petraea* and *A. arenosa* plants compared to the annual *A. thaliana* (Alexandre & Hennig, 2008; Hu et al., 2014; Kemi et al., 2013; Müller-Xing et al., 2022; Schönrock et al., 2006; Soppe et al., 2021). We detect differing expressions of *MIKC*s* between the species. *MIKC*-2* and *MIKC*-3* (*AGL30* and *AGL65*) are expressed in *A. lyrata* ssp. *petraea* and *A. arenosa* seeds but are not detected in *A. thaliana*. On the other hand, *MIKC*-6* and *MIKC*-7* (*AT4G37435* and *AGL33*) show expression in *A. arenosa* and *A. thaliana* but not in *A. lyrata* ssp. *petraea* seeds. *MIKC*s* are known for their transcription activity in pollen (Verelst et al., 2007) but have also been detected in the endosperm (Zhang et al., 2018). In addition, a high level of redundant heterodimers has been reported and double mutants reduce pollen fertility (Adamczyk & Fernandez, 2009).

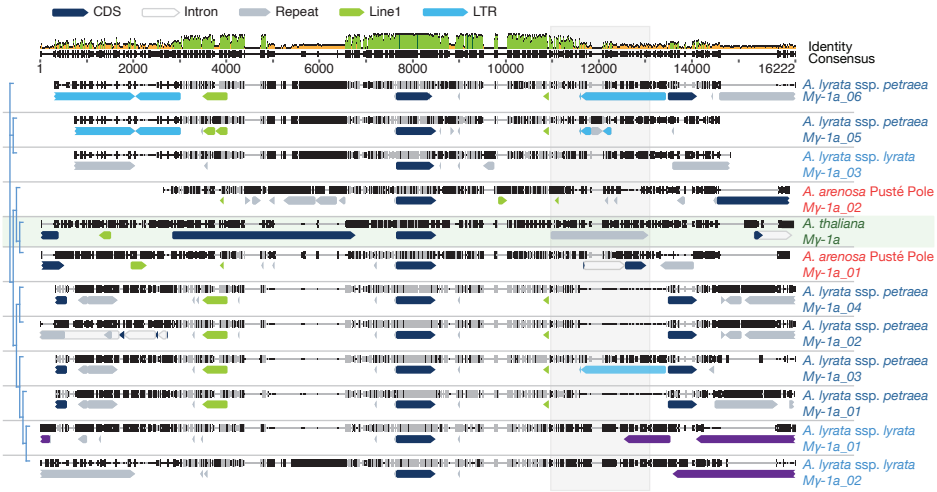


Figure S12. MAFFT alignments of genomic sequences containing and surrounding Mγ-1. Protein coding sequences (CDS), which are displayed as dark blue arrows of PHE1 orthologs from *A. thaliana* (green), *A. arenosa* (red), *A. lyrata* ssp. *lyrata* (light blue) and *A. lyrata* ssp. *petraea* (dark blue), are placed in the center. *A. thaliana* is highlighted by a light green background. On top, the consensus with identity score indicates a high similarity of the 3 prime regions downstream of the *Mγ-1* loci. Around 2200 bp 3 prime of *PHE1* lies the repeat-rich region crucial for its parental-specific expression (inside the AT1TE79790 RC/Helitron, highlighted by a transparent gray box). Similar locations of Line 1 (light green arrows) and LTR transposons (light blue arrows) can be found next to the *Mγ-1* loci between species. Further repeats are marked as light gray arrows and lncRNA as violet arrows (XR_002328948.1 and XR_002334149.1).

II

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 August 2023
DOI 10.3389/fpls.2023.1229060

OPEN ACCESS

EDITED BY
Dongfang Wang,
Spelman College, United States

REVIEWED BY
Jugou Liao,
Yunnan University, China
Filipe Borges,
INRA Centre Versailles-Grignon, France

*CORRESPONDENCE
Paul E. Grini
paul.grini@ibv.uio.no

†These authors have contributed equally to this work

RECEIVED 26 May 2023
ACCEPTED 12 July 2023
PUBLISHED 02 August 2023

CITATION
Bjerkan KN, Alling RM, Myking IV, Brysting AK and Grini PE (2023) Genetic and environmental manipulation of *Arabidopsis* hybridization barriers uncovers antagonistic functions in endosperm cellularization. *Front. Plant Sci.* 14:1229060. doi: 10.3389/fpls.2023.1229060

Genetic and environmental manipulation of *Arabidopsis* hybridization barriers uncovers antagonistic functions in endosperm cellularization

Katrine N. Bjerkan[1,2†], Renate M. Alling[1,2†], Ida V. Myking[1,2], Anne K. Brysting[1,2] and Paul E. Grini[1*]

[1]Section for Genetics and Evolutionary Biology (EVOGENE), Department of Biosciences, University of Oslo, Oslo, Norway, [2]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

Speciation involves reproductive isolation, which can occur by hybridization barriers acting in the endosperm of the developing seed. The nuclear endosperm is a nutrient sink, accumulating sugars from surrounding tissues, and undergoes coordinated cellularization, switching to serve as a nutrient source for the developing embryo. Tight regulation of cellularization is therefore vital for seed and embryonic development. Here we show that hybrid seeds from crosses between *Arabidopsis thaliana* as maternal contributor and *A. arenosa* or *A. lyrata* as pollen donors result in an endosperm based post-zygotic hybridization barrier that gives rise to a reduced seed germination rate. Hybrid seeds display opposite endosperm cellularization phenotypes, with late cellularization in crosses with *A. arenosa* and early cellularization in crosses with *A. lyrata*. Stage specific endosperm reporters display temporally ectopic expression in developing hybrid endosperm, in accordance with the early and late cellularization phenotypes, confirming a disturbance of the source-sink endosperm phase change. We demonstrate that the hybrid barrier is under the influence of abiotic factors, and show that a temperature gradient leads to diametrically opposed cellularization phenotype responses in hybrid endosperm with *A. arenosa* or *A. lyrata* as pollen donors. Furthermore, different *A. thaliana* accession genotypes also enhance or diminish seed viability in the two hybrid cross-types, emphasizing that both genetic and environmental cues control the hybridization barrier. We have identified an *A. thaliana* MADS-BOX type I family single locus that is required for diametrically opposed cellularization phenotype responses in hybrid endosperm. Loss of AGAMOUS-LIKE 35 significantly affects the germination rate of hybrid seeds in opposite directions when transmitted through the *A. thaliana* endosperm, and is suggested to be a locus that promotes cellularization as part of an endosperm based mechanism involved in post-zygotic hybrid barriers. The role of temperature in hybrid speciation and the identification of distinct loci in control of hybrid failure have great potential to aid the introduction of advantageous traits in breeding research and to support models to predict hybrid admixture in a changing global climate.

KEYWORDS

***Arabidopsis thaliana*, *A. arenosa*, *A. lyrata*, hybrid barrier, temperature, endosperm, seed development**

1 Introduction

Speciation is usually a continuous process towards increasing divergence and reproductive isolation between two lineages. Reproductive isolation can be obtained due to hybridization barriers, which act before fertilization (pre-zygotic) or after fertilization (post-zygotic) (Rieseberg and Willis, 2007; Widmer et al., 2009). A special case of post-zygotic hybridization barriers acts in the developing seed (Lafon-Placette and Köhler, 2016), resulting in developmental abnormality in the endosperm (Cooper and Brink, 1942). Seed death is accepted to be mainly due to failure of endosperm development since the embryo can be rescued in culture after microdissection (Sharma, 1999). Many studies have indicated endosperm deficiency to be the major cause of hybrid seed inviability (Brink and Cooper, 1947; Sukno et al., 1999; Dinu et al., 2005; Roy et al., 2011) and an endosperm-based hybridization barrier has been shown to be conserved across distinct species groups such as *Arabidopsis* (Lafon-Placette and Köhler, 2016), *Capsella* (Rebernig et al., 2015; Dziasek et al., 2021), rice (Ishikawa et al., 2011; Sekine et al., 2013; Zhang et al., 2016; Tonosaki et al., 2018; Wang et al., 2018), tomato (Florez-Rueda et al., 2016; Roth et al., 2019), monkeyflower (Oneal et al., 2016; Flores-Vergara et al., 2020; Kinser et al., 2021), and potato (Johnston and Hanneman, 1982; Cornejo et al., 2012). This suggests that the phenomenon represents a major mechanism of reproductive isolation in plants, however it is largely ignored in modern literature of speciation (Lafon-Placette and Köhler, 2016).

The endosperm is a triploid tissue that requires tight control of genome dosage (2:1 maternal:paternal ratio). Cellularization of the endosperm marks a transition in seed development, as up to this point, the endosperm functions as a nutrient sink. At this developmental time point, the endosperm concurrently switches from nutrient sink to source for the developing embryo (Lafon-Placette and Köhler, 2016), and manipulating the timing of endosperm cellularization through interploidy crosses arrests embryo development (Scott et al., 1998; Hehenberger et al., 2012). Similarly, hybridization between plant species was shown to result in embryo arrest due to endosperm cellularization failure (Haig and Westoby, 1988; Haig and Westoby, 1991; Comai et al., 2000; Bushell et al., 2003). Manipulating the ploidy of parents in interspecies crosses has also shown to improve the success of hybridization, demonstrating a requirement for genome balance in the endosperm (Comai et al., 2000; Bushell et al., 2003; Lafon-Placette et al., 2017).

The genus *Arabidopsis* has been widely used for studying evolutionary questions (Koenig and Weigel, 2015; Koch, 2019) including the effects of interspecific hybridization in controlled crossings (Chen et al., 1998; Comai et al., 2000; Nasrallah et al., 2000; Josefsson et al., 2006; Walia et al., 2009; Burkart-Waco et al., 2012; Burkart-Waco et al., 2013; Burkart-Waco et al., 2015; Bjerkan et al., 2020). When diploid *A. thaliana* is crossed to diploid *A. arenosa*, the endosperm shows late cellularization and high degree of seed abortion (Josefsson et al., 2006; Bjerkan et al., 2020). Furthermore, when *A. arenosa* is crossed as male to *A. lyrata*, the same phenotype can be seen with late endosperm cellularization and a very high seed lethality (Lafon-Placette et al., 2017). Interestingly, crossing *A. lyrata* as male to *A. arenosa* results in the opposite effect with early endosperm cellularization (Lafon-Placette et al., 2017).

The strength of the endosperm-based hybridization barrier can be influenced by accession specific genetic variation (Burkart-Waco et al., 2012; Burkart-Waco et al., 2013; Burkart-Waco et al., 2015; Bjerkan et al., 2020). In crosses between diploid *A. thaliana* to diploid *A. arenosa*, the choice of accessions in both species significantly acts to repress or enhance the endosperm barrier (Bjerkan et al., 2020; Burkart-Waco et al., 2012). The current knowledge on the effect of temperature in early seed development is limited (Paul et al., 2020), but an effect of temperature on hybrid seed development in reciprocal crosses of wheat and barley has previously been reported (Molnár-Láng and Sutka, 1994). The sensitivity of endosperm cellularization to heat stress during early endosperm development has been demonstrated in rice (Folsom et al., 2014) and type I MADS-box transcription factors (TFs) are deregulated during moderate heat stress (Chen et al., 2016). We have previously reported a temperature effect on endosperm based post-zygotic hybrid lethality of diploid species in the genus *Arabidopsis* (Bjerkan et al., 2020). This temperature effect was also shown using different accessions of both *A. arenosa* and *A. thaliana*, demonstrating a combinatorial effect of accessions and temperature (Bjerkan et al., 2020; Burkart-Waco et al., 2012).

Genomic imprinting is an epigenetic phenomenon, which infers parent-of-origin allele specific expression of maternally or paternally inherited alleles (Hornslien et al., 2019; Batista and Kohler, 2020). As proper endosperm development depends on a correct ratio of parental genomes, it is suggested that differences in genomic imprinting programs may be responsible for the evolution of sexual incompatibility in crosses between divergent individuals (Haig and Westoby, 1988; Haig and Westoby, 1991; Bushell et al., 2003; Schatlowski and Kohler, 2012). Alternatively, or additionally, epigenetic remodeling upon hybridization due to combination of diverged maternal and paternal siRNAs may lead to comprehensive failure of genomic imprinting and ectopic expression of transposons and imprinted genes (Martienssen, 2010; Ng et al., 2012). Recent evidence supports this emerging role of imprinted genes, and in *Arabidopsis* interspecies hybrids, paternally expressed genes (PEGs) shifted to be maternally expressed genes (MEGs) (Josefsson et al., 2006). Importantly, both PEGs and MEGs have been shown to erect hybridization barriers, and mutational loss of these genes has been reported to bypass hybridization barriers in interspecies crosses (Walia et al., 2009; Wolff et al., 2015). A major part of the MEGs and PEGs encodes proteins that activate pathways in the endosperm consistent with the prominent role of cellularization in seed survival.

Proper endosperm development in *Arabidopsis* is reliant on the FERTILIZATION INDEPENDENT SEED-Polycomb Repressive Complex 2 (FIS-PRC2) (Grossniklaus et al., 1998; Kiyosue et al., 1999; Luo et al., 1999; Köhler et al., 2003). FIS-PRC2 is important in endosperm development indirectly through the genes it regulates, which include several type I MADS-box TFs (Zhang et al., 2018). Deregulation of type I MADS-box TFs in interspecies crosses has been postulated to induce the endosperm-based hybridization barrier, but unfortunately most of these TFs have no clear function because of extensive genetic redundancy (De Bodt et al.,

2003; Parřenicová et al., 2003; Walia et al., 2009; Bemer et al., 2010). One exception is AGAMOUS-LIKE 62 (AGL62), which is found to suppress cellularization of the endosperm in *Arabidopsis* (Kang et al., 2008). The clear function of AGL62 is further emphasized through its interaction with the FIS-PRC2 complex (Hehenberger et al., 2012), with mutation of FIE, MEA, FIS2 and MSI1 resulting in an ectopic proliferation of nuclear endosperm (Grossniklaus et al., 1998; Köhler et al., 2003; Guitton et al., 2004), whereas the *agl62* mutant results in precocious cellularization (Kang et al., 2008). AGL62 mutation has also been shown to alleviate the hybridization barrier in the *A. thaliana* × *A. arenosa* cross, resulting in a higher germination rate (Walia et al., 2009; Bjerkan et al., 2020).

Here we report that hybrid seeds from crosses between *A. thaliana* mothers with *A. arenosa* or *A. lyrata* pollen donors result in diametrically opposed endosperm phenotypes, both giving rise to reduced seed germination rate, albeit caused by late cellularization in crosses with *A. arenosa* and early cellularization in crosses with *A. lyrata*. We demonstrate that the hybrid barriers are under the influence of abiotic factors, and show that a temperature gradient leads to opposed cellularization phenotype and seed viability in hybrid endosperm with *A. arenosa* or *A. lyrata* as pollen donors. In addition, *A. thaliana* accession genotypes also influence seed viability in the two hybrid cross-types in opposite directions. Using stage specific endosperm reporters, we demonstrate that the source-sink endosperm phase change is delayed or precocious in seeds of the two hybrids. Our data suggests an *A. thaliana* type I MADS-BOX family locus to act as a promoter of endosperm cellularization, affecting the germination rates of *A. arenosa* or *A. lyrata* hybrid seeds in opposite directions.

2 Materials and methods

2.1 Plant material and growth conditions

A. thaliana accessions (Col-0, C24, Ws-2 and Wa-1) and mutant lines were obtained from the Nottingham Arabidopsis Stock Center (NASC). The *A. arenosa* population MJ09-4 originates from Nízke Tatry Mts.; Pusté Pole (N 48.8855, E 20.2408) and the *A. lyrata* subsp. *petraea* population MJ09-11 originates from lower Austria; street from Pernitz to Pottenstein (N 47.9190, E 15.9755) (Jørgensen et al., 2011; Lafon-Placette et al., 2017; Bjerkan et al., 2020). Mutant lines *agl35-1* (SALK_033801), *agl40-1* (SALK_107011) and *mea-9* (SAIL_724_E07) were in Col-0 accession background (Shirzadi et al., 2011; Kirkbride et al., 2019; Bjerkan et al., 2020). The *agl35-1* T-DNA line was genotyped using primers AAACCAAAGTTTTGCCACTAAGAC, ATTTTTCAGTCAAGATTACCCACC and GCGTGGACCGCTT GCTGCAACTCTCTCAGG. Marker lines *proAT5G09370>>H2A-GFP* (EE-GFP) and proAT4G00220>>H2A-GFP (TE1-GFP) were in Col-0 accession background (van Ekelenburg et al., 2023).

Surface-sterilization of seeds was performed by treatment with 70% ethanol, bleach (20% Chlorine, 0.1% Tween20) and wash solution (0.001% Tween20) for 5 min each step (Lindsey et al., 2017). Sterilized seeds were transferred to petri dishes containing 0.5 MS growth-medium with 2% sucrose (Murashige and Skoog, 1962) and stratified at 4°C for 2 days (*A. thaliana*) or 10 days (*A. lyrata*, *A. arenosa* and hybrids) before germination at 22°C with a 16h/8h light/dark cycle. After two weeks, seedlings were transferred to soil and cultivated at 18°C (16h/8h light/dark cycle, 160 µmol/m^2/sec, 60-65% humidity). *A. arenosa* strain MJ09-4 was previously demonstrated to be diploid (Bjerkan et al., 2020). *A. lyrata* and *A. thaliana* × *A. lyrata* individual hybrid plants were confirmed diploid by flow cytometry (Supplementary Datasheet S1).

