Transcriptomic, Small RNA, and epigenetic atlas of the *Brassica napus* seed

By

Dylan Ziegler

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Department of Biological Sciences

University of Manitoba

Winnipeg, MB, Canada

Acknowledgements

I would first like to acknowledge my primary supervisor, Dr. Mark Belmonte, for all his feedback, assistance, patience, and support he has given me over the years. I am truly grateful to have had such an excellent mentor. Thank you for taking me in as a PhD student, and for helping me grow academically, professionally, and personally. I would also like to thank Dr. Bruce Ford and Dr. Diana Bizecki-Robson for their support of my academic growth in the herbaria and collections in Manitoba, and their willingness to share their knowledge and experience with me. Thank you as well to Diana Sawatzky for their support of my endeavours in the WIN herbarium throughout my graduate research. Further, I would like to thank Dr. Az Klymiuk for their academic guidance and support of me, and for sharing their valuable and deep knowledge. Thank you all for supporting me to become the botanist I want to be.

I would like to thank my previous lab members for their company, assistance, and shared experience, including but certainly not limited to: Dr. Philip Walker, Dr. Nick Wytinck, Joey Wan, Dr. Deirdre Khan, Nadège Pulgar-Vidal, Bliss Beernink, Natalie Cale, Shayna Giesbrecht, and Rachel Robinson. I would like to additionally express my gratitude to Dr. Deirdre Khan, whom I am grateful to have established (what I hope to be) a lifelong friendship. You have made me a better scientist, academic, and person. Thank you everyone.

My committee members, Dr. Claudio Stasolla, Dr. Steve Whyard, and Dr. Olivia Wilkins have graciously donated their knowledge, expertise, and time to my thesis work and research endeavours. I am grateful for all your insights, feedback, and criticism to help me grow as an academic. Thank you for all the time you have invested in me.

Potentially most of all, thank you to my loving family; my sister Amanda, my mother, my brother-in-law Connor, and my step-aunt Tammy for their support and faith in me as a person and as an academic. Thank you to my supportive friends, especially Lucas Hampel, Nicholas Piroddi, and Johanna Dansereau for being such an optimistic presence in my life. Thank you to my charming dog, Abby, for tolerating all my late nights and all our flights and moves across the country. To my father, whom this dissertation is dedicated to, I wish it was finished in time for you to have seen — may you rest in peace as the strongest, gentlest man I will ever know.

Dedication

This dissertation is dedicated to my father
who gave me all the love he had to give,
and always encouraged my curiosity of the
natural world. Thank you for everything —
love you to the moon and back, always.

Table of contents

Contents

Thesis Abstract

The diversification of flowering plants has been described as the "abominable mystery" of evolutionary biology for decades. Angiosperms diversified at unprecedented rates in the early Cretaceous, rapidly occupying new niches and pervading much of the earth's terrestrial landscape. This speciation was accompanied by dramatic genome evolution as well — more than 75% of extant angiosperm species are believed to have originated from genome duplication events. These duplications are most often in the form of endogenous chromosomal duplication (autopolyploidy) or, more frequently, via interspecific hybridization (allopolyploidy). Consequently, dynamics of polyploidy within flowering plants are of interest to the fields of plant genetics and evolution. One of the world's most important oilseeds is *Brassica napus* L. ("canola"), an allopolyploid species derived from two progenitors (*B. rapa* and *B. oleracea*). *B. napus* has two relatively complete diploid genomes from each progenitor, but also shares a whole genome triplication (autopolyploidization) with the rest of the Brassicaceae. These factors constitute *B. napus* as an informative model to study the consequences of polyploidization. Within this thesis, I describe the transcriptional and epigenetic profiles throughout *B. napus* seed development. Therein, I utilize laser microdissection, histological analysis, and bioinformatic tools to describe the transcriptional profiles of the seed. The subgenomes constituting *B. napus* contribute to gene expression within the seed asymmetrically, with many homlogs being more expressed in one subgenome than the other. Additionally, I provide evidence for striking differences in epigenetic structure of the fractionated genomes depending on genomic context. The works detailed within this thesis serve to explain the transcriptional landscapes of the *B. napus* seed, particularly through the lens of polyploidy, and the exceptional complexity of angiosperm genomes.

1 Literature Review

1.1 Seed Development in the Brassicaceae

1.1.1 Overview of *Brassica* seed development

Of all the agriculturally relevant crops, few can compare to the diversity achieved by the *Brassica* genus. The various subspecies and varieties within the *Brassica* exhibit tremendous variation both vegetatively and reproductively and are some of the most prevalent food crops in the world. *Brassica's* robust diversity is represented best within *Brassica oleracea*, which accounts for many of the crops grown in the Brassicaceae, including cauliflower, brussels sprouts, broccoli, cabbage, kohlrabi, kale, and collard greens (Maggioni et al., 2018). Canola (*Brassica napus* L.) is the second most economically important oilseed crop in the world (FAOSTAT 2018), and is particularly important in Canada, contributing 30 billion dollars to the Canadian economy and generating 207,000 jobs (Canola Council of Canada 2019). The cultivated form of *B. napus*, canola, was developed in part at the University of Manitoba as an oilseed with low erucic acid and low glucosinolate content which helped to cement canola's global importance as an oilseed. Despite the renowned relevance of this crop, the development of the seed is understudied at the anatomical and genetic level. Moreover, *B. napus* is an allopolyploid formed by the hybridization of *B. rapa* (A^rA^r) and *B. oleracea* (C^oC^o) approximately 7,500-12,500 years ago (Chalhoub et al., 2014a). The result was an amphidiploid, a species which contains both constituent diploid genomes, effectively behaving as (2+2)n ploidy. *B. napus* ($A^nA^nC^nC^n$) is therefore a globally important oilseed in a diverse plant family with relatively few studies exploring the genetic regulation underpinning the seed's development, especially through the lens of polyploidy.

Seed development is a complex and highly orchestrated process, requiring tight coordination between gene expression and anatomical modification to properly organize, nourish, and protect the embryo. The Brassicaceae bears bitegmic (two integuments), campylotropous (with the chalaza and micropyle adjacent to one another), tenuinucellate (single epidermal nucellus layer surrounding the embryo sac) ovules (OV) adjoined to the nucellar epidermis at the chalazal pole (Figure 1.1) (Wilkinson et al., 2019). Post-fertilization, the proliferation of the outer integument places the micropyle and the chalaza to be adjacent to one another, shifting to an amphitropous position (Schneitz et al., 1995; Skinner et al., 2016). The embryo proliferates and develops in two distinct phases, 1) morphogenesis, and 2) maturation. Morphogenesis is initiated when the zygote first undergoes an asymmetric division forming the basal cell and the apical cell. The larger basal cell is embedded within the micropylar pole and is the progenitor cell of the suspensor (Boscá et al., 2011a). The smaller apical cell is cytoplasmically dense and divides to form the embryo proper (EP). The apical cell is localized to the distal end of the basal cell, and proceeds through morphogenesis with equatorial divisions forming the globular embryo at the 32-cell stage. The EP organizes cell initials in the globular stage (GLOB) to transition into the heart stage (HEART) (**Figure 1.1**). During this transition, the cells at the apex of the EP divide periclinally to form the initials of the cotyledons. The heart stage is characterized by the delineation of the cotyledon primordium, the shoot apical meristem, the root apical meristem, as well as the pith rib meristems delegating the position of the vasculature – by the end of the heart stage the EP has organized the entirety of the plant's tissues. The heart phase is followed by the linear cotyledon stage, which is the transitional phase of seed development connecting morphogenesis and maturation (Belmonte et al., 2013a). Maturation begins once the EP elongates to fully occupy the volume of the seed and deposition of storage lipids and proteins initiates – characterizing the mature green stage (MG).

Once the accumulation of storage reserves completes, the seed concludes maturation in the dry seed stage (DS), wherein the seed desiccates and stores transcripts vital to germination (Nelson et al., 2017). These processes necessitate strict control over both the maternal and filial components of the seed, which adds further questions as to how seeds achieve this organization and coordination – especially in species with extensive storage reserves, as in oilseeds.

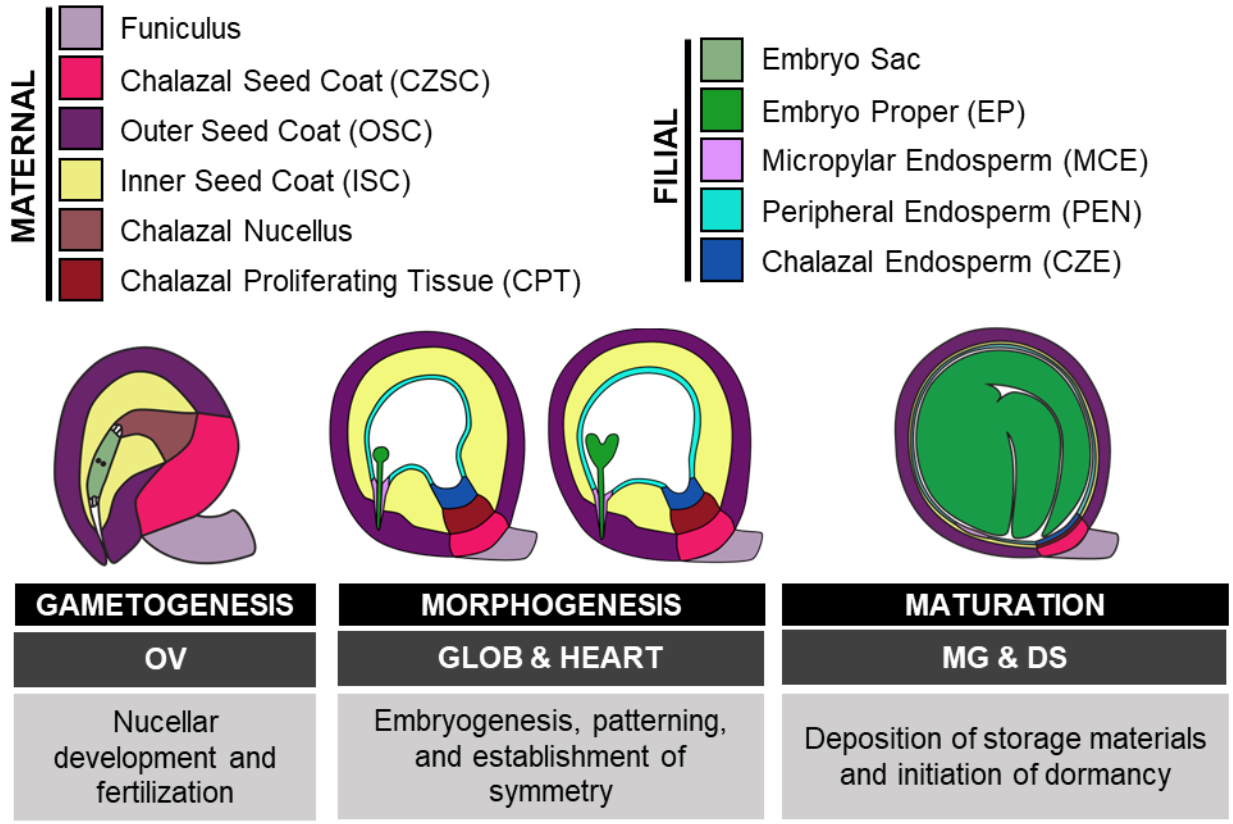


Figure 1.1. Seed development of *Brassica napus*. Seed development is initiated by gametogenesis, when chalazal nucellus is subtending the newly formed embryo sac. Post-fertilization, the EP enters morphogenesis and the CPT proliferates intramarginally to the CZE and CZSC. The ISC and OSC elongate and fuse at the micropyle immediately following fertilization, sealing the EP off from the outside. The MCE, PEN, and CZE form a continuous syncytium during morphogenesis. Gradual cellularization occurs throughout seed development, and by maturation the MCE and PEN have cellularized. After the deposition of storage materials has finished during maturation the seed enters dormancy.

1.1.2 Subregions of the seed

The genetic diversity of the seed is one of its most intriguing yet elegant features – with three highly distinct genetic identities: i) maternal, ii) endosperm, and iii) filial. How the seed accommodates this complexity is one of the overarching questions in seed development research. Furthermore, many regions of the seed are distinct anatomically, furthering the spatial complexity of the structure. In the Brassicaceae, seeds can be divided into the maternal subregions including the inner seed coat (ISC), outer seed coat (OSC), chalazal seed coat (CZSC), and the chalazal proliferating tissue (CPT) and filial subregions including the micropylar endosperm (MCE), peripheral endosperm (PEN), chalazal endosperm (CZE), and the embryo proper (EP) (**Figure 1.1.**). Each of these subregions have been shown to contribute uniquely to development in Arabidopsis (Belmonte et al., 2013a).

The integuments of flowering plant ovules fuse to form the contiguous seed coat post-fertilization. Despite the similarity between the two integuments, they operate independently and the asymmetric growth of the two integuments is integral to seed development. In the Brassicaceae, the outer integument proliferates to envelop the chalaza and cover the embryo sac, subtending the developing inner integument. Both the inner and outer integument operate autonomously from one another, requiring different developmental programs to their elongation in ovule development (Coen et al., 2017; Simon et al., 2017). Furthermore, the CZSC, which acts as the receptacle for the nucellus in early development, has distinct morphology at the cellular and ultrastructural levels (Millar et al., 2015a). The CZSC is adjacent to the funiculus and, in *Brassica napus,* extensive plasmodesmatal networks connect the surrounding maternal tissues enhancing both apoplastic and symplastic mobilization to adjacent subregions. Additionally, the subregion has abundant endomembrane systems to facilitate synthesis of

storage products as well as communication. The characteristic anatomy of the CZSC reaffirms its independence from the rest of the seed coat and supports its role as a vital transport vector between the funiculus and the developing seed. Distal to the CZSC lies the CPT, located intramarginally to the CZE and CZSC. The precise progenitor cells of the CPT are not well known, and its function has not been characterized. The morphology of the CPT varies greatly in the Brassicaceae, being diminutive in the Camelineae (including *Capsella* and *Arabidopsis*), while being comparatively massive in both the Thelypodieae (*Stanleya*) and most importantly, the oilseeds of the Brassiceae (*Brassica*) (Schulz and Jensen, 1971; Brown et al., 2004a). Despite the apparent complexity and importance of this subregion, its function is largely unknown.