2.2 Crosses, temperature- and germination assays

A. thaliana plants were emasculated 2 days before pollination. Crossed plants were placed at experimental growth temperature until silique maturity/harvesting. For each cross combination, 4-8 different individual plants were used as pollinators and 15-85 siliques (biological replicates) were harvested individually (Supplementary Datasheet S2). After short-term storage at 4°C seeds from induvidual siliques were surface-sterilized ON using chlorine gas (Lindsey et al., 2017). All seeds from the harvested individual siliques were counted and planted on individual 0.5 MS growth-medium containing petri-dishes and scored for germination by counting protrusions through the seed coat after 10 days at 22°C growth conditions. On day 20, germinated seedlings were checked for *A. thaliana* accidental self-pollination (formation of floral shoots without vernalization). These rare events occurred at a frequency less than 1% and self-plants were removed from the analyses if present.

2.3 Statistics

R-studio (version 2023.03.1 + 446) (R Core Team, 2023) was used for data analyses. Plots were generated using the *ggplot2* (Wickham, 2016) and *dplyr* (Wickham et al., 2023) packages. For statistical analyses the *car* package (Fox and Weisberg, 2019) and *ggpubr* package (Kassambara, 2023) were employed. To assess the homogeneity of variance for the germination assays we conducted Levene's test (Levene, 1960). If the null hypothesis was rejected, the Welch's t-test (Welch, 1947) was used for statistical analyses. In all other cases statistical analyses were performed using Wilcoxon rank sum test (Mann and Whitney, 1947).

2.4 Microscopy

Feulgen stained seeds were harvested at 6 days after pollination (DAP), stained using Schiff's reagent (Sigma-Aldrich S5133), fixed and embedded in LR White (London Resin) (Braselton et al., 1996). Imaging was performed using an Andor DragonFly spinning disc confocal microscope with a Zyla4.2 sCMOS 2048x2048 camera attachment and excitation 488 nm/emission 500 to 600 nm. Seeds were scored for embryo and endosperm developmental stage and the number of endosperm nuclei. Endosperm nuclei counts were assigned an endosperm division value (EDV), which estimates the number of divisions to reach the corresponding number of

endosperm nuclei (Ungru et al., 2008). Mean EDV was calculated using the formula: 2^x = mean number of nuclei, where x is the EDV (x = LOG(mean number of nuclei)/LOG(2)).

Crosses with EE-GFP and TE1-GFP markers were imaged using the Andor DragonFly as described. Whole-mount imaging of seeds was performed using an Axioplan2 imaging microscope after 24 h/4°C incubation in 8:2:1 (w/v/v) chloral hydrate:water:glycerol (Grini et al., 2002). Mature dry seeds were imaged on a 1.5 x 2.5 cm grid using a Leica Z16apoA microscope connected to a Nikon D90 camera. Seed size and circularity were measured by converting images to black and white and then using the ImageJ "Threshold" and "Analyze particles" functions (https://imagej.nih.gov/ij/).

3 Results

3.1 Antagonistic effects on endosperm barriers in hybrid seeds from *A. lyrata* and *A. arenosa* crossed to *A. thaliana*

In order to compare the success of interspecific hybrids of *A. thaliana* crossed with *A. lyrata* or *A. arenosa*, we first observed the seed-set. Hybrid crosses with *A. lyrata* fathers had significantly reduced seed set per silique compared to crosses with *A. arenosa* fathers or *A. thaliana* (Col-0) self crosses (Supplementary Figures S1A, B). The lower seed-set correlates with failure of pollen tube burst after entering the female gametophyte (Supplementary Figure S1C), a pre-zygotic barrier previously described (Escobar-Restrepo et al., 2007). Interestingly, pollen tube burst failure can also be observed in crosses between *A. thaliana* and *A. arenosa*, but to a lower degree, correlating with the observed seed-set frequency (Supplementary Figures S1A, C). Differences in seed set between the two hybrid crosses thus have a gametophytic pre-zygotic base, and therefore do not affect the postzygotic hybridization barrier in these crosses.

A. thaliana self seeds and *A. thaliana* × *A. arenosa* or *A. thaliana* × *A. lyrata* hybrid seeds were comparable in size at 6 DAP (Figure 1A), but most *A. thaliana* × *A. lyrata* hybrid embryos were at an earlier developmental stage (Figure 1B). Large variation in embryonic stages including developmental arrest was observed in 12 DAP whole-mount chloral hydrate cleared hybrid seeds (Supplementary Figure S2A). A time series of Feulgen stained *A. thaliana* × *A. lyrata* hybrid seeds further identified variation in endosperm cellularization, starting at 3 DAP and resulting in developmental arrest at the globular embryo stage (Supplementary Figure S2B). The frequency of *A. thaliana* × *A. lyrata* early cellularization was higher than for *A. thaliana* self, and in strong contrast to the late cellularization in hybrid seeds from *A. thaliana* crossed with *A. arenosa* [Figure 1B; (Josefsson et al., 2006; Bjerkan et al., 2020)]. In order to compare the observed cellularization phenotype with nuclear proliferation in the syncytial endosperm, we investigated the number of nuclei in hybrid endosperm. The endosperm nuclei-number was significantly lower in *A. thaliana* × *A. lyrata* compared to *A. thaliana* × *A. arenosa* hybrid seeds, suggesting a reduction in endosperm proliferation rate (Figure 1C). Nonetheless, germination rates of hybrid seeds from crosses with *A. lyrata* were significantly higher than crosses with *A. arenosa* (Figure 1D; Supplementary Datasheet S2), indicating that the endosperm hybrid barrier is influenced in a diametrically opposed manner, both in terms of the endosperm cellularization phenotype and the effective output measured as the ability of hybrid seeds to germinate.

3.2 Endosperm phase change and cellularization in hybrids are not synchronized with embryo development

Compared to *A. thaliana*, viable seeds resulting from crosses between *A. thaliana* mothers and *A. arenosa* or *A. lyrata* fathers appear to develop along a slower embryo and endosperm developmental path. This suggests that as long as endosperm cellularization and embryo development are synchronized, seed viability is not impacted. In *A. lyrata* × *A. arenosa* hybrid seeds with unsynchronized endosperm and embryo development, the embryo can be rescued *in vitro*, indicating that the barrier is caused by lack of nutrient support to the growing embryo (Lafon-Placette and Köhler, 2016).

To investigate synchronization of embryo and endosperm in *A. thaliana* (Col-0) hybrid seeds with *A. arenosa* or *A. lyrata* fathers, we used genetic markers of endosperm development that are expressed before and after endosperm cellularization in *A. thaliana* (van Ekelenburg et al., 2023). In *A. thaliana* the EE-GFP genetic marker is expressed after fertilization and up to endosperm cellularization at 6-7 DAP (van Ekelenburg et al., 2023) and not expressed after cellularization (Figures 2, S3). In *A. thaliana* × *A. arenosa* hybrid seeds, expression of the EE-GFP marker was observed for the full duration of the time series (15 DAP) and no visible downregulation could be observed (Figures 2, S3). This expression pattern supports the observation that *A. thaliana* × *A. arenosa* hybrid seeds fail to initiate or have delayed endosperm cellularization (Figure 1) as described previously (Josefsson et al., 2006; Bjerkan et al., 2020). In contrast, when crossed to *A. lyrata*, expression of the EE-GFP marker decreased from 5 DAP and only a low frequency of seeds expressed the marker at cellularization around 9 DAP (Figures 2, S3). Taking the developmental delay in the *A. thaliana* × *A. lyrata* hybrid into account, the EE-GFP marker is prematurely terminated (Figure 2B), in accordance with the early cellularization phenotype (Figure 1). In *A. thaliana* self seeds, downregulation of the EE-GFP marker coincides with endosperm cellularization (van Ekelenburg et al., 2023). In hybrid seeds from *A. arenosa* or *A. lyrata* fathers, continued expression or premature EE-GFP downregulation, respectively, (Figure 2B) coincides with diametrically opposed cellularization phenotypes both indicating that the endosperm phase change and cellularization is not synchronized with embryo development, leading to embryo and seed failure.

In *A. thaliana*, the TE1-GFP genetic marker is expressed after cellularization at 6-7 DAP (van Ekelenburg et al., 2023) and until seed maturation at 17-19 DAP (Figures 3, S4). Delayed activation of marker expression was observed at low frequency in the *A. thaliana* × *A. arenosa* hybrid seeds with expression from 9 to 18 DAP (Figures 3, S4). Interestingly, in *A. thaliana* × *A. lyrata* hybrid seeds the TE1-GFP marker expressed prematurely from before 6 DAP globular stage seeds lasting until 18 DAP (Figures 3, S4),

10.3389/fpls.2023.1229060

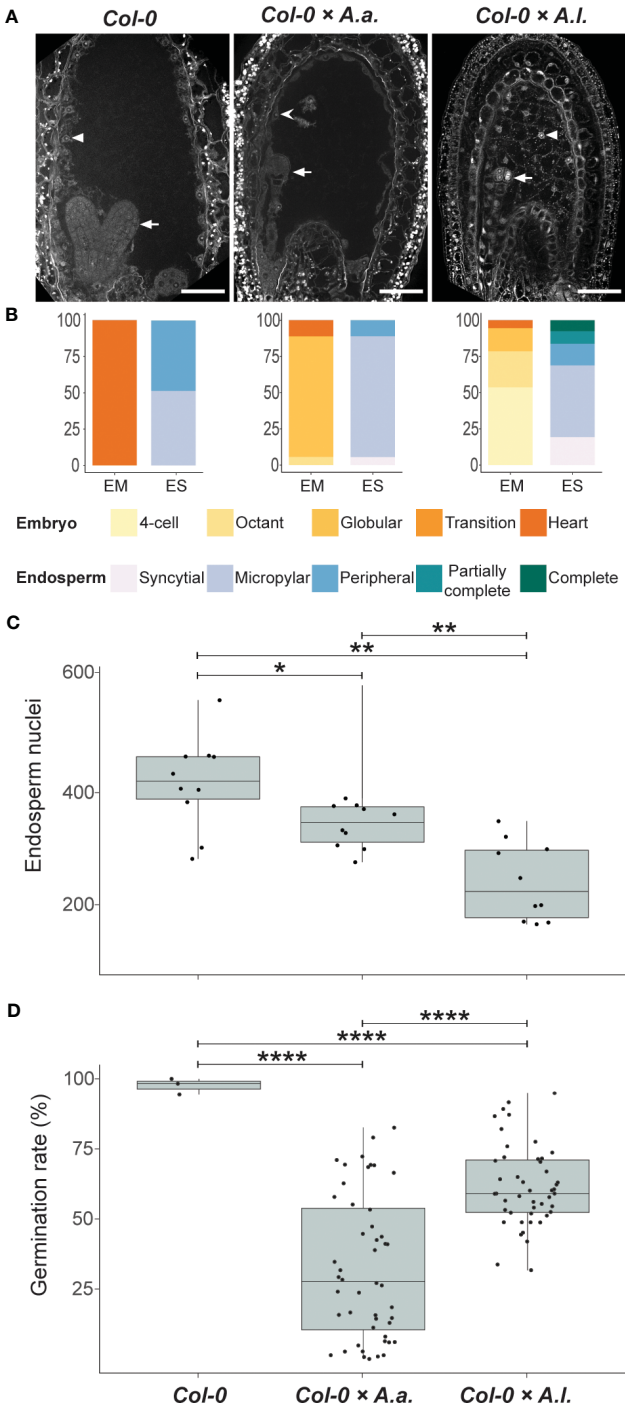


FIGURE 1
Antagonistic endosperm phenotypes in hybrid seeds. Seeds from crossing *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (Col-0) have antagonistic endosperm phenotypes and display opposing seed germination rates. **(A)** Confocal images of Feulgen-stained Col-0 and hybrid seeds 6 days after pollination (DAP) emphasizing endosperm cellularization. Open arrowhead points to syncytial endosperm nuclei, closed arrowheads point to cellularized endosperm nuclei and full arrows point to the embryo. In Col-0 self cellularization occurs at the 6 DAP embryo heart stage whereas in *A.a.* hybrid seeds the endosperm is mainly syncytial at 6 DAP. In *A.l.* hybrid seeds precocious endosperm cellularization is observed 6 DAP already at the early globular stage. All crosses are female × male. Scale bar = 50 µm. **(B)** Relative frequencies of embryo and endosperm stages in seeds from the same crosses as above. Col-0, n = 37; Col-0 × *A.a.*, n = 18; Col-0 × *A.l.*, n = 93; EM, embryo stages; ES, endosperm stages. **(C)** Number of endosperm nuclei in seeds from the same crosses as above, n = 10. Significance is indicated for the comparisons between all genotypes (Wilcoxon rank-sum test: *P ≤ 0.05; **P ≤ 0.01; ****P ≤ 0.0001). **(D)** Germination rates in seeds from the same crosses as above. Biological replicates (siliques): Col-0, n = 4; Col-0 × *A.a.*, n = 48; Col-0 × *A.l.*, n = 48. Significance is indicated for the comparisons between all genotypes (Welch's t-test: *$P \leq 0.05$; **$P \leq 0.01$; ****$P \leq 0.0001$).

10.3389/fpls.2023.1229060

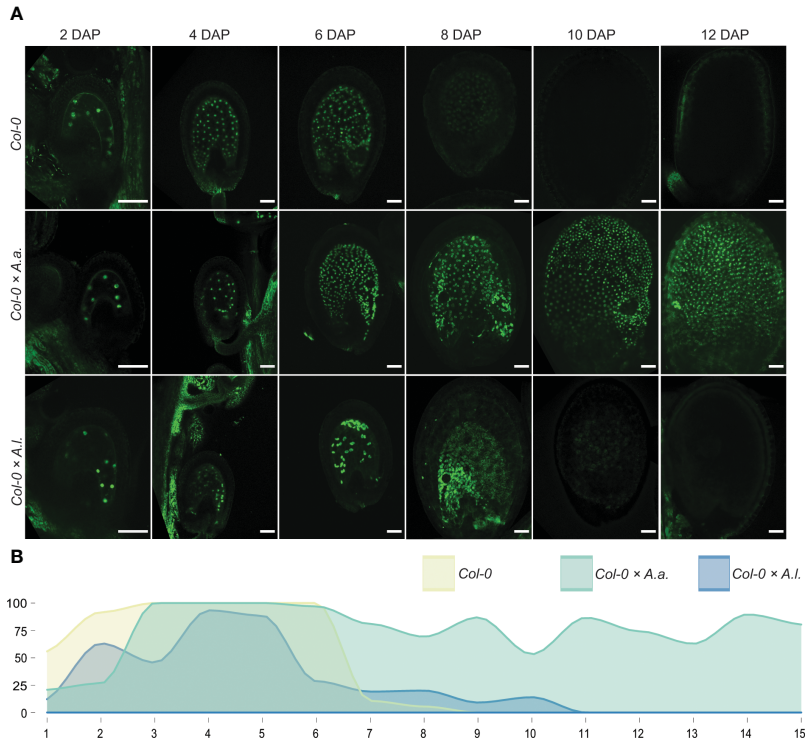


FIGURE 2
Early endosperm marker confirms aberrant cellularization timing in hybrid seeds. **(A)** Expression patterns of proAT5G09370>>H2A-GFP (EE-GFP) in seeds of *A. thaliana* (Col-0) and in hybrid seeds from crossing *A. arenosa* (*A.a.*)/*A. lyrata* (*A.l.*) as pollen donor to Col-0 at 2, 4, 6, 8, 10 and 12 days after pollination (DAP). Scale bar = 50 µm. **(B)** Percentage of hybrid seeds expressing EE-GFP at 1-15 DAP. *A. thaliana* (Col-0) self crosses and hybrid crosses are indicated with different colors.

indicating premature endosperm phase change initiation and supporting the observed precocious endosperm cellularization phenotype (Figure 1).

In *A. thaliana* seeds the decrease of EE-GFP expression and increase of TE1-GFP expression is strictly coordinated, with limited overlap. This is contrasted by the EE-GFP and TE1-GFP expression in seeds from hybrid crosses (Figure 4). Since the marker transgenes are expressed from the maternal *A. thaliana* genomes, the only difference in these crosses is the paternal contribution. For *A. thaliana* × *A. arenosa* an overlap in expression of the markers was observed from 9 DAP, caused by the prolonged expression of the EE-GFP marker (Figure 4). In *A. thaliana* × *A. lyrata* overlapping expression was observed from 6 to 10 DAP (Figure 4), showing a shift in TE1-GFP expression towards earlier developmental stages, although the *A. thaliana* × *A. lyrata* hybrid seeds develop slower compared to both *A. thaliana* selfed and *A. thaliana* × *A. arenosa* seeds (Figure 1).