The endosperm subregions and the EP comprise the filial regions and are genetically distinct. In angiosperms, the endosperm forms from a double fertilization event, involving sperm nuclei from the pollen grain fusing with both the egg cell and the polar nuclei within the embryo sac (Zhang et al., 2020). Within the Brassicaceae, one haploid sperm nucleus fuses with two haploid polar nuclei, forming a triploid endosperm progenitor cell (Brown et al., 1999). The endosperm then proliferates through karyokinesis without cytokinesis, forming a syncytium that wholly encircles the embryo. As the ISC and OSC elongate to the amphitropous position, the endosperm adheres to the inner surface of the ISC and CPT. At this point, the endosperm can be described as three distinct subregions: the MCE, PEN, and CZE. The MCE is the region adhered to the micropylar pole, and more specifically tightly associated with the suspensor adjoined to the EP in early development. The PEN comprises most of the endosperm, covering the periphery of the seed's ISC. Finally, the CZE which has been described as a cyst at the chalazal pole is directly adjacent to the CPT. While the endosperm is initially a syncytium, the entirety of all three subregions cellularize as the seed matures. Cellularization begins at the

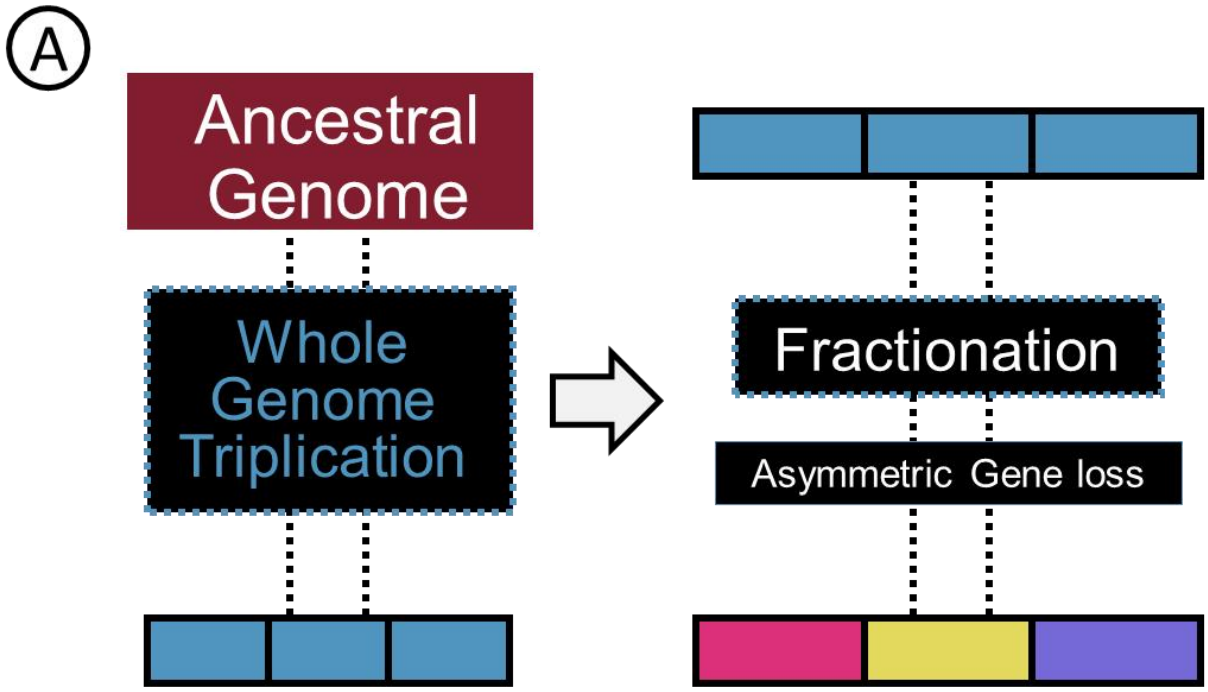
micropylar pole, and proceeds in a wave-like motion towards the chalazal pole. By consequence, the CZE cellularizes late in development (Mansfield and Briarty, 1990; Batista et al., 2019a). The subregions of the endosperm appear to show a degree of compartmentalization, as Belmonte et al., 2013 reported, the transcriptome of the CZE is distinct from that of the MCE and PEN, which both share more transcriptional similarity with the EP and suspensor than they do with the CZE. Furthermore, the CZE was shown to have the most diverse set of subregion-enriched genes in early seed development. Altogether, the chalazal end of the seed warrants more investigation into its functions coordinating seed development.

1.1.3 Anatomical diversity of the seeds of the Brassicaceae

Considerable evolutionary distance separates the *Brassica* genus from *Arabidopsis*, and by extension, there is notable diversity in seed anatomy across the breadth of the Brassicaceae. The Brassicaceae is sister to the Cleomaceae and is comprised of approximately 44 tribes, containing 307 ranked genera and 3600 species. These tribes are most easily divided by lineages; as outlined in Franzke et al., 2011, that describe three primary lineages within the family of varying taxonomic resolution. Lineages I and III diverge earlier than Lineage II, and while *Brassica* is ranked within Lineage II, *Arabidopsis* is currently assigned within Lineage I. This distance necessitates consideration when interpreting and comparing datasets across species in the Brassicaceae.

The subregions discussed prior are differ anatomically and developmentally across the Brassicaceae. In particular, the chalazal pole of the seed shows markedly different development in tribes belonging to Lineage I and III than in Lineage II (Figure 1.2). *Lepidium* and *Coronopus* (Lepideae, Lineage I) as well as *Arabidopsis* and *Capsella* (Camelineae, Lineage I) all demonstrate similar chalazal anatomy, with a smaller CZSC and a diminutive CPT consisting of only a few

cells, which are morphologically indistinct from the incipient chalazal nucellus (Nguyen et al., 2000a; Brown et al., 2004a). A scant CPT is also seen in *Erysimum* (Hesperideae, Lineage III), and is compounded with a compressed chalazal chamber and a dwarfed CZSC. Interestingly, while the CPT demonstrates this diversity across lineages, the CZE forms a cyst in almost all documented species in the Brassicaceae, even within the condensed chalazal chamber of *Erysimum* (Brown et al., 2004b). Conversely, *Brassica* species (Brassiceae, Lineage II) demonstrate a highly specialized CPT, with numerous layers of elongate fusiform cells differentiating them from the chalazal nucellus. *B. rapa, B. oleracea,* and *B. napus* all demonstrate this unique morphology (Brown et al., 2004a; Millar et al., 2015a). These species are the most relevant crops within the Brassicaceae, and as such investigating and interpreting the functions of the chalazal subregions is of high priority.


A
Ancestral Genome
Whole Genome Triplication
Fractionation
Asymmetric Gene loss

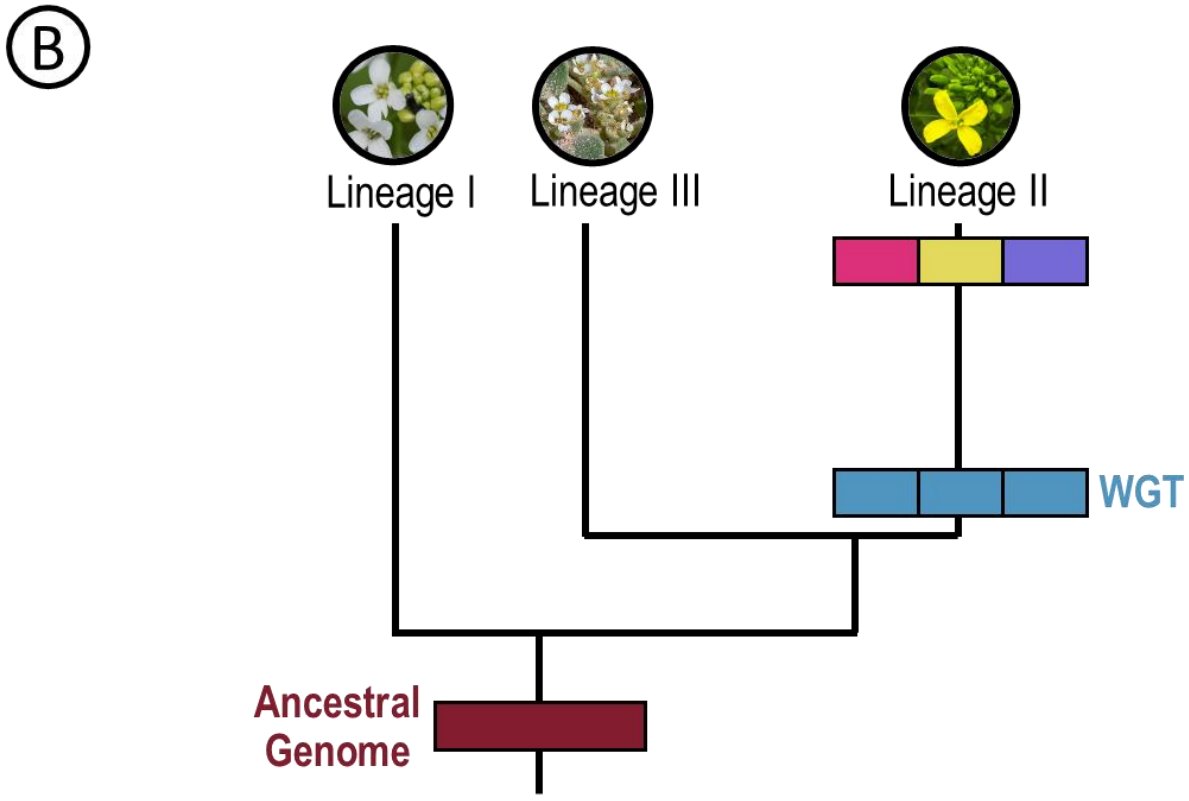

B
Lineage I
Lineage III
Lineage II
WGT
Ancestral Genome

Figure 1.2. (A) Hypothetical model of genome evolution post-WGT. Ancestral genomes undergoing triplication will frequently experience subsequent asymmetric gene loss, creating distinctive, fractionated genomes. **(B)** In the Brassicaceae, Lineage II (containing the Brassiceae) a whole genome triplication (WGT) occurred which has lead to the specific genome fractionation within *Brassica. (B) is a simplified version of the phylogeny presented in Franzke et al., 2011.*

1.2 Polyploidy in the Brassicaceae

1.2.1 Genomic architecture of the *Brassica* genus

The *Brassica* genus is founded on several ancestral genome duplications. Excluding the whole genome duplications (WGDs) documented within the dicot lineage, the *Brassica* genus has two unique WGDs. The first occurred approximately 34 million years ago prior to the divergence of Lineages I, II, and III, and is thus shared by the entirety of the core Brassicaceae (the At-α WGD event) (Franzke et al., 2011a). A unique whole genome triplication (WGT) followed At-α within Lineage II and is specific to the Brassiceae tribe (Br-α) (Lysak et al., 2005). Thus, the *Brassica* genus has highly convoluted genomic structure, and the emergence of allopolyploids within the *Brassica* genus complicates this further. Moreover, the WGT occurred millions of years ago, and as a result all three of the triplicated genomes have diverged from one another (Figure 1.2A) (Liu et al., 2014a; Parkin et al., 2014). This divergence is broadly defined by the retention (or lack thereof) of syntenic regions relative to the ancestral Brassicaceae genome using *Arabidopsis*, (Camelineae, Lineage I) as the reference. The triplicated, fractionated genomes only exist within the Brassiceae tribe, of Lineage II (Figure 1.2B). *B. napus* represent the union of three fractionated genomes and presents a unique opportunity to investigate both the effect of recent polyploidization on the genetics underpinning reproductive development, and the dynamics of the fractionated genomes within a nascent amphidiploid.

In addition to *Arabidopsis'* ancestral polyploidizations, *B. napus* possesses an overall sixfold duplication from Br-α (3x) as well as the allotetraploid hybridization (2x). Further, it is

known that the *B. oleracea* genome has followed a starkly different evolutionary trajectory than its sister species, *B. rapa*. While substantial gene loss in *Brassica* predates the divergence of *B. oleracea* from *B. rapa*, numerous chromosomal translocations characterize the *B. oleracea* genome (Liu et al., 2014b). This contrasts with the findings in other lineages, as in *Solanum lycopersicum* and *Solanum tuberosum* (Solanaceae) as well as between *Pyrus bretscheideri* and *Malus* x *domestica* (Rosaceae) (Sato et al., 2012; Wu et al., 2013), wherein the sister genomes show a tremendous degree of conservation in genome structure. *B. oleracea* contains nine chromosomal pairs and is ~630 Mb in size and *B. rapa* has ten chromosomal pairs and is ~530 Mb (Chalhoub et al., 2014b), and while syntenic genes between the two sister species seem to be conserved, genomic structure is not. Transposable element (TE) accumulation is noticeably higher in *B. oleracea* than in *B. rapa,* with TEs in the former being 3.4x greater in length (Liu et al., 2014b). While genome fractionation of polyploids leading to diploidization is common in plants, *B. oleracea's* non-collinear gene sets demonstrate substantial retention relative to *B. rapa* as well. This indicates that diploidization in *B. rapa* has been more severe than that of *B. oleracea,* which has retained more of the ancestral Br-α event. *B. napus* is the allotetraploid culmination of these two species, with a total haploid genome size of ~1130 Mb (Chalhoub et al., 2014b).

1.2.2 Genome duplications in land plants

Genome duplications are now understood to be a major driver in land plant evolution and have occurred over many time periods in many clades (Qiao et al., 2019). Genome duplications broadly arise via hybridization or aneuploidy in meiosis, passing either extra endogenous chromosomal pairs (autopolyploidy) or exogenous chromosomes from a related species (allopolyploidy) to the progeny. It is currently estimated that >70% of extant angiosperms and monilophytes are polyploids, many with several genome duplications

preceding them (Otto and Whitton, 2000b; Stuessy and Weiss-Schneeweiss, 2019). While the details of the advantages of polyploidy are largely unclear, it appears that some of the most diverse lineages within the angiosperms, including the Fabaceae, Rosaceae, Poaceae, Asteraceae, and Orchidaceae (Amich et al., 2007; Huang et al., 2016; Wang et al., 2019b) may owe their diversity to ancient polyploidization events. Larger scale analyses to test the association between polyploidy and adaptive radiation mostly confirm this tight association, with some exceptions (Landis et al., 2018). Notably, it seems that while polyploidization events are only loosely correlated with a subsequent expansion in diversity, this correlation is augmented in geologic times where there was environmental instability. These results likely indicate that while polyploidy does not drive diversification independently, polyploidizations enable niche expansion followed by divergent evolution when environmental niches expand and/or are more numerous. It is possible that this additionally applies to crop breeding and artificial selection, wherein polyploid plants make ideal crops for their capacity to acquire new characteristics and speciate (Baduel et al., 2019). This is especially pertinent to *Brassica* given that *B. oleracea* exhibits such remarkable structural diversity and has diversified into as many food crops as it has.

Genome duplications are ubiquitous in ancient angiosperms lineages especially, and have influenced all of the extant angiosperms. Polyploidization events have been documented in all important transition points within angiosperm evolution; within the Magnoliids (basal angiosperms), the Chloranthales (lineage directly preceding the monocot and dicot divergence) the Ceratophyllales (lineage preceding dicot divergence), and the Vitales (basal Rosid) (Jaillon et al., 2007; Ren et al., 2018) as well as within the Brassicaceae specifically (Franzke et al., 2011; Chalhoub et al., 2014). WGDs undergo extensive deletions over time, however, and genes that

are products of polyploidization are often lost exponentially with an estimated half-life of 21.6 million years (Ren et al., 2018). This is particularly relevant in the context of the Brassiceae tribe, a high lineage within the crucifer family that is defined by a whole genome triplication (WGT), though the timing of this event is still unclear (Franzke et al., 2011). *B. napus* is of particular interest as a recently formed allopolyploid that is comprised of two species which experienced the WGT, and as such has hypothetically sixfold duplication of genetic material compared to lower lineages in the Brassicaceae, such as *Arabidopsis* (Camelineae). The dynamic changes within the genomes of angiosperms are as captivating as they are bewildering, and unravelling that complexity is the crux of genome biology.

1.2.3 Genome fractionation and bias in allopolyploids

Due to hybridization, allopolyploid plants contain sets of chromosomes from separate progenitors thus forming a new species. Crop plants have particularly duplicated genomes from polyploidizations and/or interspecific hybridizations (Renny-Byfield and Wendel, 2014). *Triticum aestivum* is an extreme example of this, being an allohexaploid with three distinct subgenomes. *T. aestivum* was formed by the hybridization of *T. urartu* and an either undescribed or extinct relative of *Aegilops speltoides*, forming the tetraploid *Triticum turgidum*. *T. turgidum* hybridized many years later with *Aegilops tauschii* to form the allohexploid wheat, *T. aestivum* (Gardiner et al., 2015). Allopolyploid genomes, especially ones recently formed, are tremendously noisy; they have the transcriptome and epigenetic structures of multiple genomes within one nucleus, and consequently regulatory elements from one genome can theoretically affect the other and vice versa. This biological phenomenon is a marvel of organization and speaks to the resilience of plant genomes, but how these genomes interact both transcriptionally and epigenetically is mostly unknown.

In the long term, allopolyploid plants may undergo dynamic changes in the balance between the subgenomes. These changes typically arise from genome bias – where one genome exhibits higher transcriptional activity than the other(s). A subgenome experiencing greater expression of homologs in one genome than the other(s) is dominant, as the machinery of the subgenome becomes more necessary for survival. As a result, the other genomes, by nature of their machinery having less utility, undergo extensive deletions and are thus fractionated. Genome fractionation is known to have played a part in the evolution of *Zea mays* but seems to be less influential in palaeo-autopolyploids like *Populus trichocarpa* (Liu et al., 2017; Zhao et al., 2017). Interestingly, it is also known that unbiased subgenome evolution exists – where fractionation occurs, but is relatively even between subgenomes, as in the palaeo-polyploid *Pyrus bretschneideri* (Li et al., 2019).

DNA methylation architecture seems to be vulnerable to subgenome bias and follows the patterns of gene expression bias within the species (Gardiner et al., 2015; Li et al., 2019). Further, retention of DNA methylation patterns is documented in *T. aestivum,* wherein the progenitor species (*Ae. tauschii*) displays highly conserved methylation patterns as its corresponding D subgenome does in *T. aestivum*. Small RNAs are regulators of both transcription and the epigenome (Borges and Martienssen, 2015; Tang et al., 2022), and are even more infrequently studied in the context of polyploidy than DNA methylation. It is unknown how genome bias or lack thereof arises, how the epigenome and its regulators establish genomic dominance, and how the balance of the bias is established.

1.3 Transcriptional regulation of the *Brassica napus* seed

1.3.1 Transcriptional networks governing seed development

Transcription factors (TFs) are some of the most important mediators of seed development, regulating vital developmental programs required to make a healthy seed. Perhaps one of the most important regulators in seed development is that of B3 Domain TFs: LEAFY COTYLEDON (LEC1/2), FUSCA3 (FUS3), and ABSCISIC ACID INSENSITIVE3 (ABI3). LEC1 is a subunit of the nuclear factor-Y complex and governs activation of the shoot apical meristem and postgerminative development (West et al., 1994). LEC1 ectopic expression incites developmental transitions in embryonic cells and can cause abnormal proliferation of the suspensor (Lotan et al., 1998). Further, in double knockdown mutants of LEC1 and ABI3 or FUS3 the suspensor proliferates into a second embryo, causing polyembryony (Vermon and Meinke, 1994). Similarly, both *abi3* and *fus3* mutants prematurely enter postgerminative developmental programs in morphogenesis, which implies that FUS3 and ABI3 plays a pivotal role in the suppression of germination (Meinke, 1992; Nambara et al., 2000; Jo et al., 2019). FUS3 is also known to bind to miR156 gene loci, which is of particular interest given miR156's integral role in vegetative phase transitions (Nodine and Bartel, 2010; Wang and Perry, 2013). This indicates that FUS3 is an important regulator of the miR156 transitional switch. Further, some of these TFs have been documented to be important in regulating seed storage proteins.

As central regulators of seed development, LEC1/2, ABI3, and FUS3 are also heavily implicated in maturation. The repressive effect these TFs have on postgerminative programs carries over to maturation, where the maturing embryo ceases deposition of storage reserves and initiates premature germination in *lec1* and *fus3* plants (Harada, 2001), and in *abi3* plants the embryo fails to degreen and complete dessication (Nambara et al., 1995; Delmas et al., 2013).

LEC1 is also known to maintain embryonic identity of the cotyledons, wherein loss of function mutants display adaxial-abaxial polarity and bear trichomes on the cotyledons, as in true leaves (West et al., 1994; Harada, 2001). In addition to this, *LEC1*-overexpression mutants accumulate mRNAs encoding proteins involved in fatty acid biosynthesis (Mu et al., 2008). Moreover, *LEC1* depends in part on the activity of *ABI3*, *FUS3*, and *WRINKLED1 (WRI1)* in regulation of fatty acid biosynthesis during maturation. LEC2 is also implicated in maturation programs, with ectopic expression inducing nutrient and lipid deposition in the embryo precociously and is additionally known to activate auxin biosynthesis to promote maturation (Kroj et al., 2003; Stone et al., 2008). Furthermore, ABI1 and LEC2 are established to be involved in the process of seed storage deposition in late seed development. For example, LEC2 is required for the expression of OLEOSIN genes, which are encode proteins that facilitate and generate large oil bodies in seed storage (Che et al., 2009). ABI3 regulates the expression of *CRUCIFERIN 1* and *CRUCIFERIN 3* (*CRU1/3*), which are genes involved in protein body deposition (Bedi et al., 2016). As one of the world's most important seeds, *B. napus* is rich in both healthy oils and proteinaceous seed meal, seed storage mechanisms are core to its development and identity as a crop. It is currently unknown if these seed storage genes experience genomic bias or specificity for certain subregions in *B. napus*. Theoretically, genes from the dominant genome would be more expressed in seed development and would likely be the most efficient targets for biotechnological applications in bolstering oil or protein production.

1.3.2 Global gene activity in the subregions of the seed

Most transcriptional analyses of seed development are done with whole seed tissue, which loses the potential resolution gained from investigating the subregions separately. To date, the most comprehensive analysis of subregions in a developing seed are documented in

Arabidopsis (Belmonte et al., 2013). This work documents the discrete transcriptomes of the maternal (ISC, OSC, CZSC, CPT) subregions relative to that of their filial counterparts (EP, CZE, PEN, MCE). Despite arising from the same double fertilization event, the endosperm and EP are transcriptionally distinct from each other, likely due to their independent development following fertilization. As expected, the EP is enriched for genes associated with embryonic patterning, such as the establishment of symmetry and leaf polarity (Belmonte et al., 2013). The CZSC was shown to have selective enrichment for trehalose biosynthesis, which implies the segregation of a pathway vital to the maintenance of embryonic health (Gómez et al., 2006). The PEN shows particularly biased accumulation of transcripts pertaining to photosynthesis, which suggests a notable compartmentalization of the process within the endosperm (Belmonte et al., 2013b). Finally, the CZE is the most transcriptionally distinct region within *Arabidopsis*, with a transcriptome rich in gibberellic acid, abscisic acid, and cytokinin biosynthetic genes, which implies its utility as a communication vector in the seed. Interestingly, the chalazal pole also has distinct anatomy in *Arabidopsis* (as discussed in 1.1.2) relative to the rest of the Brassicaceae, which raises questions regarding how the CPT contributes to seed development in the Brassiceae especially. While the chalazal pole is a maternal tissue potentially derived from the chalazal nucellus, it is unclear if its proximity to the CZE has any relevance in the signaling cascades the CZE may be involved in.

Laser microdissection has also been carried out in *Zea mays* seeds to explore the transcriptional variation within seed subregions (Zhan et al., 2015). As in the Brassicaceae, maternal and filial subregions are highly distinct transcriptionally, but in stark contrast the endosperm regions were the least distinct regions in *Zea* seeds. Monocots, and especially the Poaceae, have complex fusion of maternal subregions, wherein the SC will fuse with the

pericarp to form the caryopsis in grasses, which may explain the increased transcriptional diversity of the maternal subregions. The authors additionally reported a high degree of transcriptional synchronicity between endosperm subregions, which may also explain their comparative lack of diversity. Further, laser microdissection has been employed in investigating embryo abortion in *Citrus*, fibre elongation in *Gossypium*, and developmental networks in *Glycine* seed development (Wu et al., 2007; Danzer et al., 2015; Ando et al., 2021; Lu et al., 2023).

1.3.4 Laser Microdissection in a world of single-cell RNA-sequencing

LMD methods are a powerful tool for isolating sections of tissue for downstream genetic analysis but requires chemical fixation in order to cut out the region of interest (Belmonte et al., 2013b). Chemical fixation of cells necessitates both the efficiency of the fixative to preserve cells quickly, but also that the fixative does not cause substantial alterations in cellular morphology in the process. Unfortunately, transcriptome profiles of plants change very rapidly in response to stimuli, which means even the most efficient fixatives will cause shifts in the mRNA profile of treated samples before fixation is complete. Further, mRNA obtained from LMD samples are often poor in quality and yield due to the extensive chemical processing preceding it (Clément-Ziza et al., 2008), though improvements have been made to mitigate RNA degradation (Bevilacqua et al., 2010; Belmonte et al., 2013b). It is reasonable to turn to single cell RNA-sequencing (scRNA-seq) as a means to overcome this limitation. scRNA-seq has made great advances in our understanding of *Arabidopsis* root development in particular (Dorrity et al., 2021; Shahan et al., 2022), largely because the method depends on digesting plant cell wall material and generating protoplasts. Creating protoplast libraries for scRNA-seq is effective for some plant tissues but runs into difficulty in complex tissues in which separation of unique cells is difficult or impossible (Jean-Baptiste et al., 2019; McFaline-Figueroa et al., 2020). Moreover,

cells with variable cell wall composition, particularly ones with secondary reinforcement, are often unable to be digested and will fail to become protoplasts. These challenges are difficult in seeds which have complex tissue layers, including syncytial endosperm which cannot form protoplasts. scRNA-seq, similar to LMD, can also alter the transcriptome profile of cells as a result of protoplasting. More recently, some methods have emerged detailing and supporting the use of nuclear isolation and sorting as the more reliable method of scRNA-seq sampling, but this approach does limit the transcriptome to that contained within the nucleus (Conde et al., 2021). While many researchers are working to optimize nuclear isolation of complex tissues, the method is still in its infancy and will require additional efforts, especially at the species-level. In the context of *B. napus* seeds, LMD allows us to capture complex tissue layers to examine gene expression within entire subregions. It is possible that hand dissections of certain subregions followed by protoplasting would enable scRNA-seq, but subregion-subregion contamination would be of greater concern. LMD datasets could also help to generate characteristic data profiles of each seed subregion in *B. napus*, which could then be applied to scRNA-seq to assign protoplast samples to specific subregions.