3.3 Temperature alters the hybrid barrier strength in diametrically opposed directions

Lowering of temperature from 22°C to 18°C ameliorates the germination efficiency of the hybrid seeds from *A. thaliana* mothers crossed to *A. arenosa* fathers (Bjerkan et al., 2020). In order to investigate if the phenotypically contrasting hybrid barrier observed in seeds from *A. lyrata* fathers was affected by temperature in a corresponding way, we performed crosses between *A. thaliana* (Col-0) mothers and *A. lyrata* or *A. arenosa* fathers at 4°C temperature windows ranging from 14°C to 26°C. Interestingly, the germination rate of *A. thaliana* × *A. arenosa* hybrid seeds was significantly enhanced by progressive lowering of the temperature, and contrasted by the germination rate of *A. thaliana* × *A. lyrata* hybrid seeds that was significantly enhanced by progressive increase of the temperature (Figure 5; Supplementary Datasheet S2). In the temperature window, the effects on the hybrid barriers followed a close to linear, but opposed reaction norm. We applied more extreme temperatures previously reported to be within the normal, and not stress inducing, growth-range of *A. thaliana* (Lloyd et al., 2018), but a further enhancement could not be obtained (Supplementary Figure S5).

3.4 Accessions of *A. thaliana* influence *A. arenosa* and *A. lyrata* hybrid barriers antagonistically

Different accessions of *A. thaliana* affect the strength of the hybrid barrier when crossed to *A. arenosa* (Burkart-Waco et al., 2012; Bjerkan et al., 2020). Having demonstrated an antagonistic temperature effect on the hybrid barrier when *A. arenosa* or *A.*

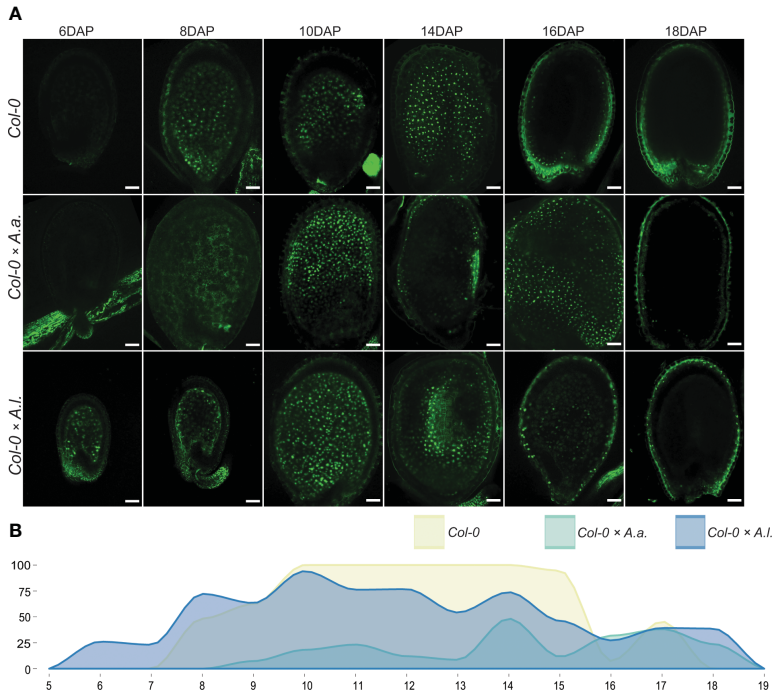


FIGURE 3
Total endosperm marker confirms aberrant cellularization timing in hybrid seeds. **(A)** Expression patterns of proAT4G00220>>H2A-GFP (TE1-GFP) in seeds of *A. thaliana* (Col-0) and in hybrid seeds from crossing *A. arenosa* (*A.a.*)/*A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (Col-0) at 6, 8, 10, 14, 16 and 18 days after pollination (DAP). Scale bar = 50 µm. **(B)** Percentage of hybrid seeds expressing TE1-GFP at 5-19 DAP. *A. thaliana* (Col-0) self crosses and hybrid crosses are indicated with different colors.

lyrata are crossed to *A. thaliana* mothers, we next investigated if different *A. thaliana* accessions influence the hybrid barrier in similar or opposite manner. We performed hybrid crosses with *A. arenosa* and *A. lyrata* fathers to the diploid *A. thaliana* accessions Col-0, C24 and Ws-2. Additionally, the tetraploid accession Wa-1 was used as a control, as tetraploid *A. thaliana* crossed to diploid *A. arenosa* has been shown to increase hybrid seed survival (Josefsson et al., 2006).

Compared to the Col-0 accession crosses, the C24 accession enhanced seed survival significantly when crossed to *A. arenosa*, contrasted by the *A. lyrata* hybrid where the germination rate declined highly significantly compared to the Col-0 cross (Figure 6; $p \leq 0.0001$). For Ws-2, the germination rate was severely reduced in the *A. arenosa* hybrid cross, contrasted by moderately high (though lower than for Col-0) germination rate in the *A. lyrata* hybrid cross (Figure 6, Supplementary Datasheet S2).

In the tetraploid *A. thaliana* Wa-1 to diploid *A. arenosa* hybrid cross, the seed germination rate was enhanced, as previously reported (Josefsson et al., 2006). In contrast, in the Wa-1 to *A. lyrata* hybrid cross, the germination rate was significantly decreased compared to the Col-0 cross (Figure 6). Notably, the effect on hybrid seed viability was higher with the diploid C24 accession than with the tetraploid Wa-1 accession.

Crosses performed in parallel at 18°C and 22°C demonstrated the same trends at both temperatures (Supplementary Figure S6, Supplementary Datasheet S2). Although large differences in the barrier strength was observed, as measured by germination and seed viability, no obvious correlation could be found between seed survival and seed size and circularity (Supplementary Figure S7).

3.5 *A. thaliana* accession effects are not readily explained by endosperm cellularization phenotype

To investigate if the endosperm phenotype reflects the influence of accessions on hybrid seed viability, we inspected Feulgen stained 6 DAP hybrid seeds by confocal microscopy. We scored the number of endosperm nuclei in *A. thaliana* accessions and accession hybrids with *A. arenosa* and *A. lyrata* at 18°C and 22°C. The endosperm division value [EDV; (Ungru et al., 2008)] was generally higher when *A. arenosa* was involved (Figure 7; Supplementary Datasheet S3). No obvious correlation between the number of endosperm nuclei and hybrid seed viability was found (Supplementary Figure S8), however a significant correlation between endosperm proliferation rates and growth temperature could be observed in all crosses except *A. thaliana* C24 x *A. lyrata*. The latter hybrid cross did indeed exhibit very low germination rates (Figure 6), however similar low germination rates were found in *A. thaliana* Ws-2 x *A. arenosa* hybrid seeds but here accompanied by a high endosperm proliferation rate (Supplementary Figure S8, Supplementary Datasheet S3).

Seed phenotypes were scored for defined stages of embryo and endosperm development. *A. thaliana* accession self crosses at 22°C displayed embryo stages in the late heart to walking stick stage, with

10.3389/fpls.2023.1229060

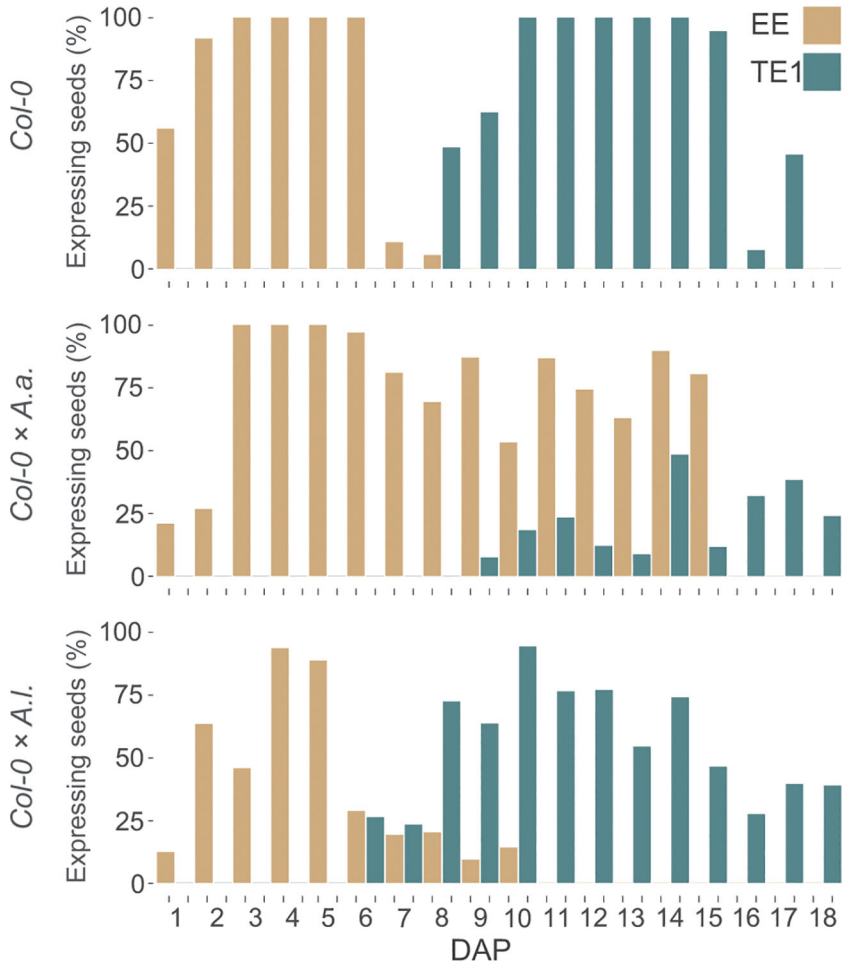


FIGURE 4
Overlapping expression of early and late endosperm markers in hybrid seeds. Percentage of seeds expressing proAT5G09370>>H2A-GFP (EE-GFP) and proAT4G00220>>H2A-GFP (TE1-GFP) at 1-18 days after pollination (DAP) in *A. thaliana* (Col-0) self-crosses and from crossing *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (Col-0). EE-GFP expression in the Col-0 x *A. arenosa* hybrid seeds were not documented after 15 DAP.

C24 exhibiting the fastest embryonic development. Endosperm cellularization was partly complete, though fully completed in most *A. thaliana* C24 self-seeds (Figure 7; Supplementary Datasheet S3). In *A. thaliana* accession x *A. arenosa* hybrid seeds, embryo development ranged from globular to transition stages. The endosperm was mainly syncytial or had initiated cellularization in the micropylar endosperm. In the C24 cross almost half of the seeds exhibited advanced cellularization stages and also complete cellularization. In this cross, higher germination rate was correlated with temporally correct timing of endosperm cellularization (Figures 6, 7).

In *A. thaliana* accession x *A. lyrata* hybrid seeds, embryonic stages ranged from globular to heart, where the C24 accession displayed a majority of heart stages, and Ws-2 a majority of globular stages. In the Col-0 and C24 accession hybrids, near uniform complete endosperm cellularization was observed, contrasted by early peripheral cellularization in Ws-2 (Figure 7). In the case of the latter hybrid cross, embryo development and endosperm cellularization appeared to be synchronized, leading to higher seed viability (Figure 6). However, large differences in seed viability between C24 (low) and Col-0 (high) hybrid crosses (Figure 6) were not reflected by the endosperm cellularization phenotype as both crosses had mostly fully cellularized endosperm and appeared to be in a similar embryonic stage (Figure 7). *A. thaliana* accession hybrid crosses at 18°C exhibited a similar pattern (Supplementary Figure S9). Major significant differences in seed viability between C24 and Ws-2 (low vs medium-high; Supplementary Figure S6) were not reflected by endosperm cellularization as both accession hybrids exhibited mostly micropylar endosperm (Supplementary Figure S9). We conclude that the effect of using different accessions in the hybrid crosses can not readily be explained by a direct effect on the endosperm cellularization phenotype alone and that a more complex interaction between different genotypes occur.

3.6 Mutation of the MADS-box transcription factor AGL35 influences *A. arenosa* and *A. lyrata* hybrid barriers antagonistically

Deregulation of type I MADS-box TFs has been correlated with endosperm-based hybridization barriers (Walia et al., 2009), and

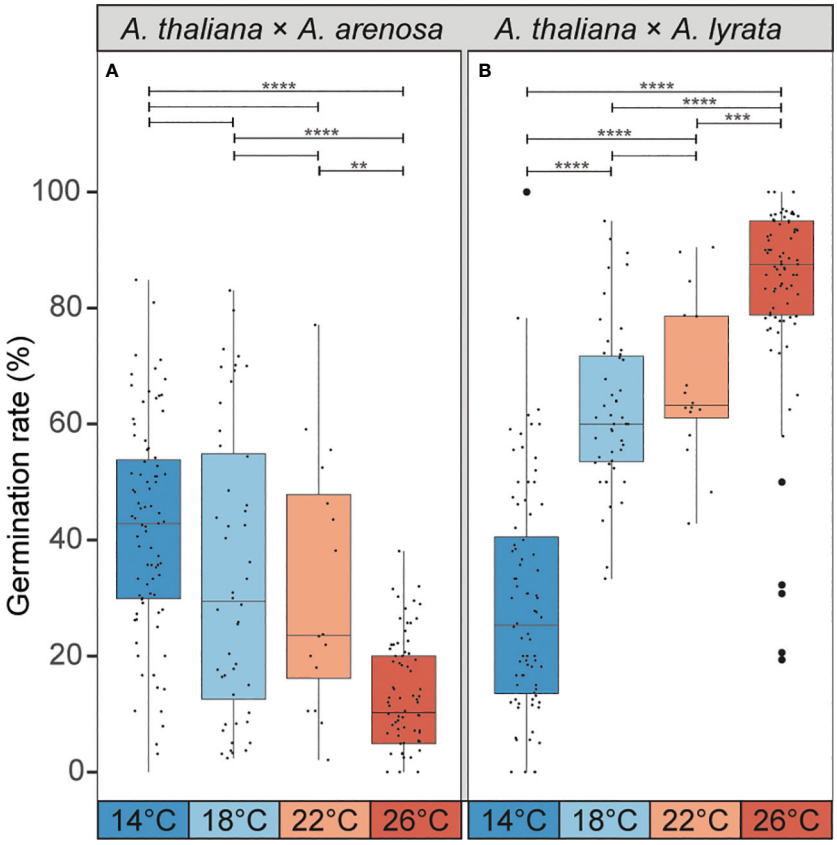


FIGURE 5
Temperature during seed development affects germination rate of hybrid seeds. **(A)** Germination rate of *A. thaliana* (Col-0) × *A. arenosa* hybrid seeds from crosses performed at 14°C, 18°C, 22°C and 26°C. Germination rate decreases with increasing temperature. Biological replicates (siliques): 14°C, n = 83 (3243 seeds); 18°C, n = 48 (2610 seeds); 22°C, n = 16 (796 seeds); 26°C, n = 78 (2713 seeds). **(B)** Germination rate of *A. thaliana* (Col-0) × *A. lyrata* hybrid seeds at 14°C, 18°C, 22°C and 26°C. Germination rate increases with increasing temperature. Biological replicates (siliques): 14°C, n = 84 (1676 seeds); 18°C, n = 47 (1540 seeds); 22°C, n = 16 (533 seeds); 26°C, n = 85 (2043 seeds). Box plot contains scattered data points representing germination rates observed per silique. Outliers are plotted as large data points. Significance is indicated for the comparisons between all temperatures (Welch's t-test: $^{**}P \leq 0.01$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$).

many of these TFs are epigenetically regulated by the so-called FIS-PRC2 and the histone methyltransferase MEDEA (MEA) (Zhang et al., 2018). Mutation of *MEA* results in ectopic proliferation of endosperm nuclei and delayed cellularization (Grossniklaus et al., 1998; Köhler et al., 2003; Guitton et al., 2004) and we therefore investigated if endosperm overproliferation in *A. thaliana mea* mutant mothers crossed to *A. arenosa* or *A. lyrata* enhances or alleviates the hybrid barriers, respectively.

Heterozygous self-crossed *A. thaliana mea* mutants resulted in a reduced germination rate of 60% meaning that 80% of seeds carrying the mutant maternal allele failed to germinate due to delayed endosperm cellularization (Supplementary Figure S10). Crossing *A. thaliana mea* to *A. arenosa* or *A. lyrata* resulted in a significant decrease in seed survival compared to Col-0 crosses (Supplementary Figure S10, Supplementary Datasheet S2). Reduced germination rate in both crosses corresponded to an additive effect of the reduced germination of the heterozygous *A. thaliana mea* mutation. Compared to expected values, single gene mutation of *mea* could not bypass the *A. thaliana* x *A. lyrata* species barrier, nor enhance the *A. thaliana* x *A. arenosa* barrier (Supplementary Figure S10, Supplementary Datasheet S2).

We studied the effect of single candidate genes regulated by FIS-PRC2 (Zhang et al., 2018). The mutant *agl35-1* in the Col-0 background was previously shown to strengthen the barrier when *A. thaliana* was crossed to *A. arenosa* (Bjerkan et al., 2020) and *AGL35* was upregulated in the same hybrid cross (Walia et al., 2009). *AGL40* is similarly expressed in the endosperm (Zhang et al., 2018), upregulated in hybrids (Walia et al., 2009) and mutant seeds have reduced seed size (Kirkbride et al., 2019). To investigate if mutation of single candidate genes could produce opposed effects on the hybrid barrier when crossing *A. thaliana* mothers to *A. lyrata* or *A. arenosa*, as observed when changing temperature or *A. thaliana* accession (Figures 5, 6), we performed *A. lyrata* and *A. arenosa* crosses to *A. thaliana agl35-1* and *agl40-1* and scored seed germination.