1.4 Small RNAs in Plants

1.4.1 miRNA biogenesis

Epigenetics has an increasingly broader definition as the science surrounding it expands, and since the discovery of micro RNAs (miRNAs) in *C. elegans*, our understanding of how non-coding elements impact the genome has been illuminated (Lee et al., 1993; Wightman et al., 1993). In plants, the two overarching types of small RNA (sRNA) exist as miRNA and small interfering RNA (siRNA), both of which are small strands of RNA involved in gene regulation. miRNAs are small (20-24 nt) molecules encoded by a variety of loci across the genome and are

involved in post-transcriptional gene silencing (Axtell and Meyers, 2018; Wang et al., 2019a). miRNAs and siRNAs are principally differentiated by their means of biogenesis that involve different precursors. miRNAs are transcribed by DNA-Dependent RNA Polymerase II (RNA Pol II) as a self-complementary single stranded miRNA precursor, often many times longer (from 60 to >500 nt) than the mature miRNA strand (20-24 nt) (**Figure 1.3.**) (Bologna and Voinnet, 2014). miRNA precursors, due to their self-complementarity, form a hairpin (hp-miRNA) secondary structure. Interestingly, miRNAs can be encoded either between protein-coding regions (intergenic miRNA) or within the introns of a genic region (intronic miRNA), and the loci encoding the miRNA affects the downstream processing of the miRNA precursor (Millar and Waterhouse, 2005). Intergenic miRNAs are independent RNA Pol II products and are processed individually. Intronic miRNAs, however, are processed alongside the cleavage of the parent mRNA transcript, yielding a functional mRNA and intronic hp-miRNA(s). Following the annotation reviewed by Wang et al. (2019a), hp-miRNAs form a long secondary structure consisting of the terminal loop, upper stem, the mature miRNA sequence and its complementary miRNA* sequence, the lower stem, and the two arms. The hp-miRNA is then recognized by the arms of the precursor and cleaved by DICER-LIKE (DCL) proteins into the mature miRNA and miRNA* strands. While there are four different DCLs in most flowering plants, DCL1 catalyzes the cleavage of most hp-miRNAs. The mature miRNA strand is then loaded unto an ARGONAUTE (AGO) protein. The complex of the mature miRNA bound to the AGO forms the RNA-Induced Silencing Complex (RISC), which is then capable of transcriptional repression. Complementary binding of the mature miRNA to a target transcript causes the cleavage of the target by the AGO, preventing its translation. These fine switches to control gene expression are vital in development and growth, and their absence can cause

severe growth deformities and developmental defects in plants both reproductively and vegetatively (Nag and Jack, 2010; de Lima et al., 2012; Samad et al., 2017).

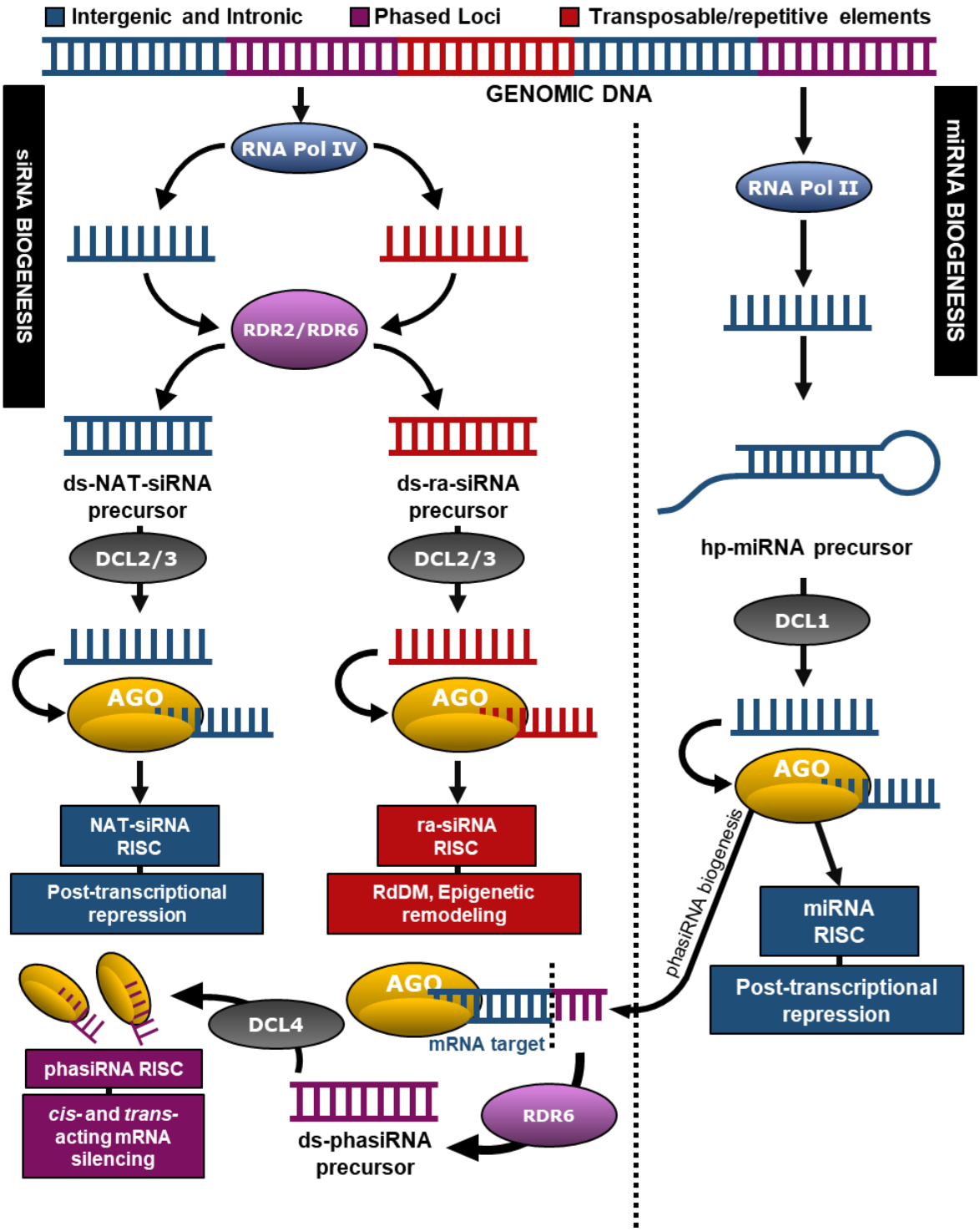


Figure 1.3. Approximated model of the sRNA biogenesis pathways in plants including both the miRNA and siRNA cascades. Genomic origin is indicated by colour, with intergenic/intronic regions in blue, transposable and repetitive elements in red, and phased loci in purple . RNA polymerase II (RNA Pol II) transcribes miRNA precursors, whereas RNA Polymerase IV (RNA Pol IV) transcribes siRNA precursors. DICER-like 1 (DCL1) cleaves hp-miRNA precursors to 21-24 nts, which are then loaded into ARGONAUTE (AGO) proteins to form the RNA Induced Silencing Complex (RISC) and represses transcription of the gene targeted by the mature miRNA. Single stranded siRNA precursors have their complementary strand synthesized by RNA-Directed RNA Polymerase 2/6 (RDR2/RDR6). The dsRNA is then cleaved by either DICER-like 2/3 and loaded into AGO proteins. DICER-like 4 cleaves phased siRNA precursors and are incorporated into phasiRNA RISCs.

1.4.2 siRNA biogenesis

While miRNAs are relatively limited in their genomic distribution to be within and proximal to protein-coding gene loci, siRNA are a highly diverse class of molecules originating from almost anywhere within the genome. siRNAs can be both intergenic and intronic, but they can also originate from repetitive elements including transposable elements and heterochromatic regions. The most discrete difference between miRNAs and siRNAs, however, is their mode of biogenesis. siRNAs are transcribed by RNA Pol IV/V, and a complementary strand is synthesized by RNA-dependent RNA polymerase (RDR) to form a double-stranded siRNA (ds-siRNA) precursor (**Figure 1.3**) (Zhou and Law, 2015). The ds-siRNA precursor is then cleaved by DCL proteins into a 20-24-nt mature siRNA. The mature siRNA then follows similar steps as those within the canonical miRNA pathway. The mature siRNA is loaded into an AGO to form the RISC. Due to their various encoding loci, the targets of siRNA are more variable than that of miRNA. siRNAs encoded by transposable elements, heterochromatic regions, and other repetitive regions are deemed repeat-associated siRNAs (ra-siRNAs) which are typically involved in epigenetic restructuring and DNA methylation cascades. siRNAs encoded by intergenic or intronic regions that target mRNAs in post-transcriptional gene silencing ala miRNA silencing are intuitively called natural antisense siRNA (NAT-siRNA).

Finally, siRNA loci with repetitive mature sequences encode a phased siRNA (phasiRNA) that is cleaved into multiple mature siRNAs from the same ds-RNA precursor by DCL4 (DCL5 in the Poaceae). siRNAs targeting the same region encoding them are termed *cis*-acting, while those targeting regions elsewhere in the genome are *trans*-acting. It is important to note, however, that ra-siRNAs can be derived from phased loci, and most phasiRNAs target mRNAs like NAT-siRNAs. Taken together, both the biogenesis and genomic architecture of siRNAs in plants is highly complex and difficult to interpret, which has slowed scientific progress on determining mode of action and function of these convoluted molecules.

1.5 Small RNAs in plant reproductive development

1.5.1 Known miRNAs governing female reproduction

The history surrounding the roles miRNAs play in seed development is a complicated one. Initially, several mutants were identified to cause seed lethality in *Arabidopsis* that were later identified to have similar sequence homology to Dicer within *Drosophila*, and the locus was renamed to be DCL1. The *embryo defective 76 (emb76)* mutant is one such case, wherein the mutant derived from a T-DNA insert population ceased embryo development past the heart stage, and resulted in full seed lethality (Errampalli et al., 1991). Similarly, the mutants *sin1* (*Short-integuments1)*, *sus1* (*suspensor1)*, and *caf (carpel factory)* all demonstrated varying developmental defects in both filial and maternal tissues. *sin1* mutants failed to establish embryo polarity and caused defects in cotyledon development (Ray et al., 1996). The *sus1* mutant was characterized by the proliferation of the suspensor after embryo development was terminated (Schwartz et al., 1994). The eventual discovery that these mutants were all the product of disrupting DCL1 activity was the initiation of the cascade of research into the role of sRNAs in seed development (Schauer et al., 2002).

The downstream effects of disruption in DCL1 expression are severely deleterious to early embryo development. Nodine and Bartel (2010) reported that cell differentiation in *dcl1* mutants are perturbed as early as the 8-cell stage with >50 miRNA targets overexpressed in the globular embryo. The targets with the greatest shift in transcript abundance in *dcl1* mutants were targets of MIR156, indicating its importance in embryonic development. Further, it seems that miRNAs within the embryo serve to repress transcripts responsible for developmental transition, and their inactivity causes early developmental defects due to premature transitional changes. This is supported in somatic embryogenesis studies (SE), wherein a majority of miRNAs are invoked during SE, including 98% of the miRNA genes in *Arabidopsis* being active and 64% being differentially expressed during SE (Szyrajew et al., 2017; Siddiqui et al., 2019). Altogether, miRNAs culminate as a pivotal component of developmental transitions in reproductive tissues.

1.5.2 phasiRNAs in reproductive development

Phasing within siRNA loci is common in flowering plants, and especially so within the Poaceae. Phased loci are loci in which the mature sequence of the siRNA is repeated numerous times, and processing via DCL proteins leads to the generation of many mature phasiRNAs. While DCL2/3/4 function partially redundantly with each other in processing ra-siRNA, NAT-siRNA, and phasiRNA precursors, DCL5, found exclusively within the Poaceae, also processes phasiRNAs. phasiRNAs follow a slightly different biogenesis pathway than that of other classes of siRNAs involving either a "one-hit" or "two-hit" system (Chen et al., 2018). In the "one-hit" system, a miRNA-AGO RISC cleaves a single target transcript into two fragments, and the ds-siRNA precursor is synthesized via RDR6 from a single fragment (**Figure 1.3**). In the "two-hit" system, the transcript has two miRNA targeted sites, and is bound by two miRNA-AGO

complexes and cleaved. The fragment upstream of the transcript post-cleavage is synthesized into a ds-siRNA by RDR6 and processed by DCL4/5 to generate the mature phasiRNAs. Importantly, phasiRNAs can be *cis*- or *trans*- acting in their targeting and silence mRNAs ala miRNA or NAT-siRNA.

There are few phased loci involved in reproduction that are well understood in *Arabidopsis* or even within the Brassicaceae. It is known that phasiRNAs targeted by miR828 in *Arabidopsis* are involved in the regulation of MYB transcription factors which are required for successful seed development (Zhang et al., 2013; Yang et al., 2018), but most interactions of phasiRNAs are unverified or purely theoretical. Perhaps the most comprehensive phasiRNA regulatory network is that of the miR2275 targeting 24-nt phasiRNAs in grasses, but this was later found to be prevalent in eudicot lineages as well, including the Solanaceae (Polydore et al., 2018). miR2275 accumulates in the tapetum of anthers early in development and recedes post-meiosis, and absence of this activity causes developmental lag in the anther (Ono et al., 2018). Fascinatingly, not only does miR2275 seem to be absent from the Brassicaceae in general, 24-nt phasiRNAs invoked in reproductive tissues seem to have been lost in the entirety of the Brassicales. Despite this, phasiRNAs are still prevalent in the Brassicaceae, though their role in reproduction, if any, remains enigmatic.

1.5.3 Roles of ra-siRNAs in DNA methylation

The diversity of flowering plants is truly exemplified in their genome structure. Many lineages of angiosperms have multiple WGDs which precede them, and by extension, non-coding elements are prolific in many species. Transposable elements (TEs) and genomic redundancy are commonplace in plants with varying benefits and challenges. While TEs have been shown to enable duplication of important genes or regions (Krasileva, 2019), they are also

implicated in genomic damage (Contreras et al., 2015). Methylated TEs are also involved in silencing neighbouring coding genes (Freeling et al., 2015) and consequently, methylation of TEs is a crucial process to reign in the chaotic nature of polyploid angiosperm genomes. Epigenetic remodeling and genome maintenance are accomplished by way of ra-siRNAs (referred to in some literature as heterochromatic siRNAs, het-siRNA). ra-siRNAs are encoded by transposable elements and/or heavily repetitive regions within the genome and are ubiquitously 24-nt in length. Their most well characterized function is their integral role in RNA-dependent DNA Methylation (RdDM). Plants are able to methylate cytosines in three separate contexts: CG, CHG, and CHH (where H = A, C, or T). DNA methylation is enabled by DNA Methyltransferases (DNMT), which bind methyl groups symmetrically to CG and CHG contexts, and asymmetrically to the CHH context. Land plants possess four ancient DNMTs: methyltransferase 1 (MET1), DNA methyltransferase 3 (DNMT3), chromomethylase (CMT), and domain rearranged methyltransferase (DRM) (Springer et al., 2015). CMT and DRM are land plant-specific, and date back to ancient bryophyte lineages (Bewick et al., 2017; Pei et al., 2019). These components of DNA methylation machinery are all recruited in RdDM, wherein a 24-nt ra-siRNA guides DNMT to a locus to be methylated. This methylation can occur in any cytosine context and is the primary way methylation architecture is sustained in plants. ra-siRNAs are currently the only class of sRNA known to be involved in pre-transcriptional regulation and are thus hypothesized to be integral in maintaining genome structure.

RdDM is required for plants to uphold stringent control over TE methylation during reproduction. It is known that plants do not undergo the mass erasure of CG methylation that mammals do during embryogenesis, and instead heavily methylate CHG and CHH contexts as seed development progresses (Bouyer et al., 2017). Further, DNMTs are scarcely expressed

during megagametogenesis in *Arabidopsis,* and climb steadily in expression throughout embryogenesis. The flux of both DNA methylation machinery and methylation itself suggests not only that ra-siRNAs are heavily implicated in seed development, but also that the versatility of the RdDM pathway is likely pivotal in embryogenesis.

DNA methylation in plants does not coincide as cleanly with silencing of gene expression as it does in animals. In plants, DNA methylation has even been shown to promote gene expression with the assistance of SUVH1 and SUVH3, which recruit transcriptional machinery to methylated areas of the genome (Harris et al., 2018). However, DNA methylation in plants is ubiquitous in repetitive genome regions, especially during reproduction. In seed development, DNA methylation increases across the genome upon reaching maturation in *Arabidopsis*, wherein much of the plant's repetitive regions are heavily methylated (Kawakatsu et al., 2017). This shift in architecture is believed to silence these repetitive regions of the genome, particularly transposable elements, to reduce genomic noise during the reproductive process. *B. napus* represents the union of two genomes accumulated repetitive regions from a WGT and allopolyploidization, and consequently must regulate and manage these extensively repetitive genomes. As such, the *B. napus* seed presents a unique research opportunity to uncover the subgenomic and fractionated genomic biases of DNA methylation in a complex amphidiploid.

1.6 Identification of small RNA species

1.6.1 Current methods in identifying miRNAs

Predictions of candidate miRNAs rely on their modes of biogenesis dictating the miRNA's structure. Unlike discrete coding regions, miRNA loci are not flanked by promoters, initiation sites, nor terminated with a stop codon thus complicating their annotation.

Computational predictions rely on the hp-miRNA precursor's secondary structure and the presence of the miRNA* strand within the sequencing dataset to accurately predict both the precursor and mature miRNA sequences. In the past, miRNA predictions have been highly vulnerable to false positives above all else, causing overestimation of the active miRNAs within species, and consequently complicating laboratory confirmations of their presence/activity (Alptekin et al., 2017; Axtell and Meyers, 2018). Criteria covered by Axtell and Meyers (2018) aimed to reduce these false positives by suggesting higher stringency when using predictive software. Previously, miRNA annotation permitted extensive secondary structures (many loops) within the precursor sequence and generally allowed for >300-nt precursors. It is now understood that these structures are unstable and unlikely to form *in planta* (Polydore and Axtell, 2018). Additionally, it was clarified that the miRNA/miRNA* duplex would require minimal secondary structure (<5 mismatches, no asymmetric bulges in the duplex), and that the miRNA must be predicted within at least two biological replicates. These criteria hope to minimize the false positives within public databases, especially miRBase.

Much of the updated software available for miRNA identification pipelines work with small RNA-seq data and follow many basic principles that must be met to identify novel miRNAs. Software like miRDeep-P2 and miRplant are based off of the popular miRDeep-P, which was adapted for plants from miRDeep (Friedländer et al., 2008; Yang and Li, 2011; An et al., 2014; Kuang et al., 2019). The most recent rendition of this program, miRDeep-P2, optimizes for speed and extra plant-specific criteria. In brief, the program considers a small size range (19-24-nt) as valid miRNA candidates, then maps their encoding location to determine if a precursor both a) appears in the dataset and b) can form a valid hp-miRNA. The hp-miRNA must be capable of forming a valid secondary stem-loop structure, and finally, the miRNA*

strand must appear in the dataset. Together, these criteria effectively filter for all precursors and iterations of a mature miRNA molecule. Target prediction of mRNAs can be done through various open-source software, but psRNATarget is a well-performing tool with many genome references readily available (Dai and Zhao, 2011; Srivastava et al., 2014). This tool enables full customization of the parameters of target identification, and defaults to parameters highly important to plants, such as a 100% match of the seed region of the miRNA to its target, a known important feature of miRNA targeting in plants (Axtell et al., 2011).

1.6.2 Criteria for siRNA classification

siRNAs, encoded by a ds-siRNA precursor and being ubiquitous throughout the genome, are much harder to accurately predict than miRNAs. Additionally, many siRNAs do not target mRNAs for post-transcriptional gene silencing, which furthers the difficulty in properly identifying siRNA loci. While many labs had used in-house scripts to identify all sRNAs within a dataset for sometime, Axtell (2013) had published an open source software Shortstack which has been consistently updated since its release. Shortstack is now widely used as an siRNA identification tool in plants. Shortstack assigns miRNAs in a similar but more stringent manner than miRDeep-P, but it also calls siRNA loci as well. To call siRNAs, locations with high coverage are first deemed candidate loci. Loci with high coverage are then flanked by a stretch of nucleotides to mimic the ds-RNA precursor and form a cluster. Clusters containing reads within the size limitations of 20-24-nt are then filtered to reflect the size ranges of mature siRNAs. In short, siRNAs are broadly classified based on their a) size and b) clustering from sequencing data. While miRNA predictions can rely on both 1) the prediction of a hairpin secondary structure, and 2) the presence of a miRNA* strand for further confirmation, the ds-siRNA duplex is much harder to accurately predict and annotate. As siRNAs are highly

abundant in most genomes, this results in many more novel siRNAs present than that of miRNA.

1.7 Current state of sRNA annotations in land plants

miRNAs in plants are currently databased for public access mostly within miRbase, but also in a new database PmiREN (Kozomara et al., 2019; Guo et al., 2020). miRbase has served as the primary repository for miRNA annotations since its inception in 2006 (Griffiths-Jones, 2006) with miRNA species documented across 23 flowering plant families. miRBase catalogs both the precursor sequences and the mature sequence of every documented miRNA. However, miRBase does not require chromosome location or strand specification. In addition to this, as discussed in 1.6.1, miRNA prediction is highly vulnerable to false positives and as a result, some records within miRBase are questionable. With new attention being given to these important epigenetic regulators, more stringent guidelines have arisen to minimize false positive documentation in public databases, and current miRNA prediction pipelines are prioritizing these criteria (Axtell, 2013; Axtell and Meyers, 2018; Kuang et al., 2019).

siRNAs are comparatively less extensively documented than miRNAs. This is due in part to their novelty and difficulty in computational prediction, but also a consequence of being markedly more difficult to validate. Many siRNAs are encoded by notoriously low coverage regions (TEs, repetitive elements), and ra-siRNAs function explicitly in RdDM and epigenetic modification. As a result, these sRNAs are difficult to characterize and predict. Nevertheless, there are some databases which annotate siRNAs. The most comprehensive of them PlantNATsDB (NAT-siRNA), but has not been updated since 2014 (Chen et al., 2012). Otherwise, the few loci known to be phased and/or *cis*- or *trans*-acting are documented within The *Arabidopsis* Information Resource (TAIR) (Berardini et al., 2015).

1.8 Research Objectives

The research outlined in this thesis aims to profile the transcriptomic and epigenetic shifts of the *B. napus* seed, particularly in the context of polyploidy. Herein, I describe the transcriptional atlas that underlies seed development and use anatomical tools to explore how the complex *B. napus* seed coordinates its development and balances subgenomic bias.

The works included in this thesis are only possible due to the work of previous students and collaborative efforts. The chapters hereafter are the synthesis of these collaborations —the many research projects that endeavoured to untangle the exceptional complexity of the seed. I describe the transcriptome profile of the early *B. napus* seed first, describing the anatomical transitions accompanying genetic changes in the seed subregions. Following that, I detail a small portion of the immense complexity of epigenetic and siRNA bias through the lens of polyploidy and investigate how these biases change throughout development and in which genomic contexts. Finally, I combine the ideas from the previous two chapters and examine subgenomic bias of seed subregions across four stages of development spanning pre-fertilization to maturation. Together, this work represents the substantial undertaking of myself and many others before me to describe, quantify, and appreciate the genomic balance of the *B. napus* seed.

Contributions of other co-authors are indicated within their respective chapters.

1.9 References

Alptekin B, Akpinar BA, Budak H (2017) A comprehensive prescription for plant miRNA identification. Front Plant Sci **7**: 2058

Amich F, García-Barriuso M, Bernardos S (2007) Polyploidy and speciation in the orchid flora of the Iberian Peninsula. Botanica Helvetica **117**: 143–157

An J, Lai J, Sajjanhar A, Lehman ML, Nelson CC (2014) MiRPlant: An integrated tool for identification of plant miRNA from RNA sequencing data. BMC Bioinformatics **15**: 275

Ando A, Kirkbride RC, Jones DC, Grimwood J, Chen ZJ (2021) LCM and RNA-seq analyses revealed roles of cell cycle and translational regulation and homoeolog expression bias in cotton fiber cell initiation. BMC Genomics **22**: 1–16

Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA **19**: 740–751

Axtell MJ, Meyers BC (2018) Revisiting criteria for plant microRNA annotation in the Era of big data. Plant Cell **30**: 272–284

Axtell MJ, Westholm JO, Lai EC (2011) Vive la différence: Biogenesis and evolution of microRNAs in plants and animals. Genome Biol **12**: 1–13

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V (2019) Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun. doi: 10.1038/S41467-019-13730-0

Batista RA, Figueiredo DD, Santos-González J, Köhler C (2019a) Auxin regulates endosperm cellularization in Arabidopsis. Genes Dev **33**: 466–476

Batista RA, Figueiredo DD, Santos-González J, Köhler C (2019b) Auxin regulates endosperm cellularization in Arabidopsis. Genes Dev **33**: 466–476

Bedi S, Sengupta S, Ray A, Nag Chaudhuri R (2016) ABI3 mediates dehydration stress recovery response in Arabidopsis thaliana by regulating expression of downstream genes. Plant Science **250**: 125–140

Beeckman T, De Rycke R, Viane R, Inzé D (2000) Histological study of seed coat development in Arabidopsis thaliana. J Plant Res **113**: 139–148

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, et al (2013a) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci U S A **110**: E435–E444

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, et al (2013b) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci U S A **110**: E435–E444

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015) The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis **53**: 474–485

Bevilacqua C, Makhzami S, Helbling JC, Defrenaix P, Martin P (2010) Maintaining RNA integrity in a homogeneous population of mammary epithelial cells isolated by Laser Capture Microdissection. BMC Cell Biol **11**: 1–13

Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ (2017) The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol **18**: 65

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP (2021) Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. New Phytologist **230**: 354–371

Bleckmann A, Alter S, Dresselhaus T (2014) The beginning of a seed: Regulatory mechanisms of double fertilization. Front Plant Sci. doi: 10.3389/FPLS.2014.00452/ABSTRACT

Bologna NG, Voinnet O (2014) The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis . Annu Rev Plant Biol **65**: 473–503

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nature Reviews Molecular Cell Biology 2015 16:12 **16**: 727–741

Boscá S, Knauer S, Laux T (2011a) Embryonic development in arabidopsis thaliana: From the zygote division to the shoot meristem. Front Plant Sci. doi: 10.3389/fpls.2011.00093

Boscá S, Knauer S, Laux T (2011b) Embryonic development in arabidopsis thaliana: From the zygote division to the shoot meristem. Front Plant Sci **2**: 16239

Bouyer D, Kramdi A, Kassam M, Heese M, Schnittger A, Roudier F, Colot V (2017) DNA methylation dynamics during early plant life. Genome Biol. doi: 10.1186/s13059-017-1313-0