Interestingly, in crosses where *agl35-1* was crossed to *A. arenosa* or *A. lyrata* a highly significant decrease or increase in germination rate was observed, respectively, compared to wild type Col-0 crosses (Figure 8A). Single mutation of *AGL35* affected the hybrid barrier strength in diametrically opposed directions, as in crosses to *A. lyrata* the germination rate was significantly enhanced, in contrast to *A. thaliana* x *A. arenosa* crosses where the germination rate was significantly reduced (Figure 8A). Mutation of *AGL40* crossed to *A.*

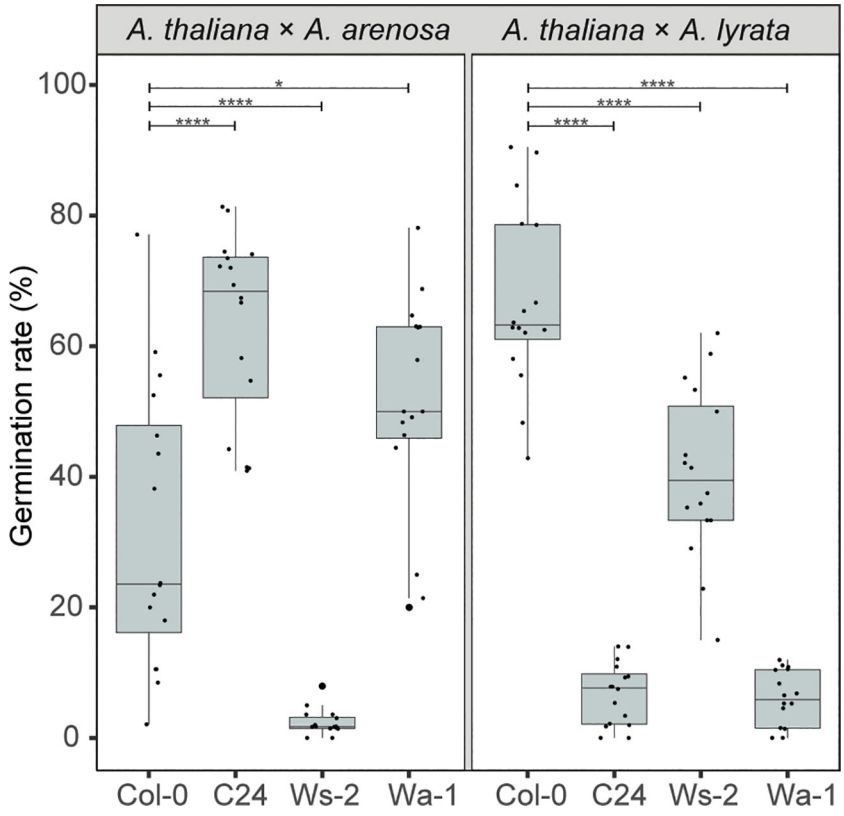


FIGURE 6
A. thaliana accessions affect hybrid barriers antagonistically. Germination rate of seeds from crossing *A. arenosa* (*A.a.*)/*A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (Col-0/C24/Ws-2/Wa-1) at 22°C. Biological replicates (siliques): Col-0 × *A.a.*, n = 16 (533 seeds); C24 × *A.a.*, n = 16 (805 seeds); Ws-2 × *A.a.*, n = 16 (947 seeds); Wa-1 × *A.a.*, n = 16 (805 seeds); Col-0 × *A.l.*, n = 16 (533 seeds); C24 × *A.l.*, n = 16 (821 seeds); Ws-2 × *A.l.*, n = 16 (572 seeds); Wa-1 × *A.l.*, n = 16 (759 seeds). Box plot contains scattered data points representing germination rates observed per silique. Outliers are plotted as large data points. Significance is indicated for comparisons between Col-0 × *A.a.*/Col-0 × *A.l.* and crosses involving other *A. thaliana* accessions (Welch's t-test; *$P \leq 0.05$; ****$P \leq 0.0001$).

arenosa did not significantly affect germination rate, but in crosses to *A. lyrata* germination rate was significantly reduced. Col-0 crossed to *A. lyrata* displayed reduced seed size (Figures S2, S7) due to early endosperm cellularization (Figure 1), and thus the mutation of *AGL40* may increase the frequency of early cellularization.

To test if cryptic genetic variation in the *agl35-1* mutant line could account for the observed phenotype, heterozygous *agl35-1* was introgressed twice to Col-0 and segregating progeny of selfed heterozygotes were crossed to *A. arenosa*. The segregating *agl35-1 A. thaliana* mothers were genotyped for the *agl35-1* insert and germination rate of hybrid seeds was scored (Figure 8B). Importantly, the germination rate of segregating wildtype plants (*agl35+/+)* was not significantly different from wildtype (Col-0) when crossed to *A. arenosa*. Both homozygous (*agl35-1 -/-)* and heterozygous *agl35-1* +/- plants crossed to *A. arenosa* displayed significantly lower germination rate than Col-0 crossed to *A. arenosa*, indicating that the increased strength of the hybrid barrier was caused by mutation of *AGL35*. Furthermore, the observation that the strength of the hybrid barrier can be caused by heterozygous (*agl35-1* +/-) plants crossed to *A. arenosa* indicates that the observed phenotype is caused by genetic interaction occurring in the fertilization products, the embryo or the endosperm.

4 Discussion

Recent advances in elucidating molecular mechanisms and genetic networks in hybrid endosperm lethality suggest that imprinted genes and genetic variation in the hybrid parents are important factors that can enhance or repress the frequency of the endosperm-based barrier (Bushell et al., 2003; Josefsson et al., 2006; Walia et al., 2009; Burkart-Waco et al., 2012; Burkart-Waco et al., 2013; Kradolfer et al., 2013; Schatlowski et al., 2014; Burkart-Waco et al., 2015; Rebernig et al., 2015; Wolff et al., 2015; Bjerkan et al., 2020). However, the mechanistic role of these factors and the interplay of genetic networks is largely unknown. For instance, ploidy can bypass the endosperm-based post-zygotic barrier, but a general role of ploidy cannot be attributed since ploidy plays a different role in maternal and paternal cross settings (Lafon-Placette et al., 2017). Furthermore, the role of gene dosage and genomic imprinting is supported by reports suggesting that mutation of MEGs and PEGs can overcome the endosperm-based post-zygotic barrier (Dilkes et al., 2008; Walia et al., 2009; Wolff et al., 2015; Borges et al., 2018), but a general role of imprinted genes cannot be defined since only some imprinted genes appear to have this effect (Walia et al., 2009; Burkart-Waco et al., 2015; Rebernig et al., 2015; Wolff et al., 2015). Since different accessions of the parental

10.3389/fpls.2023.1229060

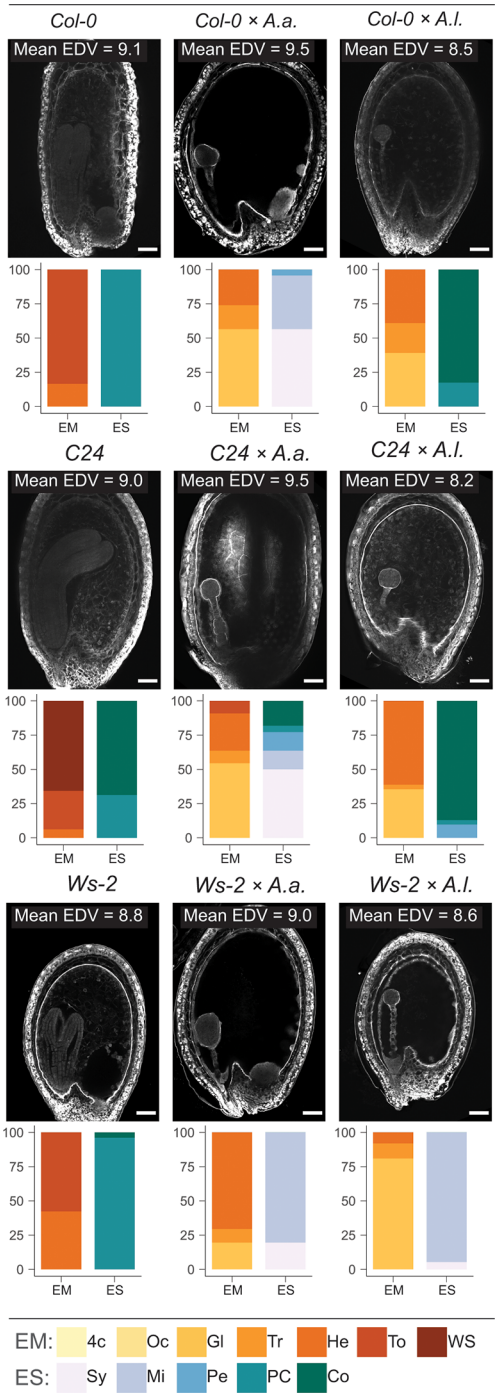


FIGURE 7
Effect of temperature, accession and hybridization on endosperm development. Confocal images showing endosperm cellularization of 6 DAP Feulgen-stained seeds from crossing *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (accession Col-0 or C24 or Ws-2) at 22°C. Scale bar = 50 µm. Mean endosperm division value (EDV) is shown within each image, n_{EDV} = 10 seeds. Beneath each image quantification of the described embryo and endosperm stages is shown as bar charts: Col-0, n = 36; Col-0 × A.a., n = 23; Col-0 × A.l., n = 23; C24, n = 32; C24 × *A.a.*, n = 22; C24 × *A.l.*, n = 31; WS-2, n = 26; WS-2 × *A.a.*, n = 51; WS-2 × *A.l.*, n = 37. Embryo stages (EM): 4c, 4-cell; Oc, Octant; Gl, Globular; Tr, Transition; He, Heart; To, Torpedo; WS, Walking stick; Endosperm cellularization stages (ES): Sy, Syncytial endosperm; Mi, Micropylar endosperm cellularization; Pe, Peripheral endosperm cellularization; PC, Partially complete endosperm cellularization; Co, Complete endosperm cellularization.

individuals in a hybrid cross can partly bypass the endosperm-based post-zygotic barrier without any change in ploidy (Walia et al., 2009; Wolff et al., 2015), the role of maternal and paternal genomes cannot be generalized.

Here we demonstrate that interspecies hybrid seeds from crossing *A. lyrata* or *A. arenosa* as the paternal parent to *A. thaliana* mothers show antagonistic endosperm cellularization phenotypes, with late cellularization in crosses with *A. arenosa*

10.3389/fpls.2023.1229060

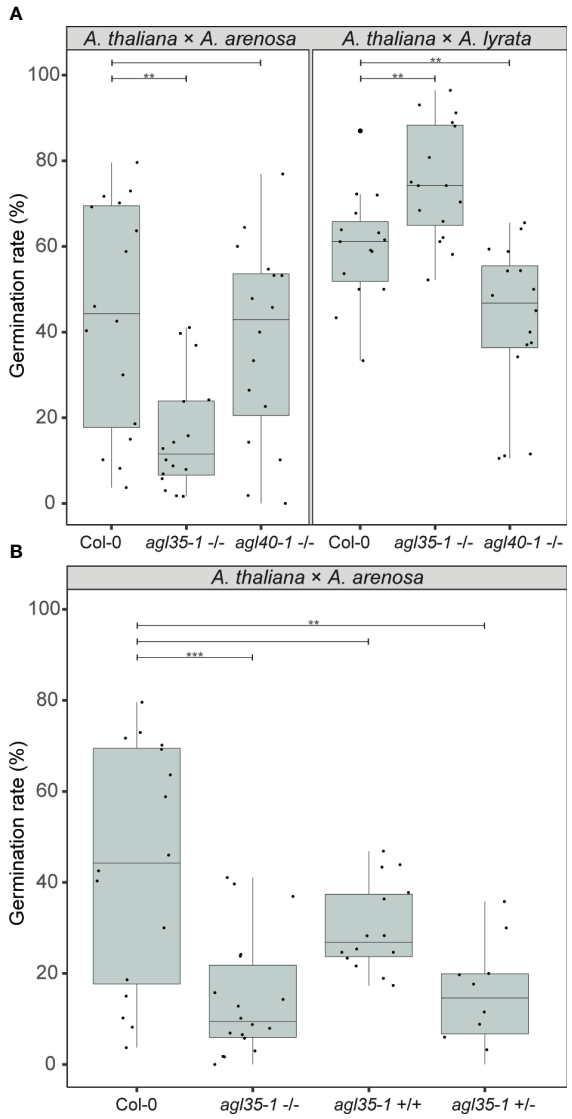


FIGURE 8
Genetic parameters influence the hybrid barrier. **(A)** Germination rate of seeds from crossing *A. arenosa* (*A.a.*)/*A. lyrata* (*A.l.*) as pollen donors to *A. thaliana* (Col-0), single mutants *agl35-1*-/- and *agl40-1*-/- at 18°C. Biological replicates (siliques): Col-0 × *A.a.*, n = 16 (833 seeds); *agl35-1*-/- × *A.a.*, n = 16 (904 seeds); *agl40-1*-/- × *A.a.*, n = 16 (879 seeds); Col-0 × *A.l.*, n = 16 (457 seeds); *agl35-1*-/- × *A.l.*, n = 16 (509 seeds); *agl40-1*-/- × *A.l.*, n = 16 (525 seeds). **(B)** Germination rate of seeds from crossing *A. arenosa* (*A.a.*) as pollen donor to *A. thaliana* (Col-0), homozygous *agl35-1*-/-, twice introgressed *agl35*-/- into Col-0 (*agl35*+/+), and heterozygous *agl35-1* +/- at 18°C. Biological replicates (siliques): Col-0 × *A.a.*, n = 16 (833 seeds); *agl35-1*-/- × *A.a.*, n = 18 (1064 seeds); *agl35*+/+ × *A.a.*, n = 14 (1562 seeds); *agl35-1* +/- × *A.a.*, n = 10 (525 seeds). Box plot contains scattered data points representing germination rates observed per silique. Outliers are plotted as large data points. Significance is indicated for the comparisons between Col-0 crosses and the mutant crosses (Welch's t-test; **P ≤ 0.01; ***P ≤ 0.001).

and early cellularization in crosses with *A. lyrata*. In both cases, cellularization failure results in an endosperm-based hybrid barrier and reduced viability of germinating seeds. This compares to previous studies where timing of endosperm cellularization is influenced by the paternal species in reciprocal *A. arenosa* and *A. lyrata* interspecies crosses and in crosses within the genus *Capsella* (Rebernig et al., 2015; Lafon-Placette et al., 2017). Intriguingly, we find that a temperature gradient leads to diametrically opposed cellularization phenotype responses in hybrid endosperm with *A. arenosa* or *A. lyrata* as pollen donors. In addition, *A. thaliana* accession genotypes also influence hybrid seed viability in opposite directions. To this end, we demonstrate that single gene mutation in *A. thaliana* MADS-box TF *AGL35* independently can affect the germination rates of *A. arenosa* or *A. lyrata* hybrid seeds in opposite directions.

4.1 Ectopic timing of endosperm developmental phase change

The endosperm genetic markers EE-GFP and TE1-GFP mark syncytial endosperm development before cellularization and cellular endosperm stages, respectively (van Ekelenburg et al., 2023). We show that timing of the developmental phase change

connected to endosperm cellularization is disturbed in hybrid seeds. In the Col-0 × *A. arenosa* hybrid seeds that fail to cellularize, the EE-GFP marker continues to be expressed throughout endosperm development, indicating phase change failure. In the Col-0 × *A. lyrata* hybrid seeds, characterized by early cellularization, the developmental time point of TE1-GFP expression indicates occurrence of a premature phase change. These findings support that not only the timing of endosperm cellularization is affected in these developing hybrid seeds, but also the developmental timing of the genetic network associated with endosperm phase change and maturation occurring at cellularization. In accordance with the incomplete endosperm hybridization barrier, prolonged expression of EE-GFP in Col-0 × *A. arenosa* seeds and precocious expression of TE1-GFP in Col-0 × *A. lyrata* seeds were not observed in all individual seeds. This suggests that gene regulation associated with the endosperm phase change within each hybrid seed varies and potentially is affected by genetic or epigenetic variation that modulates threshold levels for gene activation or repression.

In our system the sole difference between the two hybrids is the paternal parent, indicating a trans-acting mechanism, where differential expression from *A. lyrata* and *A. arenosa* genomes regulates the genetic markers expressed from the *A. thaliana* genome. Supporting this hypothesis, paternal transmission of mutants in NUCLEAR RNA POLYMERASE D1 (NRPD1) can bypass the cellularization phenotype in paternal excess inter-ploidy crosses (Erdmann et al., 2017; Martinez et al., 2018). NRPD1 is a main component in the the RNA-directed DNA methylation (RdDM) pathway, resulting in small RNA directed gene regulation by *de novo* DNA methylation (Law et al., 2013; Kirkbride et al., 2019) and could be a potential trans-acting regulatory mechanism (Erdmann et al., 2017). Future experiments to identify transcriptional differences from the parental genomes in hybrid seeds may point at key genes and mechanisms responsible for ectopic timing of the endosperm developmental phase change.