Brown RC, Lemmon BE, Nguyen H (2004a) Comparative anatomy of the chalazal endosperm cyst in seeds of the Brassicaceae. Botanical Journal of the Linnean Society **144**: 375–394

Brown RC, Lemmon BE, Nguyen H (2004b) Comparative anatomy of the chalazal endosperm cyst in seeds of the Brassicaceae. Botanical Journal of the Linnean Society **144**: 375–394

Brown RC, Lemmon BE, Nguyen H, Olsen OA (1999) Development of endosperm in Arabidopsis thaliana. Sex Plant Reprod **12**: 32–42

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014a) Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. Science (1979) **345**: 950–953

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014b) Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. Science (1979) **345**: 950–953

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014c) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345**: 950–953

Che NY, Yang Y, Li YD, Wang LL, Huang P, Gao Y, An CC (2009) Efficient LEC2 activation of OLEOSIN expression requires two neighboring RY elements on its promoter. Sci China C Life Sci **52**: 854–863

Chen C, Zeng Z, Liu Z, Xia R (2018) Small RNAs, emerging regulators critical for the development of horticultural traits. Hortic Res. doi: 10.1038/s41438-018-0072-8

Chen D, Yuan C, Zhang J, Zhang Z, Bai L, Meng Y, Chen L-L, Chen M (2012) PlantNATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acid Res **40**: 1187–1193

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X (2018) Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants 2018 4:5 **4**: 258–268

Clément-Ziza M, Munnich A, Lyonnet S, Jaubert F, Besmond C (2008) Stabilization of RNA during laser capture microdissection by performing experiments under argon atmosphere or using ethanol as a solvent in staining solutions. RNA **14**: 2698

Coen O, Fiume E, Xu W, De Vos D, Lu J, Pechoux C, Lepiniec L, Magnani E (2017) Developmental patterning of the sub-epidermal integument cell layer in Arabidopsis seeds. Development (Cambridge) **144**: 1490–1497

Conde D, Triozzi PM, Balmant KM, Doty AL, Miranda M, Boullosa A, Schmidt HW, Pereira WJ, Dervinis C, Kirst M (2021) A robust method of nuclei isolation for single-cell RNA sequencing of solid tissues from the plant genus Populus. PLoS One **16**: e0251149

Contreras B, Vives C, Castells R, Casacuberta JM (2015) The impact of transposable elements in the evolution of plant genomes: From selfish elements to key players. Evolutionary Biology: Biodiversification from Genotype to Phenotype. Springer International Publishing, pp 93–105

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res **39**: 155–159

Danzer J, Meilott E, Bui AQ, Le BH, Martin P, Hashimoto M, Perez-Lesher J, Chen M, Pelletier JM, Somers DA, et al (2015) Down-Regulating the Expression of 53 Soybean Transcription Factor Genes Uncovers a Role for SPEECHLESS in Initiating Stomatal Cell Lineages during Embryo Development. Plant Physiol **168**: 1025

Delmas F, Sankaranarayanan S, Deb S, Widdup E, Bournonville C, Bollier N, Northey JGB, McCourt P, Samuel MA (2013) ABI3 controls embryo degreening through Mendel's i locus. Proc Natl Acad Sci U S A **110**: E3888

DeMarini DM, Mutlu E, Warren SH, King C, Gilmour MI, Linak WP (2019) Mutagenicity Emission Factors of Canola Oil and Waste Vegetable Oil Biodiesel: Comparison to Soy Biodiesel. Mutat Res **846**: 403057

Doll NM, Royek S, Fujita S, Okuda S, Chamot S, Stintzi A, Widiez T, Hothorn M, Schaller A, Geldner N, et al (2020) A two-way molecular dialogue between embryo and endosperm is required for seed development. Science **367**: 431–435

Dorrity MW, Alexandre CM, Hamm MO, Vigil AL, Fields S, Queitsch C, Cuperus JT (2021) The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nature Communications 2021 12:1 **12**: 1–12

Errampalli D, Patton D, Castle L, Mickelson L, Hansen K, Schnall J, Feldmann K, Meinke D (1991) Embryonic Lethals and T-DNA lnsertional Mutagenesis in Ara bidopsis.

Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C (2016) Auxin production in the endosperm drives seed coat development in Arabidopsis. Elife. doi: 10.7554/ELIFE.20542

Figueiredo DD, Batista RA, Roszak PJ, Köhler C (2015) Auxin production couples endosperm development to fertilization. Nature Plants 2015 1:12 **1**: 1–6

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011a) Cabbage family affairs: The evolutionary history of Brassicaceae. Trends Plant Sci **16**: 108–116

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011b) Cabbage family affairs: The evolutionary history of Brassicaceae. Trends Plant Sci **16**: 108–116

Freeling M, Xu J, Woodhouse M, Lisch D (2015) A solution to the c-value paradox and the function of junk DNA: The genome balance hypothesis. Mol Plant **8**: 899–910

Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N (2008) Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol **26**: 407–415

Gao P, Quilichini TD, Yang H, Li Q, Nilsen KT, Qin L, Babic V, Liu L, Cram D, Pasha A, et al (2022) Evolutionary divergence in embryo and seed coat development of U's Triangle Brassica species illustrated by a spatiotemporal transcriptome atlas. New Phytologist **233**: 30–51

Gardiner LJ, Quinton-Tulloch M, Olohan L, Price J, Hall N, Hall A (2015) A genome-wide survey of DNA methylation in hexaploid wheat. Genome Biol **16**: 273

Gómez LD, Baud S, Gilday A, Li Y, Graham IA (2006) Delayed embryo development in the ARABIDOPSIS TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. Plant Journal **46**: 69–84

Griffiths-Jones S (2006) miRBase: the microRNA sequence database. Methods Mol Biol **342**: 129–138

Guo Z, Kuang Z, Wang Y, Zhao Y, Tao Y, Cheng C, Yang J, Lu X, Hao C, Wang T, et al (2020) PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Res **48**: D1114–D1121

Harada JJ (2001) Role of arabidopsis LEAFY COTYLEDON genes in seed development. J Plant Physiol **158**: 405–409

Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM, Vaughan RM, Li X, Chen W, Xue Y, Zhong Z, et al (2018) A DNA methylation reader complex that enhances gene transcription. Science (1979) **362**: 1182–1186

Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J, et al (2017) Transcriptomic analysis of seed coats in yellow-seeded Brassica napus reveals novel genes that influence proanthocyanidin biosynthesis. Front Plant Sci **8**: 286392

Huang AHC (2018) Plant Lipid Droplets and Their Associated Proteins: Potential for Rapid Advances. Plant Physiol **176**: 1894

Huang C-H, Zhang C, Liu M, Hu Y, Gao T, Qi J, Ma H (2016) Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics. Mol Biol Evol **33**: 2820–2835

Ingram GC, Penfield S (2017) Dying to live: cell elimination as a developmental strategy in angiosperm seeds. J Exp Bot **68**: 785–796

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463–467

Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperusa JT (2019) Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. Plant Cell **31**: 993–1011

Jo L, Pelletier JM, Harada JJ (2019) Central role of the LEAFY COTYLEDON1 transcription factor in seed development. J Integr Plant Biol **61**: 564–580

de Jong GW, Adams KL (2023) Subgenome-dominant expression and alternative splicing in response to Sclerotinia infection in polyploid Brassica napus and progenitors. Plant J **114**: 142–158

Kawakatsu T, Nery JR, Castanon R, Ecker JR (2017) Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol **18**: 1–12

Khan D, Millar JL, Girard IJ, Belmonte MF (2014) Transcriptional circuitry underlying seed coat development in Arabidopsis. Plant Science **219–220**: 51–60

Khan D, Ziegler DJ, Kalichuk JL, Hoi V, Huynh N, Hajihassani A, Parkin IAP, Robinson SJ, Belmonte MF (2022) Gene expression profiling reveals transcription factor networks and subgenome bias during Brassica napus seed development. The Plant Journal **109**: 477–489

Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res **47**: D155–D162

Krasileva K V. (2019) The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. Curr Opin Plant Biol **48**: 18–25

Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. Development **130**: 6065–6073

Kuang Z, Wang Y, Li L, Yang X (2019) miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. Bioinformatics **35**: 2521–2522

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS (2018) Impact of whole-genome duplication events on diversification rates in angiosperms. Am J Bot **105**: 348–363

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75**: 843–854

Li M, Wang R, Wu X, Wang J (2020) Homoeolog expression bias and expression level dominance (ELD) in four tissues of natural allotetraploid Brassica napus. BMC Genomics **21**: 1–15

Li Q, Qiao X, Yin H, Zhou Y, Dong H, Qi K, Li L, Zhang S (2019) Unbiased subgenome evolution following a recent whole-genome duplication in pear (Pyrus bretschneideri Rehd.). Hortic Res **6**: 1–12

Liao P, Woodfield HK, Harwood JL, Chye ML, Scofield S (2019) Comparative Transcriptomics Analysis of Brassica napus L. during Seed Maturation Reveals Dynamic Changes in Gene Expression between Embryos and Seed Coats and Distinct Expression Profiles of Acyl-CoA-Binding Proteins for Lipid Accumulation. Plant Cell Physiol **60**: 2812–2825

de Lima JC, Loss-Morais G, Margis R (2012) MicroRNAs play critical roles during plant development and in response to abiotic stresses. Genet Mol Biol **35**: 1069–1077

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al (2014a) The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nature Communications 2014 5:1 **5**: 1–11

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al (2014b) The brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun **5**: 1–11

Liu Y, Wang J, Ge W, Wang Z, Li Y, Yang N, Sun S, Zhang L, Wang X (2017) Two highly similar poplar paleo-subgenomes suggest an autotetraploid ancestor of salicaceae plants. Front Plant Sci. doi: 10.3389/fpls.2017.00571

Liu Z, Wang R, Wang J (2020) Comprehensive Transcriptomic Analysis for Developing Seeds of a Synthetic Brassica Hexaploid. Plants **9**: 1–17

Lotan T, Ohto MA, Matsudaira Yee K, West MAL, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93**: 1195–1205

Lu J, Magnani E (2018a) Seed tissue and nutrient partitioning, a case for the nucellus. Plant Reprod **31**: 309–317

Lu J, Magnani E (2018b) Seed tissue and nutrient partitioning, a case for the nucellus. Plant Reprod **31**: 309–317

Lu L, Yang H, Xu Y, Zhang L, Wu J, Yi H (2023) Laser capture microdissection-based spatiotemporal transcriptomes uncover regulatory networks during seed abortion in seedless Ponkan (Citrus reticulata). The Plant Journal **115**: 642–661

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. Genome Res **15**: 516–525

Maggioni L, von Bothmer R, Poulsen G, Lipman E (2018) Domestication, diversity and use of Brassica oleracea L., based on ancient Greek and Latin texts. Genet Resour Crop Evol **65**: 137–159

Mansfield S, Briarty L (1990) Endosperm cellularization in Arabidopsis thaliana L. Arab Inf Serv **27**: 65–72

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264**: 421–424

Mathews S, Kramer EM (2012) The evolution of reproductive structures in seed plants: a re-examination based on insights from developmental genetics. New Phytologist **194**: 910–923

McFaline-Figueroa JL, Trapnell C, Cuperus JT (2020) The promise of single-cell genomics in plants. Curr Opin Plant Biol **54**: 114–121

Meade LE, Plackett ARG, Hilton J (2021) Reconstructing development of the earliest seed integuments raises a new hypothesis for the evolution of ancestral seed-bearing structures. New Phytologist **229**: 1782–1794

Meinke DW (1992) A homoeotic mutant of Arabidopsis thaliana with leafy cotyledons. Science (1979) **258**: 1647–1650

Millar AA, Waterhouse PM (2005) Plant and animal microRNAs: Similarities and differences. Funct Integr Genomics **5**: 129–135

Millar JL, Khan D, Becker MG, Chan A, Dufresne A, Sumner M, Belmonte MF (2015a) Chalazal seed coat development in Brassica napus. Plant Science **241**: 45–54

Millar JL, Khan D, Becker MG, Chan A, Dufresne A, Sumner M, Belmonte MF (2015b) Chalazal seed coat development in Brassica napus. Plant Science **241**: 45–54

Miquel M, Trigui G, D'Andréa S, Kelemen Z, Baud S, Berger A, Deruyffelaere C, Trubuil A, Lepiniec L, Dubreucq B (2014) Specialization of Oleosins in Oil Body Dynamics during Seed Development in Arabidopsis Seeds. Plant Physiol **164**: 1866

Mu J, Tan H, Zheng Q, Fu Y, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, et al (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol **148**: 1042–1054

Nag A, Jack T (2010) Sculpting the flower; the role of microRNAs in flower development. Curr Top Dev Biol. Academic Press Inc., pp 349–378

Nambara E, Hayama R, Tsuchiya Y, Nishimura M, Kawaide H, Kamiya Y, Naito S (2000) The role of ABI3 and FUS3 loci in Arabidopsis thaliana on phase transition from late embryo development to germination. Dev Biol **220**: 412–423

Nambara E, Keith K, McCourt P, Naito S (1995) A regulatory role for the ABI3 gene in the establishment of embryo maturation in Arabidopsis thaliana. Development **121**: 629–636

Nelson SK, Ariizumi T, Steber CM (2017) Biology in the dry seed: Transcriptome changes associated with dry seed dormancy and dormancy loss in the arabidopsis GA-insensitive sleepy1-2 mutant. Front Plant Sci. doi: 10.3389/fpls.2017.02158

Nguyen H, Brown RC, Lemmon BE (2000a) The specialized chalazal endosperm in Arabidopsis thaliana and Lipidium virginicum (Brassicaceae). Protoplasma **212**: 99–110

Nguyen H, Brown RC, Lemmon BE (2000b) The specialized chalazal endosperm in Arabidopsis thaliana and Lipidium virginicum (Brassicaceae). Protoplasma **212**: 99–110

Nguyen TP, Cueff G, Hegedus DD, Rajjou L, Bentsink L (2015) A role for seed storage proteins in Arabidopsis seed longevity. J Exp Bot **66**: 6399–6413

Niñoles R, Ruiz-Pastor CM, Arjona-Mudarra P, Casañ J, Renard J, Bueso E, Mateos R, Serrano R, Gadea J (2022) Transcription Factor DOF4.1 Regulates Seed Longevity in Arabidopsis via Seed Permeability and Modulation of Seed Storage Protein Accumulation. Front Plant Sci **13**: 915184

Nodine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev **24**: 2678–2692

Ono S, Liu H, Tsuda K, Fukai E, Tanaka K, Sasaki T, Nonomura KI (2018) EAT1 transcription factor, a non-cell-autonomous regulator of pollen production, activates meiotic small RNA biogenesis in rice anther tapetum. PLoS Genet **14**: e1007238

Otto SP, Whitton J (2000a) Polyploid incidence and evolution. Annu Rev Genet **34**: 401–437

Otto SP, Whitton J (2000b) Polyploid Incidence and Evolution. Annu Rev Genet **34**: 401–437

Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol **15**: 1–18

Pei L, Zhang L, Li J, Shen C, Qiu P, Tu L, Zhang X, Wang M (2019) Tracing the origin and evolution history of methylation-related genes in plants. BMC Plant Biol **19**: 307

Polydore S, Axtell MJ (2018) Analysis of RDR1/RDR2/RDR6-independent small RNAs in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of short interfering RNA loci. Plant Journal **94**: 1051–1063

Polydore S, Lunardon A, Axtell MJ (2018) Several phased siRNA annotation methods can frequently misidentify 24 nucleotide siRNA-dominated *PHAS* loci. Plant Direct **2**: e00101

Povilus RA, Gehring M (2022) Maternal-filial transfer structures in endosperm: A nexus of nutritional dynamics and seed development. Curr Opin Plant Biol **65**: 102121

Qi X, An H, Hall TE, Di C, Blischak PD, McKibben MTW, Hao Y, Conant GC, Pires JC, Barker MS (2021) Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in Brassica rapa. New Phytologist **230**: 372–386

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biology 2019 20:1 **20**: 1–23

Radchuk V, Borisjuk L (2014) Physical, metabolic and developmental functions of the seed coat. Front Plant Sci **5**: 102496

Ray S, Golden T, Ray A (1996) Maternal effects of the short integument mutation on embryo development in Arabidopsis. Dev Biol **180**: 365–369

Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J, Ma hongma H (2018) Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. Mol Plant **11**: 414–428

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: Polyploidy and crop plants. Am J Bot **101**: 1711–1725

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 Biol **21**: 1–17

Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z, Ismail I (2017) MicroRNA and transcription factor: Key players in plant regulatory network. Front Plant Sci **8**: 565

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, et al (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641

Schauer SE, Jacobsen SE, Meinke DW, Ray A (2002) DICER-LIKE1: Blind men and elephants in Arabidopsis development. Trends Plant Sci **7**: 487–491

Schneitz K, Hülskamp M, Pruitt RE (1995) Wild-type ovule development in Arabidopsis thaliana: a light microscope study of cleared whole-mount tissue. The Plant Journal **7**: 731–749

Schulz P, Jensen WA (1971) Capsella embryogenesis: the chalazal proliferating tissue. J Cell Sci **8**: 201–27

Schwartz BW, Yeung EC, Meinke DW (1994) Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis. Development 120:

Shahan R, Hsu CW, Nolan TM, Cole BJ, Taylor IW, Greenstreet L, Zhang S, Afanassiev A, Vlot AHC, Schiebinger G, et al (2022) A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. Dev Cell **57**: 543-560.e9

Siddiqui ZH, Abbas ZK, Ansari MW, Khan MN (2019) The role of miRNA in somatic embryogenesis. Genomics **111**: 1026–1033

Simon MK, Skinner DJ, Gallagher TL, Gasser CS (2017) Integument development in Arabidopsis depends on interaction of YABBY protein INNER NO OUTER with coactivators and corepressors. Genetics **207**: 1489–1500

Skinner DJ, Brown RH, Kuzoff RK, Gasser CS (2016) Conservation of the role of INNER NO OUTER in development of unitegmic ovules of the Solanaceae despite a divergence in protein function. BMC Plant Biol **16**: 143

SMITH DL (1964) THE EVOLUTION OF THE OVULE. Biological Reviews **39**: 137–159

So KKY, Duncan RW (2021) Breeding Canola (Brassica napus L.) for Protein in Feed and Food. Plants. doi: 10.3390/PLANTS10102220

Springer NM, Lisch D, Li Q (2015) Creating order from chaos: Epigenome dynamics in plants with complex genomes. Plant Cell **28**: 314–325

Srivastava PK, Moturu TR, Pandey P, Baldwin IT, Pandey SP (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. BMC Genomics **15**: 348

Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. Proc Natl Acad Sci U S A **105**: 3151–3156

Stuessy T, Weiss-Schneeweiss H (2019) What drives polyploidization in plants? New Phytologist **223**: 1690–1692

Szyrajew K, Bielewicz D, Dolata J, Wójcik AM, Nowak K, Szczygieł-Sommer A, Szweykowska-Kulinska Z, Jarmolowski A, Gaj MD (2017) MicroRNAs are intensively regulated during induction of somatic embryogenesis in arabidopsis. Front Plant Sci **8**: 18

Tan H, Zhang J, Qi X, Shi X, Zhou J, Wang X, Xiang X (2019) Correlation analysis of the transcriptome and metabolome reveals the regulatory network for lipid synthesis in developing Brassica napus embryos. Plant Mol Biol **99**: 31–44

Tang Y, Yan X, Gu C, Yuan X (2022) Biogenesis, Trafficking, and Function of Small RNAs in Plants. Front Plant Sci **13**: 825477

Traver MS, Bartel B (2023) The ubiquitin-protein ligase MIEL1 localizes to peroxisomes to promote seedling oleosin degradation and lipid droplet mobilization. Proc Natl Acad Sci U S A **120**: e2304870120

Tsai AYL, Kunieda T, Rogalski J, Foster LJ, Ellis BE, Haughn GW (2017) Identification and Characterization of Arabidopsis Seed Coat Mucilage Proteins. Plant Physiol **173**: 1059–1074

Vermon DM, Meinke DW (1994) Embryogenic transformation of the suspensor in twin, a polyembryonic mutant of Arabidopsis. Dev Biol **165**: 566–573

Wan L, Ross ARS, Yang J, Hegedus DD, Kermode AR (2007) Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant Arabidopsis thaliana (thale cress) seeds. Biochemical Journal **404**: 247

Wang F, Perry SE (2013) Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol **161**: 1251–1264

Wang J, Mei J, Ren G (2019a) Plant microRNAs: Biogenesis, homeostasis, and degradation. Front Plant Sci **10**: 360

Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, Sun S, Lei T, Meng F, Wei C, et al (2019b) Polyploidy Index and Its Implications for the Evolution of Polyploids. Front Genet. doi: 10.3389/fgene.2019.00807

Wei L, Du H, Li X, Fan Y, Qian M, Li Y, Wang H, Qu C, Qian W, Xu X, et al (2022) Spatio-temporal transcriptome profiling and subgenome analysis in Brassica napus. The Plant Journal **111**: 1123–1138

West MAL, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. Plant Cell **6**: 1731–1745

Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell **75**: 855–862

Wilkinson LG, Yang X, Burton RA, Würschum T, Tucker MR (2019) Natural Variation in Ovule Morphology Is Influenced by Multiple Tissues and Impacts Downstream Grain Development in Barley (Hordeum vulgare L.). Front Plant Sci **10**: 1374

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y (2018) Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. BMC Genomics **19**: 1–13

Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, et al (2013) The genome of the pear (Pyrus bretschneideri Rehd.). Genome Res **23**: 396–408

Wu Y, Llewellyn DJ, White R, Ruggiero K, Al-Ghazi Y, Dennis ES (2007) Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of the rapidly expanding fibre initial cells on the surface of cotton ovules. Planta **226**: 1475–1490

Xu TT, Ren SC, Song XF, Liu CM (2015) CLE19 expressed in the embryo regulates both cotyledon establishment and endosperm development in Arabidopsis. J Exp Bot **66**: 5217–5227

Xu W, Fiume E, Coen O, Pechoux C, Lepiniec L, Magnani E (2016) Endosperm and Nucellus Develop Antagonistically in Arabidopsis Seeds. Plant Cell **28**: 1343–1360

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

Yang B, Song Z, Li C, Jiang J, Zhou Y, Wang R, Wang Q, Ni C, Liang Q, Chen H, et al (2018) RSM1, an Arabidopsis MYB protein, interacts with HY5/HYH to modulate seed germination and seedling development in response to abscisic acid and salinity. PLoS Genet **14**: e1007839

Yang X, Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics **27**: 2614–2615

Yuan J, Jiao W, Liu Y, Ye W, Wang X, Liu B, Song Q, Chen ZJ (2020) Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. BMC Biol **18**: 1–13

Zhan J, Thakare D, Ma C, Lloyd A, Nixon NM, Arakaki AM, Burnett WJ, Logan KO, Wang D, Wang X, et al (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. Plant Cell **27**: 513–531

Zhang MX, Zhu SS, Xu YC, Guo YL, Yang WC, Li HJ (2020) Transcriptional repression specifies the central cell for double fertilization. Proc Natl Acad Sci U S A **117**: 6231–6236

Zhang Y, Liang W, Shi J, Xu J, Zhang D (2013) MYB56 Encoding a R2R3 MYB Transcription Factor Regulates Seed Size in Arabidopsis thaliana. J Integr Plant Biol **55**: 1166–1178

Zhang Y, Zhang L, Xiao Q, Wu C, Zhang J, Xu Q, Yu Z, Bao S, Wang J, Li Y, et al (2022) Two independent allohexaploidizations and genomic fractionation in Solanales. Front Plant Sci **13**: 1001402

Zhao M, Zhang B, Lisch D, Ma J (2017) Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. Plant Cell **29**: 2974–2994

Zheng P, Zheng C, Otegui MS, Li F (2022) Endomembrane mediated-trafficking of seed storage proteins: from Arabidopsis to cereal crops. J Exp Bot **73**: 1312–1326

Zhou M, Law JA (2015) RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules. Curr Opin Plant Biol **27**: 154–164

Ziegler DJ, Khan D, Kalichuk JL, Becker MG, Belmonte MF (2019) Transcriptome landscape of the early Brassica napus seed. J Integr Plant Biol **61**: 639–650

Ziegler DJ, Khan D, Pulgar-Vidal N, Parkin IAP, Robinson SJ, Belmonte MF (2023) Genomic asymmetry of the Brassica napus seed: epigenetic contributions of DNA methylation and small RNAs to subgenome bias. The Plant Journal **115**: 690–708

(2018) FAOSTAT. http://www.fao.org/faostat/en/#data/QC

(2017) Canola now worth $26.7 billion to Canadian economy - Canola Council of Canada. https://www.canolacouncil.org/news/canola-now-worth-$267-billion-to-canadian-economy/

2 Transcriptome Landscape of the early *Brassica napus* seed

Dylan J. Ziegler[1], Deirdre Khan[1], Jenna L. Kalichuk[1], Michael G. Becker[1], Mark F. Belmonte[1]*
[1]University of Manitoba, Winnipeg, Manitoba

This chapter is published in the Journal of Integrative Plant Biology as:
Ziegler DJ, Khan D, Kalichuk JL, Becker MG, Belmonte MF. 2019. Transcriptome landscape of the early *Brassica napus* seed. *J Integ Plant Biol*. 61(5):639-650. DOI: 10.1111/jipb.12812.

2.1 Abstract

Brassica napus L. (canola) is one of the world's most economically important oilseeds. Despite our growing knowledge of Brassica genetics, we still know little about the genes and gene regulatory networks underlying early seed development. In this work, we use laser microdissection coupled with RNA sequencing to profile gene activity of both the maternal and filial subregions of the globular seed. We find subregions of the chalazal end including the chalazal endosperm, chalazal proliferating tissue, and chalazal seed coat, have unique transcriptome profiles associated with hormone biosynthesis and polysaccharide metabolism. We confirm that the chalazal seed coat is uniquely enriched for sucrose biosynthesis and transport, and that the chalazal endosperm may function as an important regulator of the maternal region through brassinosteroid synthesis. The chalazal proliferating tissue, a poorly understood subregion, was specifically enriched in transcripts associated with megasporogenesis and trehalose biosynthesis, suggesting this ephemeral structure plays an important role in both sporophytic development and carbon nutrient balance, respectively. Finally, compartmentalization of transcription factors and their regulatory circuits has uncovered previously unknown roles for the chalazal pole in early seed development.

2.2 Introduction

Oilseed crops are defined by their substantial storage reserves within the seed. One of the most globally coveted oilseeds is *Brassica napus* L. (canola) and is the second most important oil crop worldwide (FAOSTAT 2014). The seed itself is composed of approximately 45% oil by seed mass and 40% protein-rich seed meal, establishing *B. napus* as an important nutritional food source (Rahman 2013). The oil from the *B. napus* seed is consumed worldwide, is considered heart-healthy, and is also used as a source of bio-fuel (Fore et al. 2011). Despite its importance, our understanding of *B. napus* seed development is incomplete, wherein our knowledge of the precise transcriptomic profiles of individual cells and tissues is lacking.