4.2 Temperature and accession affect the viability of hybrid seeds

We found that by increasing the temperature from 14°C to 26°C, Col-0 × *A. arenosa* seeds display significantly decreased germination rates (more than 30%). The same temperature range has an opposite effect in Col-0 × *A. lyrata* seeds resulting in increased germination rates (more than 50%). This demonstrates a temperature dependent genetic mechanism that acts antagonistically when *A. thaliana* is crossed to *A. arenosa* or *A. lyrata* and produces diametrically opposed cellularization phenotype responses in the hybrid endosperm.

Interestingly, in *Brassica oleracea*, temperature affects abscisic acid (ABA) levels specifically in the endosperm and cooler temperatures obstruct the breakdown of ABA in the desiccating endosperm (Chen et al., 2021). This is consistent with a recent report demonstrating that *A. thaliana* inter-ploidy uncellularized endosperm induced by paternal excess is correlated with increased ABA levels, suggesting that endosperm cellularization is connected to dehydration responses in the developing embryo (Xu et al., 2023). ABA catabolism in response to temperature may therefore be a potent mechanism to explain the temperature influence on the hybrid barrier when *A. thaliana* is crossed to *A. arenosa*. In a similar manner, we further speculate that precocious cellularization in crosses with *A. lyrata* may be associated with a similar mechanism that triggers ABA breakdown, but this needs further investigation.

Notably, the effect of using different *A. thaliana* accessions in the hybrid crosses is larger than the temperature effect (close to 70% difference), and even larger than the interploidy effect. While the Col-0 and Ws-2 *A. thaliana* accessions resulted in a generally higher germination rate when hybridized with *A. lyrata* compared to *A. arenosa*, the C24 accession had the opposite effect. The way the accessions affected hybrid seed viability in opposite directions may point at a similar mechanism as observed in the temperature experiment. However, our results do not readily explain the observed germination rates by cellularization phenotype alone, and further investigations are required to resolve these observations.

Crossing tetraploid *A. thaliana* Wa-1 to diploid *A. arenosa* increases hybrid seed survival (Josefsson et al., 2006). In addition, ploidy affects the strength of the hybrid barrier in crosses between *A. arenosa* and *A. lyrata*, where higher ploidy in *A. lyrata* increases the hybrid seed survival rate, while higher ploidy in *A. arenosa* causes total seed lethality (Lafon-Placette et al., 2017). Our data show that a similar effect is found using the diploid accession C24, suggesting that C24 may have a higher effective ploidy and endosperm balance number [EBN; (Johnston and Hanneman, 1982)] compared to Col-0 and Ws-2. This corresponds well with the hypothesis that *A. lyrata* has a lower EBN compared to *A. arenosa* (Lafon-Placette and Köhler, 2016), explaining why crosses with C24 or Wa-1 decrease seed viability in the *A. lyrata* hybrid. However, it does not explain why *A. lyrata* crosses with the diploid C24 is more detrimental than crosses with the tetraploid Wa-1, suggesting that accession genotypes, in addition to ploidy, has an effect on the endosperm-based hybridization barrier.

4.3 AGL35 influences endosperm cellularization in hybrid seeds

AGAMOUS-LIKE (AGL) type I MADS-box TFs are highly expressed in the seed, specifically during endosperm cellularization (Bemer et al., 2010; Zhang et al., 2018; Bjerkan et al., 2020). Their importance in the endosperm-based hybridization barrier has been suggested by several studies (Josefsson et al., 2006; Walia et al., 2009; Bjerkan et al., 2020) and it has been hypothesized that timing of endosperm cellularization requires a stoichiometric balance between members of different MADS-box protein complexes (Batista et al., 2019). In this study we demonstrate that mutation in *A. thaliana AGL35* has a highly significant and opposite effect on the hybrid barrier phenotype when crossed to *A. lyrata* and *A. arenosa*, respectively. *AGL35* is bi-allelicly expressed in the chalazal

endosperm (Bemer et al., 2010; Bjerkan et al., 2020) and upregulated in crosses between *A. thaliana* and *A. arenosa* compared with compatible crosses (Walia et al., 2009). Our results indicate that AGL35 is involved in the transition from syncytial to a cellularized endosperm, and may function as a promoter of cellularization, as mutant crosses to *A. arenosa* result in lower seed survival, while mutant crosses to *A. lyrata* result in increased survival compared to Col-0 crosses. Interestingly, a massive-multiplexed yeast two-hybrid study identified interaction between AGL62 and AGL35 (Trigg et al., 2017). These AGL TFs have seemingly antagonistic functions as AGL62 is a suppressor of endosperm cellularization (Kang et al., 2008). Paternal excess interploidy crosses cause increased *AGL62* expression, correlated with endosperm cellularization failure (Erilova et al., 2009). *AGL62* is also a direct target of the FIS PRC2 complex (Hehenberger et al., 2012) whereas we could see no direct effect on the *A. arenosa* or *A. lyrata* hybrid with *A. thaliana* by mutation of FIS PRC2. The antagonistic effects of single gene mutation of *AGL35* is intriguing, and we speculate that expression differences between *A. arenosa* and *A. lyrata* in the hybrid endosperm may account for our observations but future investigation of this interaction and the role of AGL35 in regulation of endosperm-based hybridization barriers is required.

4.4 Conclusions

The findings in this study introduce a rigorous model system for the dissection of the influence of abiotic and genetic parameters in hybrid admixture, and have a large potential to support breeding and climate research. Further examination and usage of these approaches could help pinpoint genes, networks or gene dosage balances that are involved in overcoming the endosperm-based hybridization barrier. Species previously thought to be unable to hybridize due to postzygotic seed lethality may be able to do so given favorable conditions, and a similar effect could also apply to interploidy hybrids.

Currently, it is not known if the temperature effect on hybridization success is mediated by the same genetic network that is operated by changes in ploidy or genetic variation. Phenotypically, the temperature effect restores defects in timing of cellularization, but it is not known if the trigger is upstream or downstream of the causative genetic network. Elucidation of the genetic, epigenetic and mechanistic basis for this cross talk between the genic and environmental factors is therefore essential for our understanding of the plasticity of endosperm-based hybridization barriers.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Files. Further inquiries can be directed to the corresponding author/s.

Author contributions

PG, RA, KB, and AB designed the research; RA, IM, and KB performed the experiments; PG, RA, KB, and AB analyzed and discussed the data; PG, RA, KB, and AB wrote the article; All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Norwegian Research Council (FRIPRO grants no. 276053 and 262247) to PG and AB. RA and IM were supported by the Norwegian Ministry of Education.

Acknowledgment

We thank the Laboratory of Flow Cytometry, Institute of Botany, Academy of Sciences (Czech Republic) for assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1229060/full#supplementary-material

SUPPLEMENTARY DATA SHEET 1
Flow Cytometry.

SUPPLEMENTARY DATA SHEET 2
Germination assays.

SUPPLEMENTARY DATA SHEET 3
Number of nuclei in crosses

SUPPLEMENTARY DATA SHEET 4
Feulgen phenotype observations

References

Batista, R. A., and Kohler, C. (2020). Genomic imprinting in plants-revisiting existing models. *Genes Dev.* 34, 24–36. doi: 10.1101/gad.332924.119

Batista, R. A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D. D., et al. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *Elife* 8, e50541. doi: 10.7554/eLife.50541.sa2

Bemer, M., Heijmans, K., Airoldi, C., Davies, B., and Angenent, G. C. (2010). An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. *Plant Physiol.* 154, 287–300. doi: 10.1104/pp.110.160770

Bjerkan, K. N., Hornslien, K. S., Johannessen, I. M., Krabberød, A. K., van Ekelenburg, Y. S., Kalantarian, M., et al. (2020). Genetic variation and temperature affects hybrid barriers during interspecific hybridization. *Plant J.* 101, 122–140. doi: 10.1111/tpj.14523

Borges, F., Parent, J.-S., van Ex, F., Wolff, P., Martínez, G., Köhler, C., et al. (2018). Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in *Arabidopsis*. *Nat. Genet.* 50, 186–192. doi: 10.1038/s41588-017-0032-5

Braselton, J. P., Wilkinson, M. J., and Clulow, S. A. (1996). Feulgen staining of intact plant tissues for confocal microscopy. *Biotech. Histochem* 71, 84–87. doi: 10.3109/10520299609117139

Brink, R. A., and Cooper, D. C. (1947). The endosperm in seed development. *Bot. Rev.* 13, 479–541. doi: 10.1007/BF02861549

Burkart-Waco, D., Josefsson, C., Dilkes, B., Kozloff, N., Torjek, O., Meyer, R., et al. (2012). Hybrid incompatibility in *Arabidopsis* is determined by a multiple-locus genetic network. *Plant Physiol.* 158, 801–812. doi: 10.1104/pp.111.188706

Burkart-Waco, D., Ngo, K., Dilkes, B., Josefsson, C., and Comai, L. (2013). Early disruption of maternal-zygotic interaction and activation of defense-like responses in *Arabidopsis* interspecific crosses. *Plant Cell* 25, 2037–2055. doi: 10.1105/tpc.112.108258

Burkart-Waco, D., Ngo, K., Lieberman, M., and Comai, L. (2015). Perturbation of parentally biased gene expression during interspecific hybridization. *PloS One* 10, e0117293. doi: 10.1371/journal.pone.0117293

Bushell, C., Spielman, M., and Scott, R. J. (2003). The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *Plant Cell* 15, 1430–1442. doi: 10.1105/tpc.010496

Chen, C., Begcy, K., Liu, K., Folsom, J. J., Wang, Z., Zhang, C., et al. (2016). Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol.* 171, 606–622. doi: 10.1104/pp.15.01992

Chen, Z. J., Comai, L., and Pikaard, C. S. (1998). Gene dosage and stochastic effects determine the severity and direction of uniparental ribosomal RNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. *Proc. Natl. Acad. Sci. U. S. A* 95, 14891–14896. doi: 10.1073/pnas.95.25.14891

Chen, X., Yoong, F.-Y., O'Neill, C. M., and Penfield, S. (2021). Temperature during seed maturation controls seed vigour through ABA breakdown in the endosperm and causes a passive effect on DOG1 mRNA levels during entry into quiescence. *New Phytol.* 232, 1311–1322. doi: 10.1111/nph.17646

Comai, L., Tyagi, A. P., Winter, K., Holmes-Davis, R., Reynolds, S. H., Stevens, Y., et al. (2000). Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12, 1551–1568. doi: 10.1105/tpc.12.9.1551

Cooper, D. C., and Brink, R. A. (1942). The endosperm as a barrier to interspecific hybridization in flowering plants. *Science* 95, 75–76. doi: 10.1126/science.95.2455.75

Cornejo, P., Camadro, E. L., and Masuelli, R. W. (2012). Molecular bases of the postzygotic barriers in interspecific crosses between the wild potato species *Solanum acaule* and *Solanum commersonii*. *Genome* 55, 605–614. doi: 10.1139/g2012-047

De Bodt, S., Raes, J., Van de Peer, Y., and Theißen, G. (2003). And then there were many: MADS goes genomic. *Trends Plant Sci.* 8, 475–483. doi: 10.1016/j.tplants.2003.09.006

Dilkes, B. P., Spielman, M., Weizbauer, R., Watson, B., Burkart-Waco, D., Scott, R. J., et al. (2008). The maternally expressed WRKY transcription factor TTG2 controls lethality in interploidy crosses of *Arabidopsis*. *PloS Biol.* 6, 2707–2720. doi: 10.1371/journal.pbio.0060308

Dinu, I. I., Hayes, R. J., Kynast, R. G., Phillips, R. L., and Thill, C. A. (2005). Novel inter-series hybrids in *Solanum*, section *Petota*. *Theor. Appl. Genet.* 110, 403–415. doi: 10.1007/s00122-004-1782-x

Dziasek, K., Simon, L., Lafon-Placette, C., Laenen, B., Wärdig, C., Santos-González, J., et al. (2021). Hybrid seed incompatibility in *Capsella* is connected to chromatin condensation defects in the endosperm. *PloS Genet.* 17, e1009370. doi: 10.1371/journal.pgen.1009370

Erdmann, R. M., Satyaki, P. R. V., Klosinska, M., and Gehring, M. (2017). A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep.* 21, 3364–3372. doi: 10.1016/j.celrep.2017.11.078

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Mittelsten Scheid, O., et al. (2009). Imprinting of the polycomb group gene *MEDEA* serves as a ploidy sensor in *Arabidopsis*. *PloS Genet.* 5, e1000663. doi: 10.1371/journal.pgen.1000663

Escobar-Restrepo, J.-M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W.-C., et al. (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* 317, 656–660. doi: 10.1126/science.1143562

Flores-Vergara, M. A., Oneal, E., Costa, M., Villarino, G., Roberts, C., De Luis Balaguer, M. A., et al. (2020). Developmental analysis of *Mimulus* seed transcriptomes reveals functional gene expression clusters and four imprinted, endosperm-expressed genes. *Front. Plant Sci.* 11, 132. doi: 10.3389/fpls.2020.00132

Florez-Rueda, A. M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U., and Städler, T. (2016). Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol. Biol. Evol.* 33, 2935–2946. doi: 10.1093/molbev/msw175

Folsom, J. J., Begcy, K., Hao, X., Wang, D., and Walia, H. (2014). Rice *Fertilization-Independent Endosperm1* regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* 165, 238–248. doi: 10.1104/pp.113.232413

Fox, J., and Weisberg, S. (2019). An R companion to applied regression, *3rd ed.* (Thousand Oaks, CA: Sage Publication Inc.).

Grini, P. E., Jürgens, G., and Hülskamp, M. (2002). Embryo and endosperm development is disrupted in the female gametophytic *capulet* mutants of *Arabidopsis*. *Genetics* 162, 1911–1925. doi: 10.1093/genetics/162.4.1911

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M. A., and Gagliano, W. B. (1998). Maternal control of embryogenesis by *MEDEA*, a *Polycomb* group gene in *Arabidopsis*. *Science* 280, 446 LP–446450. doi: 10.1126/science.280.5362.446

Guitton, A.-E., Page, D. R., Chambrier, P., Lionnet, C., Faure, J.-E., Grossniklaus, U., et al. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. *Development* 131, 2971–2981. doi: 10.1242/dev.01168

Haig, D., and Westoby, M. (1988). On limits to seed production. *Am. Nat.* 131, 757–759. doi: 10.1086/284817

Haig, D., and Westoby, M. (1991). Genomic imprinting in endosperm: Its effect on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. *Philos. Trans. R. Soc Lond. B Biol. Sci.* 333, 1–13. doi: 10.1098/rstb.1991.0057

Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development* 139, 2031–2039. doi: 10.1242/dev.077057

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. doi: 10.1104/pp.19.00320

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., and Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 65, 798–806. doi: 10.1111/j.1365-313X.2010.04466.x

Johnston, S. A., and Hanneman, R. E. (1982). Manipulations of endosperm balance number overcome crossing barriers between diploid *Solanum* species. *Science* 217, 446–448. doi: 10.1126/science.217.4558.446

Jørgensen, M. H., Ehrich, D., Schmickl, R., Koch, M. A., and Brysting, A. K. (2011). Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evol. Biol.* 11, 1–13. doi: 10.1186/1471-2148-11-346

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16, 1322–1328. doi: 10.1016/j.cub.2006.05.045

Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20, 635–647. doi: 10.1105/tpc.107.055137

Kassambara, A. (2023) *ggpubr: "ggplot2" based publication ready plots. R package version 0.6.0.* Available at: https://rpkgs.datanovia.com/ggpubr/.

Kinser, T. J., Smith, R. D., Lawrence, A. H., Cooley, A. M., Vallejo-Marín, M., Conradi Smith, G. D., et al. (2021). Endosperm-based incompatibilities in hybrid monkeyflowers. *Plant Cell* 33, 2235–2257. doi: 10.1093/plcell/koab117

Kirkbride, R. C., Lu, J., Zhang, C., Mosher, R. A., Baulcombe, D. C., and Chen, Z. J. (2019). Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 116, 2761–2766. doi: 10.1073/pnas.1807621116

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., et al. (1999). Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A* 96, 4186–4191. doi: 10.1073/pnas.96.7.4186

Koch, M. A. (2019). The plant model system *Arabidopsis* set in an evolutionary, systematic, and spatio-temporal context. *J. Exp. Bot.* 70, 55–67. doi: 10.1093/jxb/ery340

Koenig, D., and Weigel, D. (2015). Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nat. Rev. Genet.* 16, 285–298. doi: 10.1038/nrg3883

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003). *Arabidopsis* MSI1 is a component of the MEA/FIE *Polycomb* group complex and required for seed development. *EMBO J.* 22, 4804–4814. doi: 10.1093/emboj/cdg444

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A., and Köhler, C. (2013). An imprinted gene underlies postzygotic reproductive isolation in *Arabidopsis thaliana*. *Dev. Cell* 26, 525–535. doi: 10.1016/j.devcel.2013.08.006

Lafon-Placette, C., Johannessen, I. M., Hornslien, K. S., Ali, M. F., Bjerkan, K. N., Bramsiepe, J., et al. (2017). Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc. Natl. Acad. Sci. U. S. A* 114, E1027–E1035. doi: 10.1073/pnas.1615123114

Lafon-Placette, C., and Köhler, C. (2016). Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* 25, 2620–2629. doi: 10.1111/mec.13552

Law, JA, Du, J, Hale, CJ, Feng, S, Krajewski, K, Palanca, AMS, et al (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498, 385-389.

Levene, H. (1960). *Robust tests for equality of variances. In I. Olkin (Ed.), Contributions to probability and statistics* (Palo Alto, CA: Stanford University Press).

Lindsey, B. E.III, AU-Rivero, L., AU-Calhoun, C. S., AU-Grotewold, E., and AU-Brkljacic, J. (2017). Standardized method for high-throughput sterilization of *Arabidopsis* seeds. *J. Vis. Exp.*, 2017 (128), e56587. doi: 10.3791/56587

Lloyd, A., Morgan, C., H Franklin, F. C., and Bomblies, K. (2018). Plasticity of meiotic recombination rates in response to temperature in *Arabidopsis*. *Genetics* 208, 1409–1420. doi: 10.1534/genetics.117.300588

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E. S., Peacock, W. J., and Chaudhury, A. M. (1999). Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A* 96, 296–301. doi: 10.1073/pnas.96.1.296

Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* 18, 50–60. doi: 10.1214/aoms/1177730491

Martienssen, R. A. (2010). Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. *New Phytol.* 186, 46–53. doi: 10.1111/j.1469-8137.2010.03193.x

Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L. L., et al. (2018). Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat. Genet.* 50, 193–198. doi: 10.1038/s41588-017-0033-4

Molnár-Láng, M., and Sutka, J. (1994). The effect of temperature on seed set and embryo development in reciprocal crosses of wheat and barley. *Euphytica* 78, 53–58. doi: 10.1007/BF00021397

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nasrallah, M. E., Yogeeswaran, K., Snyder, S., and Nasrallah, J. B. (2000). *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. *Plant Physiol.* 124, 1605–1614. doi: 10.1104/pp.124.4.1605

Ng, D. W., Lu, J., and Chen, Z. J. (2012). Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr. Opin. Plant Biol.* 15, 154–161. doi: 10.1016/j.pbi.2012.01.007

Oneal, E., Willis, J. H., and Franks, R. G. (2016). Disruption of endosperm development is a major cause of hybrid seed inviability between *Mimulus guttatus* and *Mimulus nudatus*. *New Phytol.* 210, 1107–1120. doi: 10.1111/nph.13842

Parřenicová, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: New openings to the MADS world. *Plant Cell* 15, 1538–1551. doi: 10.1105/tpc.011544

Paul, P., Dhatt, B. K., Sandhu, J., Hussain, W., Irvin, L., Morota, G., et al. (2020). Divergent phenotypic response of rice accessions to transient heat stress during early seed development. *Plant Direct* 4, e00196. doi: 10.1002/pld3.196

R Core Team (2023). *R: A language and environment for statistical computing* (R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Rebernig, 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. doi: 10.1371/journal.pgen.1005295

Rieseberg, L. H., and Willis, J. H. (2007). Plant speciation. *Science* 317, 910–914. doi: 10.1126/science.1137729

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. doi: 10.1534/genetics.119.302056

Roy, A. K., Malaviya, D. R., and Kaushal, P. (2011). Generation of interspecific hybrids of *Trifolium* using embryo rescue. *Methods Mol. Biol.* 710:141–151. doi: 10.1007/978-1-61737-988-8_12

Schatlowski, N., and Kohler, C. (2012). Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. *J. Exp. Bot.* 63, 6059–6067. doi: 10.1093/jxb/ers288

Schatlowski, N., Wolff, P., Santos-Gonzalez, J., Schoft, V., Siretskiy, A., Scott, R., et al. (2014). Hypomethylated pollen bypasses the interploidy hybridization barrier in *Arabidopsis*. *Plant Cell* 26, 3556–3568. doi: 10.1105/tpc.114.130120

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. doi: 10.1242/dev.125.17.3329

Sekine, D., Ohnishi, T., Furuumi, H., Ono, A., Yamada, T., Kurata, N., et al. (2013). Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J.* 76, 792–799. doi: 10.1111/tpj.12333

Sharma, H. C. (1999). Embryo rescue following wide crosses. *Methods Mol. Biol.* 111, 293–307. doi: 10.1385/1-59259-583-9:293

Shirzadi, R., Andersen, E. D., Bjerkan, K. N., Gloeckle, B. M., Heese, M., Ungru, A., et al. (2011). Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of *AGAMOUS-LIKE36*. *PloS Genet.* 7, e1001303. doi: 10.1371/journal.pgen.1001303

Sukno, S., Ruso, J., Jan, C. C., Melero-Vara, J. M., and Fernandez-Martinez, J. M. (1999). Interspecific hybridization between sunflower and wild perennial *Helianthus* species via embryo rescue. *Euphytica* 106, 69–78. doi: 10.1023/A:1003524822284

Tonosaki, K., Sekine, D., Ohnishi, T., Ono, A., Furuumi, H., Kurata, N., et al. (2018). Overcoming the species hybridization barrier by ploidy manipulation in the genus *Oryza*. *Plant J.* 93, 534–544. doi: 10.1111/tpj.13803

Trigg, S. A., Garza, R. M., MacWilliams, A., Nery, J. R., Bartlett, A., Castanon, R., et al. (2017). CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat. Methods* 14, 819–825. doi: 10.1038/nmeth.4343

Ungru, A., Nowack, M. K., Reymond, M., Shirzadi, R., Kumar, M., Biewers, S., et al. (2008). Natural variation in the degree of autonomous endosperm formation reveals independence and constraints of embryo growth during seed development in *Arabidopsis thaliana*. *Genetics* 179, 829–841. doi: 10.1534/genetics.107.084889

van Ekelenburg, Y. S., Hornslien, K. S., Van Hautegem, T., Fendrych, M., Van Isterdael, G., Bjerkan, K. N., et al. (2023). Spatial and temporal regulation of parent-of-origin allelic expression in the endosperm. *Plant Physiol.* 191, 986–1001. doi: 10.1093/plphys/kiac520

Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J., and Comai, L. (2009). Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* 19, 1128–1132. doi: 10.1016/j.cub.2009.05.068

Wang, L., Yuan, J., Ma, Y., Jiao, W., Ye, W., Yang, D.-L., et al. (2018). Rice interploidy crosses disrupt epigenetic regulation, gene expression, and seed development. *Mol. Plant* 11, 300–314. doi: 10.1016/j.molp.2017.12.006

Welch, B. L. (1947). The generalisation of student's problems when several different population variances are involved. *Biometrika* 34, 28–35. doi: 10.1093/biomet/34.1-2.28

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York: Springer-Verlag).

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023) *dplyr: a grammar of data manipulation*. Available at: https://dplyr.tidyverse.orghttps://github.com/tidyverse/dplyr.

Widmer, A., Lexer, C., and Cozzolino, S. (2009). Evolution of reproductive isolation in plants. *Heredity* 102, 31–38. doi: 10.1038/hdy.2008.69

Wolff, P., Jiang, H., Wang, G., Santos-Gonzalez, J., and Köhler, C. (2015). Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *Elife* 4, e10074. doi: 10.7554/eLife.10074.020

Xu, W., Sato, H., Bente, H., Santos-González, J., and Köhler, C. (2023). Endosperm cellularization failure induces a dehydration-stress response leading to embryo arrest. *Plant Cell* 35, 874–888. doi: 10.1093/plcell/koac337

Zhang, H.-Y., Luo, M., Johnson, S. D., Zhu, X.-W., Liu, L., Huang, F., et al. (2016). Parental genome imbalance causes post-zygotic seed lethality and deregulates imprinting in rice. *Rice* 9, 1–12. doi: 10.1186/s12284-016-0115-4

Zhang, S., Wang, D., Zhang, H., Skaggs, M. I., Lloyd, A., Ran, D., et al. (2018). FERTILIZATION-INDEPENDENT SEED-Polycomb Repressive Complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. *Plant Physiol.* 177, 285–299. doi: 10.1104/pp.17.00534

Supplementary Material

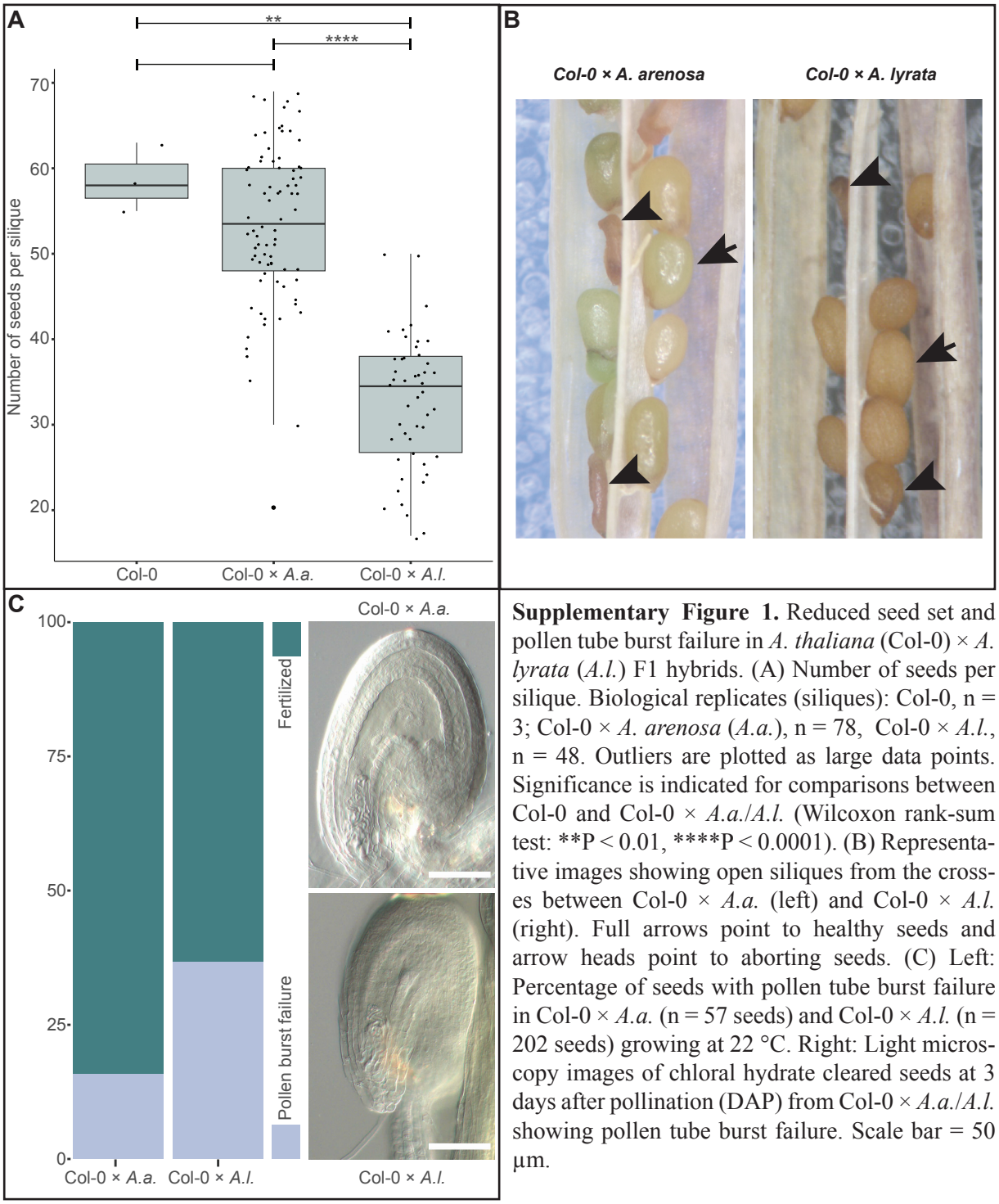
Genetic and environmental manipulation of *Arabidopsis* hybridization barriers uncover antagonistic functions in endosperm cellularization

Katrine N. Bjerkan[1,2*], Renate M. Alling[1,2*], Ida V. Myking[1,2], Anne K. Brysting[1,2] and Paul E. Grini[1]

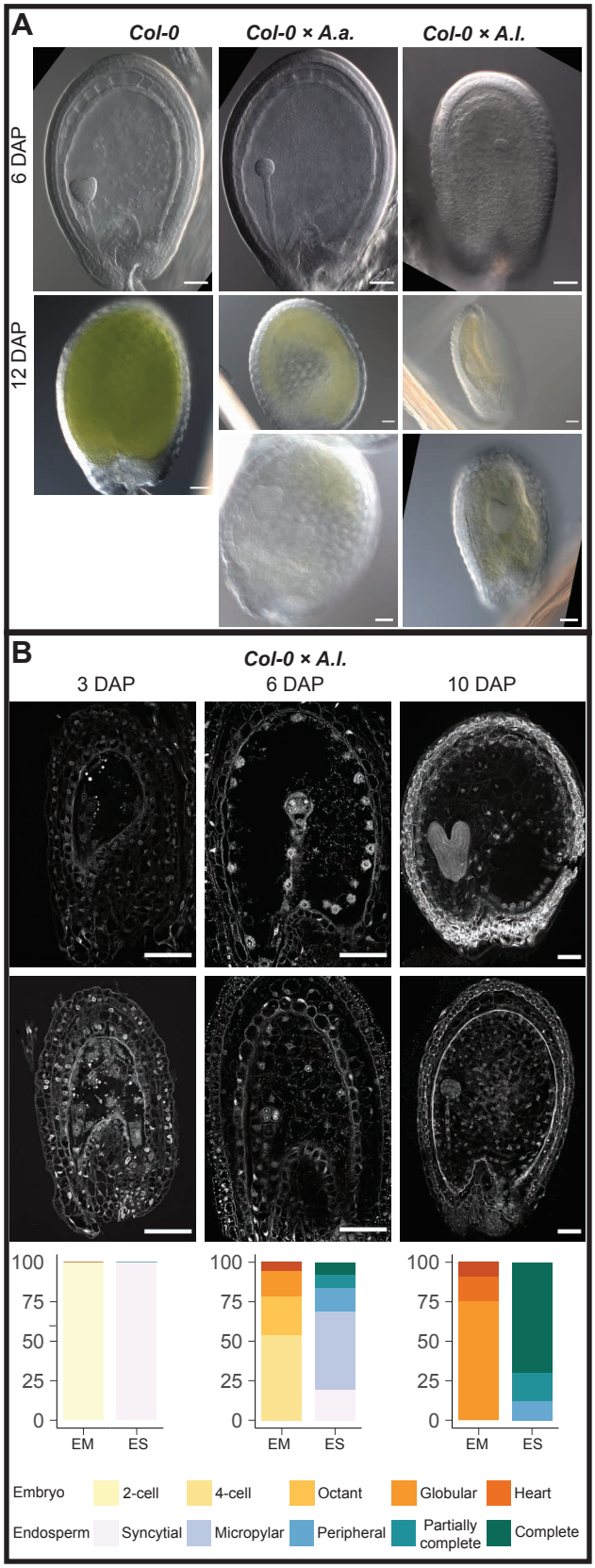
[1] Section for Genetics and Evolutionary Biology (EVOGENE), Department of Biosciences, University of Oslo, Oslo, Norway.
[2] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway.

Correspondence: Paul E. Grini: paul.grini@ibv.uio.no

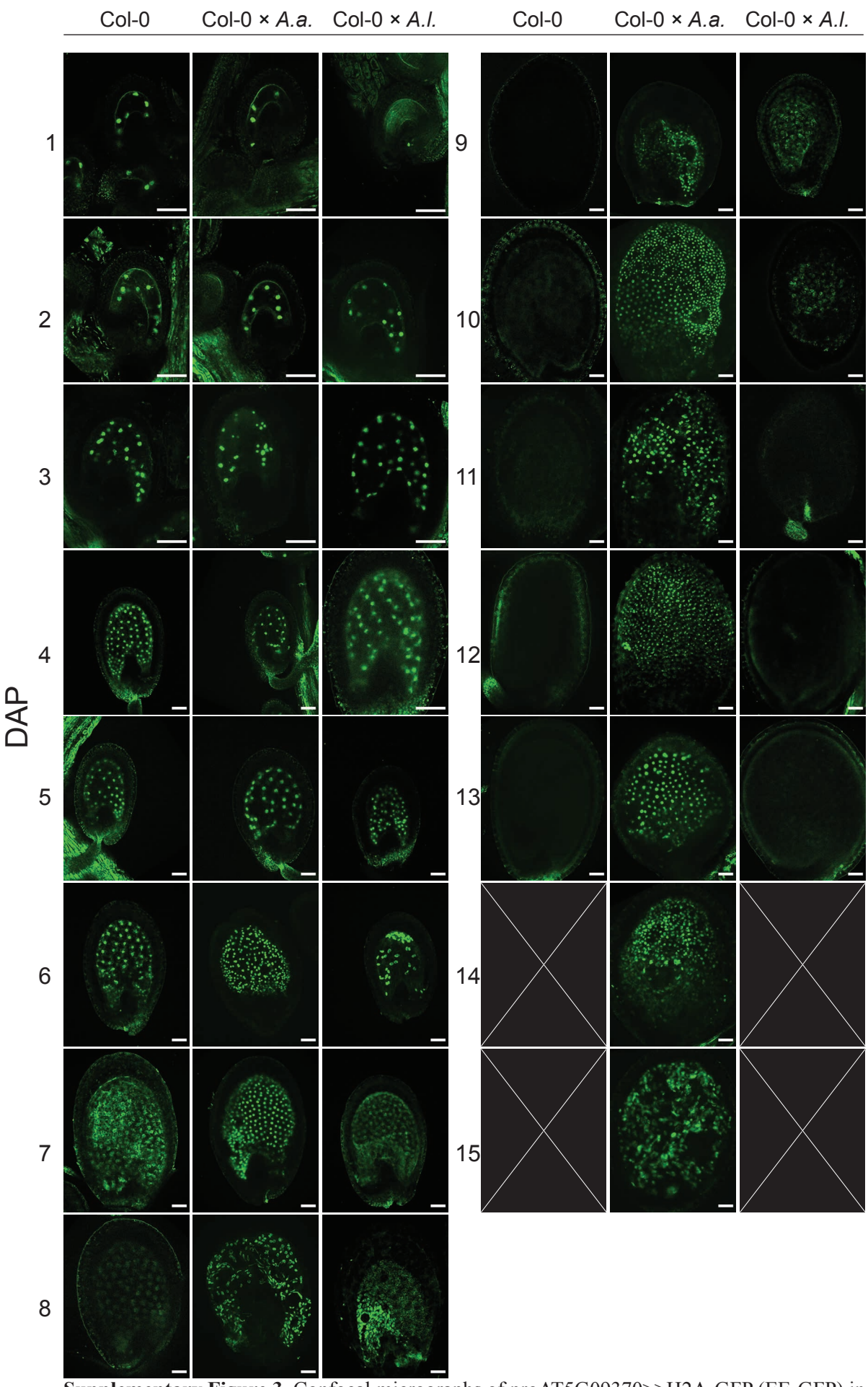
*These authors contributed equally to this work.



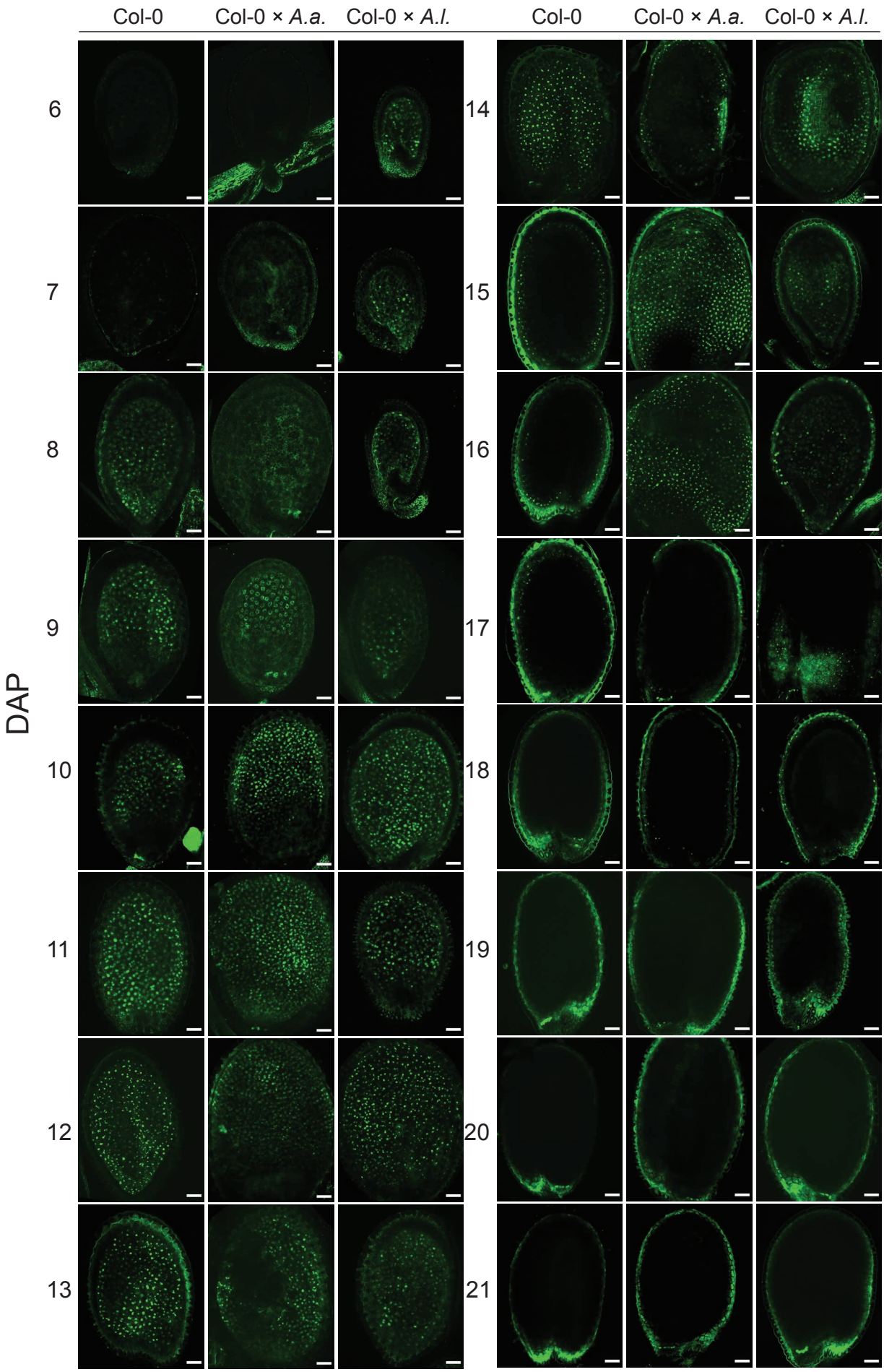
Supplementary Figure 1. Reduced seed set and pollen tube burst failure in *A. thaliana* (Col-0) × *A. lyrata* (*A.l.*) F1 hybrids. (A) Number of seeds per silique. Biological replicates (siliques): Col-0, n = 3; Col-0 × *A. arenosa* (*A.a.*), n = 78, Col-0 × *A.l.*, n = 48. Outliers are plotted as large data points. Significance is indicated for comparisons between Col-0 and Col-0 × *A.a.*/*A.l.* (Wilcoxon rank-sum test: **P < 0.01, ****P < 0.0001). (B) Representative images showing open siliques from the crosses between Col-0 × *A.a.* (left) and Col-0 × *A.l.* (right). Full arrows point to healthy seeds and arrow heads point to aborting seeds. (C) Left: Percentage of seeds with pollen tube burst failure in Col-0 × *A.a.* (n = 57 seeds) and Col-0 × *A.l.* (n = 202 seeds) growing at 22 °C. Right: Light microscopy images of chloral hydrate cleared seeds at 3 days after pollination (DAP) from Col-0 × *A.a.*/*A.l.* showing pollen tube burst failure. Scale bar = 50 µm.



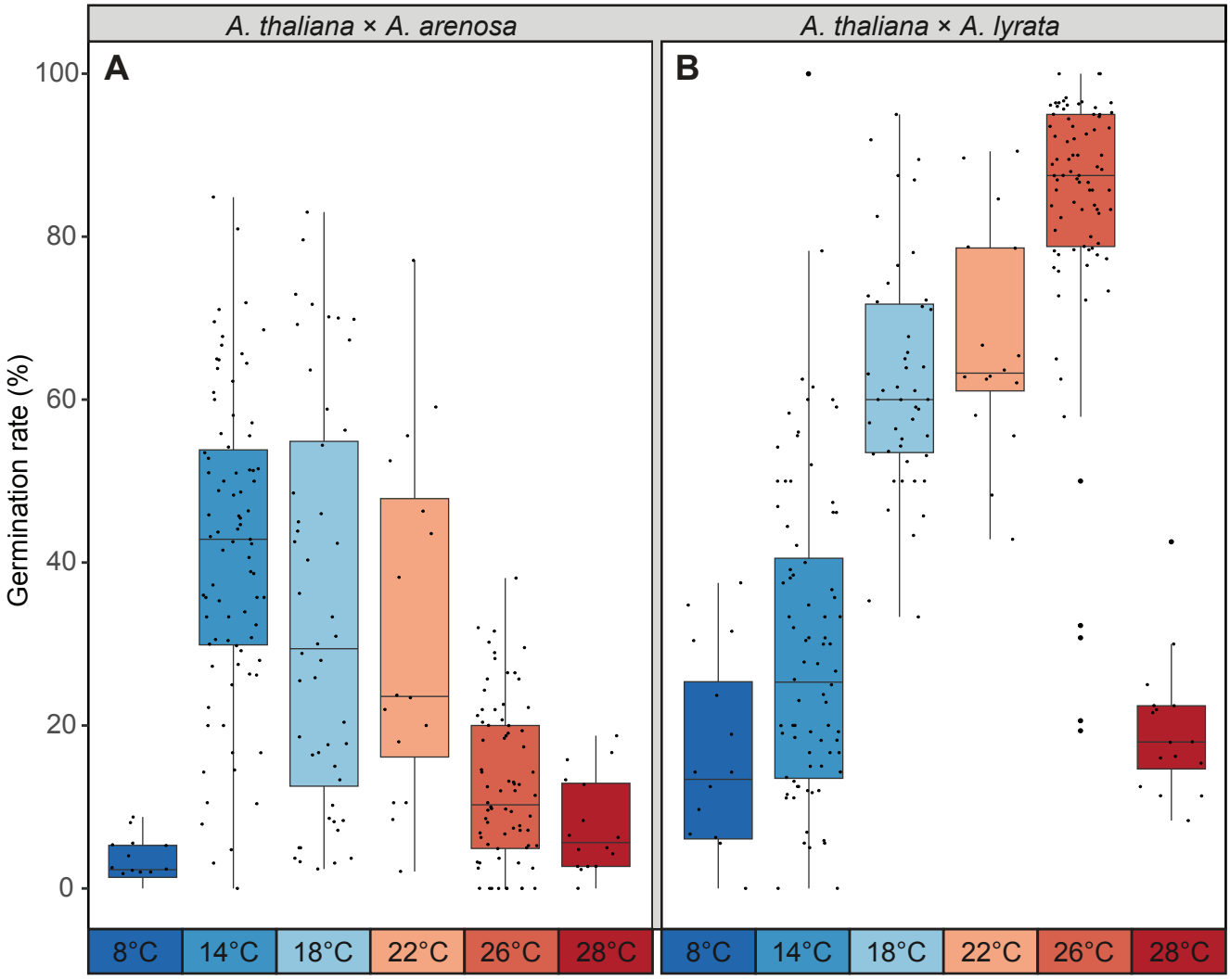
Supplementary Figure 2. Seed phenotypes in F1 hybrids. (A) Light microscopy images of chloral hydrate cleared seeds at 6 and 12 days after pollination (DAP) from A. thaliana (Col-0), Col-0 × A. arenosa (A.a.) and Col-0 × A. lyrata (A.l.). () Confocal micrographs showing endosperm cellularization of Feulgen-stained seeds at 3, 6 and 10 DAP from Col-0 × A.l. Frequency of endosperm cellularization and embryo stages is indicated by color code. Biological replicates (siliques): 3 DAP, n = 72; 6 DAP, n = 93; 10 DAP, n = 33; EM (embryo stages); ES (endosperm stages). Scale bar = 50 µm.



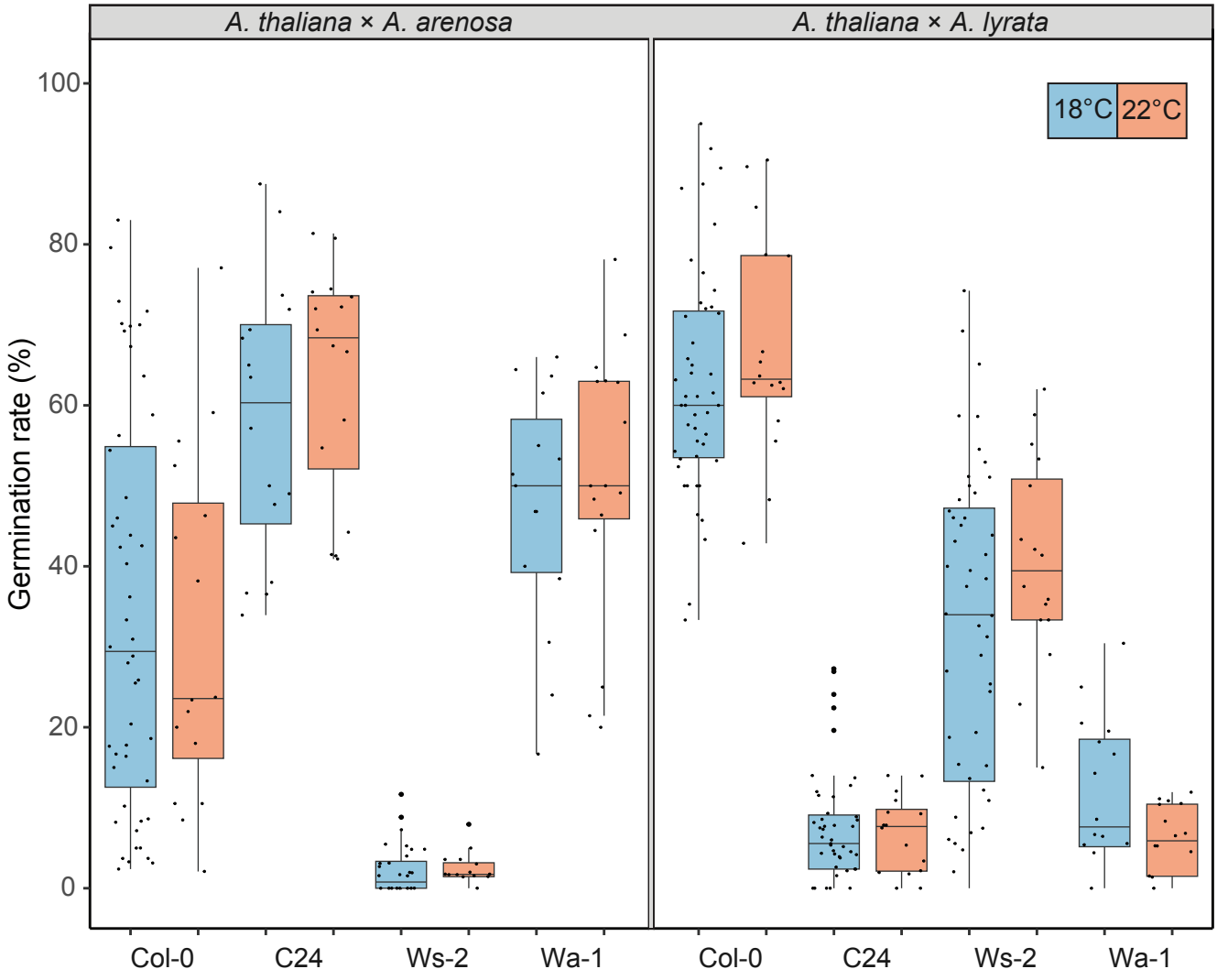
Supplementary Figure 3. Confocal micrographs of proAT5G09370>>H2A-GFP (EE-GFP) in seeds of *A. thaliana* (Col-0) and in hybrid seeds. Crosses of *A. thaliana* (Col-0) with *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) used as pollen donors from 1 to 15 days after pollination (DAP) are shown. Scale bar = 50 µm.



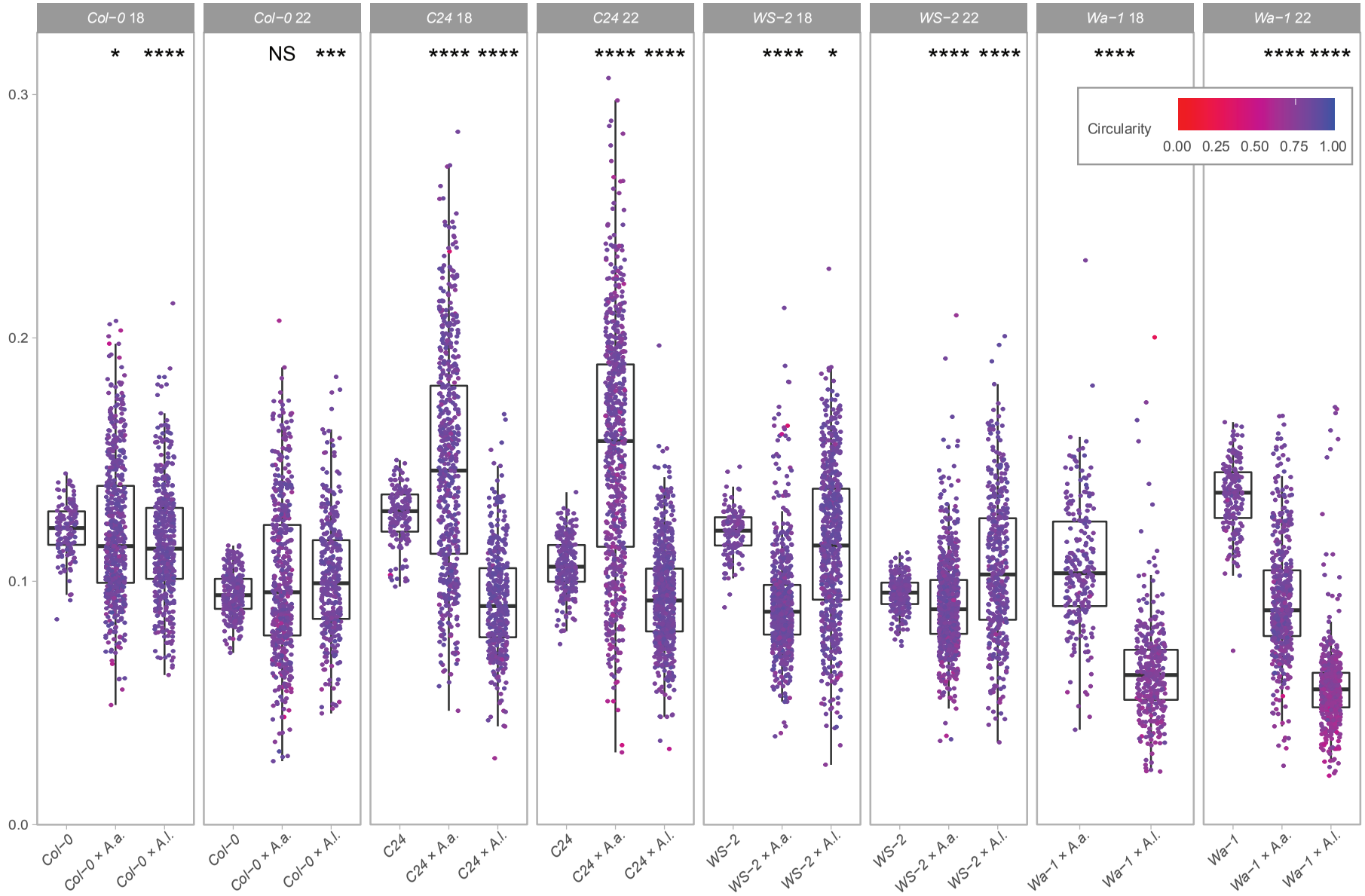
Supplementary Figure 4. Confocal micrographs of proAT4G00220>>H2A-GFP (TE1-GFP) in seeds of *A. thaliana* (Col-0) and in hybrid seeds. Crosses of *A. thaliana* (Col-0) with *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) used as pollen donors from 6 to 21 days after pollination (DAP) are shown. Scale bar = 50 μm.



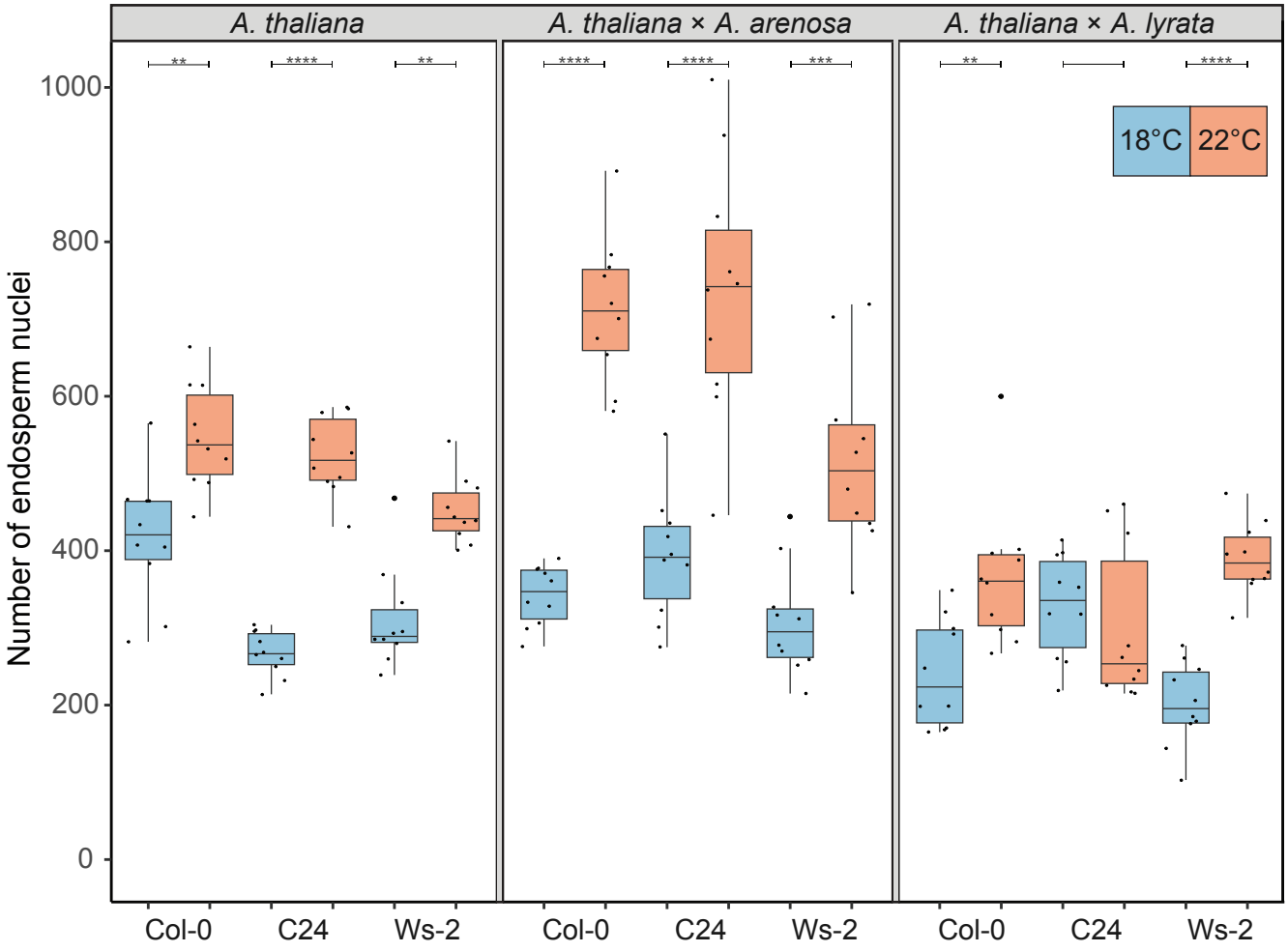
Supplementary Figure 5. Temperature affects germination rate of hybrid seeds. (A) Germination rate of *A. thaliana* (Col-0) × *A. arenosa* (*A.a.*) hybrid seeds at 8°C, 14°C, 18°C, 22°C, 26°C and 28°C. Biological replicates (siliques): 8°C, n = 16 (811 seeds); 14°C, n = 83 (3243 seeds); 18°C, n = 48 (2610 seeds); 22°C, n = 16 (796 seeds); 26°C, n = 78 (2713 seeds); 28°C, n = 16 (663 seeds). (B) Germination rate of *A. thaliana* (Col-0) × *A. lyrata* (*A.l.*) hybrid seeds at 8°C, 14°C, 18°C, 22°C, 26°C and 28°C. Biological replicates (siliques): 8°C, n = 16 (345 seeds); 14°C, n = 84 (1676 seeds); 18°C, n = 47 (1540 seeds); 22°C, n = 16 (533 seeds); 26°C, n = 85 (2043 seeds); 28°C, n = 16 (699 seeds). Box plot contains scattered data points representing germination rates observed per silique. Outliers are plotted as large data points.



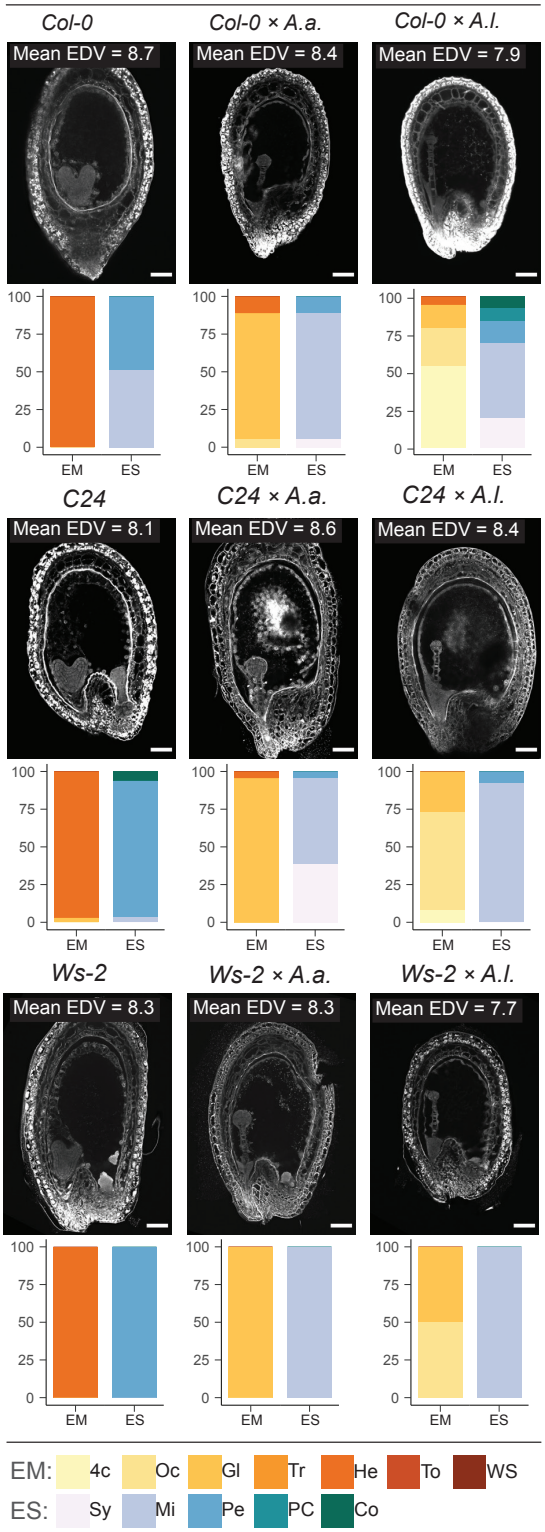
Supplementary Figure 6. Temperature effect on germination rate in hybrid seeds varies with *A. thaliana* accession. Germination rate of seeds from crossing *A. arenosa* (*A.a.*) / *A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (Col-0/C24/Ws-2/Wa-1) at 18°C and 22°C. Biological replicates (siliques): Col-0 × *A.a.* 18°C, n = 48 (2610 seeds); Col-0 × A.a. 22°C, n = 16 (533 seeds); C24 × *A.a.* 18°C, n = 16 (925 seeds); C24 × *A.a.* 22°C, n = 16 (805 seeds); Ws-2 × *A.a.* 18°C, n = 16 (987 seeds); Ws-2 × *A.a.* 22°C, n = 16 (947 seeds); Wa-1 × *A.a.* 18°C, n = 15 (546 seeds); Wa-1 × *A.a.* 22°C, n = 16 (805 seeds); Col-0 × *A.l.* 18°C, n = 47 (1540 seeds); Col-0 × *A.l.* 22°C, n = 16 (533 seeds); C24 × *A.l.* 18°C, n = 16 (801 seeds); C24 × *A.l.* 22°C, n = 16 (821 seeds); Ws-2 × *A.l.* 18°C, n = 16 (843 seeds); Ws-2 × *A.l.* 22°C, n = 16 (572 seeds); Wa-1 × *A.l.* 18°C, n = 16 (546 seeds); Wa-1 × *A.l.* 22°C, n = 16 (759 seeds). Box plot contains scattered data points representing germination rates observed per silique. Outliers are plotted as large data points.



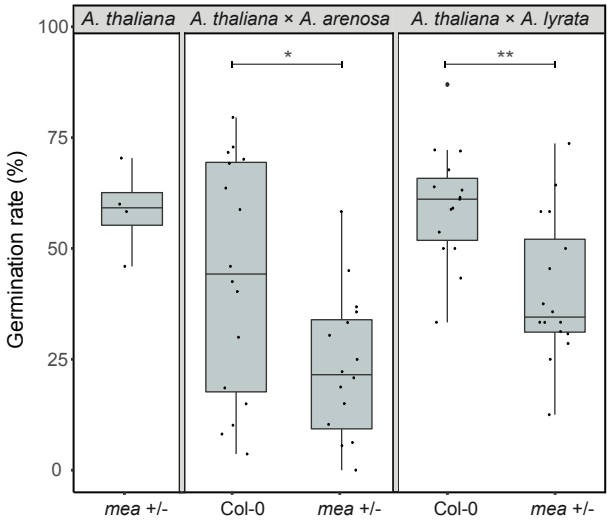
Supplementary Figure 7. Effect of temperature, *A. thaliana* accession and hybridization on seed size and circularity. Boxplot of seed size with circularity shown in a color gradient from 0 (low degree of circularity) to 1 (high degree of circularity) from A. thaliana (Col-0/C24/WS-2/Wa-1) self-crosses and from crossing *A. arenosa* (*A.a.*) / *A. lyrata* (*A.l.*) as pollen donor to Col-0/C24/WS-2/Wa-1. Col-0 18°C, n = 147; Col-0 x *A.a.* 18°C, n = 590; Col-0 x *A.l.* 18°C, n = 483; Col-0 22°C, n = 278; Col-0 x *A.a.* 22°C, n = 524; Col-0 x *A.l.* 22°C, n = 343; C24 18°C, n = 153; C24 x *A.a.* 18°C, n = 571; C24 x *A.l.* 18°C, n = 399; C24 22°C, n = 237; C24 x *A.a.* 22°C, n = 711; C24 x *A.l.* 22°C, n = 671; WS-2 18°C, n = 125; WS-2 x *A.a.* 18°C, n = 640; WS-2 x *A.l.* 18°C, n = 652; WS-2 22°C, n = 236; WS-2 x *A.a.* 22°C, n = 768; WS-2 x *A.l.* 22°C, n = 405. Significant differences in seed size are indicated for the comparisons between *A. thaliana* self-cross and accession crosses for each accession and temperature, except for Wa-1 at 18°C where the comparison was between Wa-1 × *A.a.* and Wa-1 × *A.l.* (Wilcoxon rank-sum test: NS $P > 0.05$; *$P \leq 0.05$; ***$P \leq 0.001$); ****$P \leq 0.0001$.



Supplementary Figure 8. Number of endosperm nuclei varies with temperature, accession and hybridization. Number of endosperm nuclei was counted in seeds (n = 10) from *A. thaliana* (Col-0/C24/Ws-2) self-crosses and from crossing *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) as pollen donor to *A. thaliana* (Col-0/C24/Ws-2). Significance is indicated for the comparisons of 18°C and 22°C between all accessions (Wilcoxon rank-sum test: **P ≤ 0.01; ***P ≤ 0.001; ****P ≤ 0.0001). Outliers are plotted as large data points.



Supplementary Figure 9. Effect of temperature, accession and hybridization on endosperm development. Confocal images showing endosperm cellularization of Feulgen-stained seeds from *A. arenosa* (*A.a.*) or *A. lyrata* (*A.l.*) crossed as pollen donor to *A. thaliana* (accession Col-0 or C24 or Ws-2) at 18°C. Scale bar = 50 µm. Mean endosperm division value (EDV) is shown within each image, nEDV = 10 seeds. Quantification of the described embryo and endosperm stages in each cross are shown as bar charts: Col-0, n = 37; Col-0 × *A.a.*, n = 18; Col-0 × *A.l.*, n = 93; C24, n = 32; C24 × *A.a.*, n = 23; C24 × *A.l.*, n = 37; WS-2, n = 23; WS-2 × *A.a.*, n = 16; WS-2 × *A.l.*, n = 12. Embryo stages (EM): 4c: 4-cell, Oc: Octant, Gl: Globular, Tr: Transition, He: Heart, To: Torpedo, WS: Walking stick. Endosperm cellularization stages (ES): Sy: Syncytial, Mi: Micropylar, Pe: Peripheral, PC: Partially complete, Co: Complete.



Supporting Information Fig. S10: Germination rate of *A. thaliana mea* and WT hybrid seeds from crosses with *A. arenosa* or *A. lyrata.* The x-axis indicates the genotype of the *A. thaliana* cross partner. Crossing *mea* to *A. arenosa* or *A. lyrata* results in a significant decrease in seed survival when compared to Col-0 crossed to *A. arenosa* or *A. lyrata*. The *mea* mutant crossed to self results in 60% viable seeds, indicating that the crosses to *A. arenosa* and *A. lyrata* do not have an impact on the seed survival compared to the Col-0 background. Welch's t-test: *P ≤ 0.05; **P ≤ 0.01. Outliers are plotted as large data points.

III

Low parental conflict, no endosperm hybrid barriers, and maternal bias in genomic imprinting in selfing *Draba* species

Renate M. Alling[1,2], Katrine N. Bjerkan[1,2], Jonathan Bramsiepe[1], Michael D. Nowak[3], A. Lovisa S. Gustafsson[3], Christian Brochmann[3], Anne K. Brysting[1,2] and Paul E. Grini[1,4]

[1] Section for Genetics and Evolutionary Biology, University of Oslo, 0316 Oslo, Norway.

[2] CEES, Department of Biosciences, University of Oslo, 0316 Oslo, Norway.

[3] Natural History Museum, University of Oslo, 0318 Oslo, Norway

[4] Corresponding author, paul.grini@ibv.uio.no

Author Contributions: P.E.G., K.N.B, R.M.A, & A.K.B. designed the research; K.N.B, J.B. & R.M.A performed the experiments; P.E.G., K.N.B, R.M.A, & A.K.B. analyzed and discussed the data; P.E.G., K.N.B, R.M.A, & A.K.B. wrote the article; All authors revised and approved the article.

Key words: Mating systems, Genomic imprinting, hybridization barriers, arctic, *Draba*, endosperm