The *B. napus* seed can be divided into distinct anatomical and morphological regions that each uniquely contribute to seed development. Ovules of the Brassicaceae are bitegmic and are amphitropous post-fertilization (Brown et al. 2004), meaning the ovule is partially inverted so that the micropyle and chalaza are located proximally to each other. The filial regions of the seed include the embryo proper (EP), the micropylar endosperm (MCE) embedded at the micropylar pole, the peripheral endosperm (PEN) which encircles the central chamber of the seed, and the chalazal endosperm (CZE), which is located proximally to the chalazal proliferating tissue (CPT) (Figure 2.1A). The maternal subregions of the seed include the inner (ISC) and outer seed coat (OSC), the chalazal seed coat (CZSC) and also includes the CPT, which is inserted intramarginally to the CZSC (Figure 2.1B). The CPT is a structure that is diminutive in Arabidopsis (Nguyen et al. 2000) but is prominent in Capsella bursa-pastoris (Schulz and Jensen 1971; Schulz et al. 1974) Stanleya pinnata, Erysimum asperum, Chorispora tenella, Descurainia richardsonii, and other Brassicaceae species (Brown et al. 2004), yet most of its biological function remains unexplored.

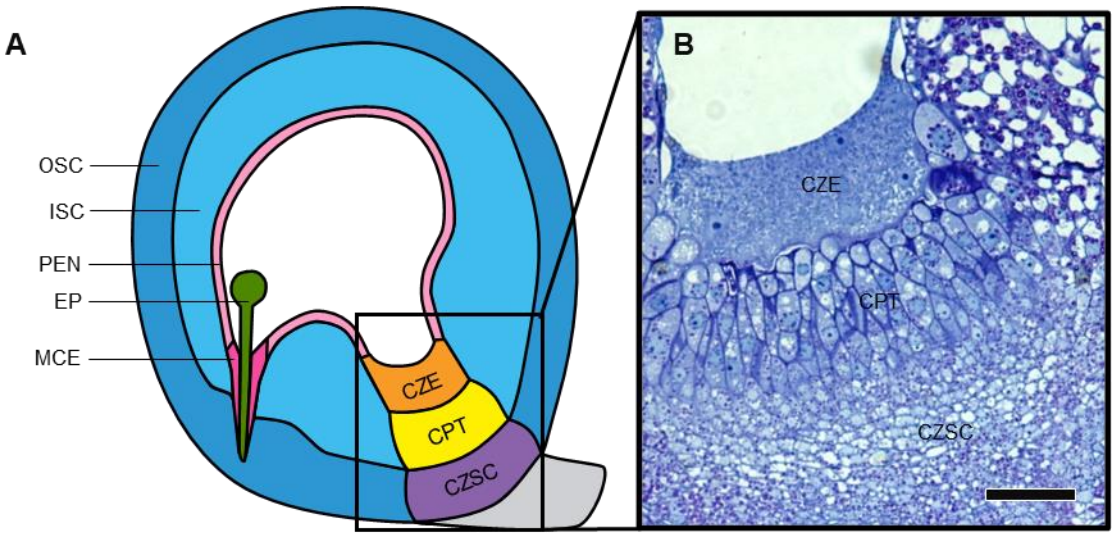


Figure 2.1. The *B. napus* globular seed (A) Schematic diagram of the *B. napus* globular seed illustrating the embryo proper (EP), micropylar endosperm (MCE), peripheral endosperm (PEN), chalazal endosperm (CZE), outer seed coat (OSC), inner seed coat (ISC), chalazal seed coat (CZSC), and the chalazal proliferating tissue (CPT). **(B)** longitudinal section of the chalazal pole stained with toluidine blue showing compact cells of the CZSC, elongated cytoplasmically-dense cells of the CPT, and syncytial cells of the CZE. Scale = 100 µm.

In amphitropous ovules, the chalazal pole is directly adjacent to the attachment of the funiculus, which is the vector of transport between the placenta and the developing seed (Khan et al. 2015). Previous work on *B. napus* has confirmed the abundant endomembrane networks and transporter activity within the funiculus of the globular seed and has also shown enrichment of genes involved in the production and organization of vascular tissues (Chan et al. 2016). The same work reported strong tissue-specific compartmentalization of biological processes within this structure including water and phloem transport of the vasculature, and growth regulation being confined to the cortex. Though the vascular trace of the funiculus terminates at the funiculus-CZSC interface in Arabidopsis, the vascular system penetrates the CZSC in *B. napus* (Millar et al. 2015). Furthermore, transport linking the maternal and filial regions have also been reported in Arabidopsis and genes associated with the transport of

amino acids (Besnard et al. 2018) and phosphates (Vogiatzaki et al. 2017) have been shown to preferentially accumulate in the chalazal end of the seed. Thus, the chalazal pole of *B. napus* may function as a gateway for nutrient unloading and seed filling thereby guiding developmental processes.

While the chalazal pole of Brassica species has been well documented anatomically (Nguyen et al. 2000; Brown et al. 2004), the transcriptome profile of the chalazal subregions is not well understood in *B. napus*. The chalazal transcriptome has been analyzed comprehensively across seed development in Arabidopsis by Belmonte et al. (2013) who reported that the CZE possessed highly diverse mRNA populations. Transcripts associated with the biosynthesis of gibberellic acid, abscisic acid, and cytokinin were all preferentially expressed in the CZE; as a result, it was hypothesized that the CZE may act as a communication hub for the seed. The same work had reported the CZSC to be enriched for trehalose biosynthesis, an essential pathway for seed development (Figueroa and Lunn, 2016). Millar et al. (2015) explored the CZSC in *B. napus* and characterized the subregion anatomically and reported increased expression of sugar transporters localized to the CZSC early in seed development. These results suggest that the chalaza of the seed in the Brassicaceae may play important regulatory and transport roles in seed development, incentivizing further exploration into its specific functionality. Though Millar et al. (2015) did highlight the distinctiveness of the CZSC in *B. napus* seed development, no current publications have fully described the transcriptome landscape of all subregions in the early seed of *B. napus*.

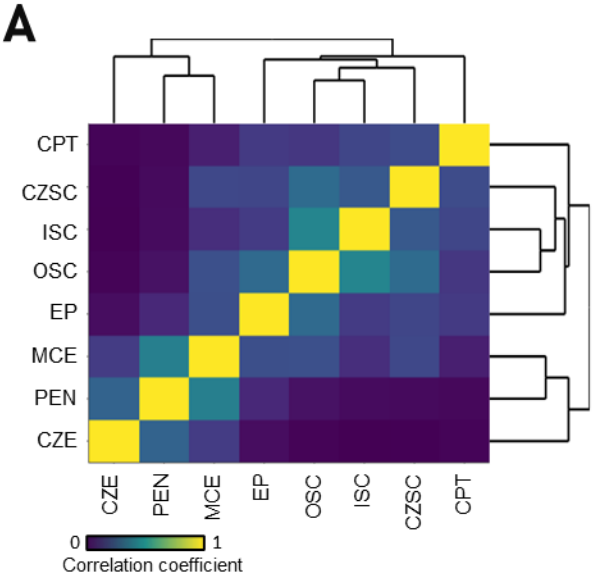
In the current study, we used laser microdissection (LMD) to capture target cells and tissues of the early *B. napus* seed. Combined with next generation sequencing technologies, we profiled the transcriptome landscape of eight seed subregions and validated our gene

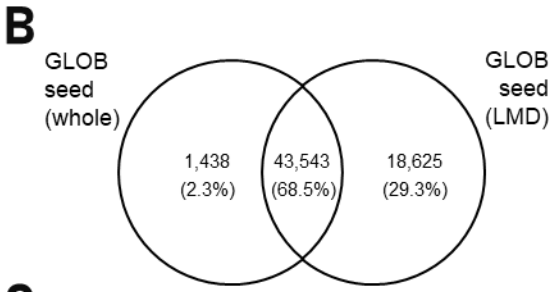
expression analysis with detailed anatomical analysis and qPCR. In particular, we discuss the capacity of the chalazal pole in integrating regulatory signals between maternal and filial subregions.

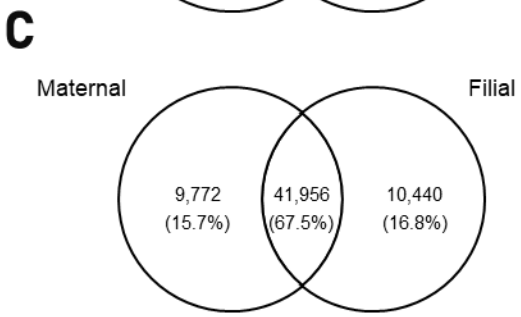
2.3 Results

2.3.1 Subregions of the *B. napus* seed are transcriptomically distinct

To examine the relationships between seed subregion of the *B. napus* globular seed, we performed clustering analysis on genes with FPKM values lower than 10,000 and greater than 10 in at least one of the seed subregions. Clustering shows all three endosperm subregions (MCE, PEN, CZE) clustered together suggesting a similar global transcriptomic profile (Figure 2.2A). Similarly, the three seed coat subregions (ISC, OSC, CZSC) clustered together. In contrast, the EP did not cluster closely with the other filial subregions, nor did the CPT with the maternal subregions.


A
CPT
CZSC
ISC
OSC
EP
MCE
PEN
CZE
CZE
PEN
MCE
EP
OSC
ISC
CZSC
CPT
0
1
Correlation coefficient


B
GLOB seed (whole)
GLOB seed (LMD)
1,438 (2.3%)
43,543 (68.5%)
18,625 (29.3%)


C
Maternal
Filial
9,772 (15.7%)
41,956 (67.5%)
10,440 (16.8%)

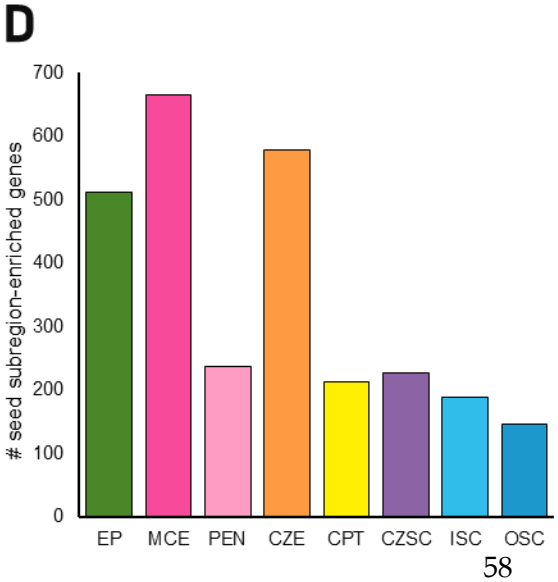

D
seed subregion-enriched genes
700
600
500
400
300
200
100
0
EP
MCE
PEN
CZE
CPT
CZSC
ISC
OSC

Figure 2.2. Comparison of subregion transcriptome profiles of the B. napus globular seed (A) Dissimilarity matrix of seed subregion transcriptomes. **(B)** Comparison of gene activity from whole seeds compared to seed subregions collected via LMD. **(C)** Comparison of gene activity in maternal and filial seed subregions. **(D)** Number of subregion-enriched.

We then compared globular stage RNA-seq data from isolated whole seeds with our LMD subregion dataset to determine the amount of overlap in transcript detection between the two methods (Figure 2.2B). The whole seed dataset detected a total of 44,981 genes at 1 ≤ FPKM, of which more than 97% were also found in the LMD dataset. The LMD dataset detected an additional 18,625 genes (using the same cut-off criteria) in any subregion that were not present in the whole seed RNA-seq expression data (Table S2.1). The use of LMD increased the total number of genes detected in the *B. napus* globular seed 1.4-fold.

Next, we compared gene activity between maternal and filial regions of the seed. Maternal subregions (ISC, OSC, CZSC, CPT) had fewer subregion-enriched genes than the filial subregions (EP, MPE, PEN, CZE) (Figure 2.2C). Together, the maternal and filial subregions shared 41,956 genes detected at 1 ≤ FPKM, with the maternal subregions having 9,772 unique transcripts and the filial having 10,440. Furthermore, our analysis detected more subregion-enriched genes (FPKM≥25 in subregion, FPKM<10 in all other subregions) in the filial subregions, particularly the EP, MCE, and CZE (Figure 2.2D). The EP, MCE, PEN, and CZE had 511, 665, 238, and 578 subregion-enriched genes, respectively while the maternal subregions contained notably less, with the CPT, CZSC, ISC, and OSC having 213, 226, 189, and 146 subregion-enriched genes, respectively (Tables S2.2, S2.3).

2.3.2 Gene Ontology terms characteristic of each subregion

We used gene ontology (GO) term enrichment to predict the biological processes, molecular functions, and subcellular locations of genes likely operative in each subregion of the

B. napus seed. A selection of GO terms enriched in each of the seed subregions is found in Figure 2.3A and a complete list of GO terms and their enrichment is found in Table S4 and Table S5. Biological processes associated with chromatin silencing (log10p = –6.87) and genes associated with developmental organization (xylem and phloem patterning (log10p = –4.17), organ morphogenesis (log10p = –5.53) were all enriched in the EP. In the endosperm region, the MCE was enriched for genes related to gibberellin response (log10p = –3.60), cytokinin response (log10p = –4.72), lipid transport (log10p = –8.23), lipid binding (log10p = –6.17), lipid storage (log10p =–6.17), nutrient reservoir activity (log10p = –7.44), and seed oilbody biogenesis (log10p = –9.78). The PEN was the only endosperm subregion enriched with photosynthesis-related GO terms while CZE was significantly enriched for biological processes associated with brassinosteroid biosynthesis (log10p = –9.95) and homeostasis (log10p = –4.85).

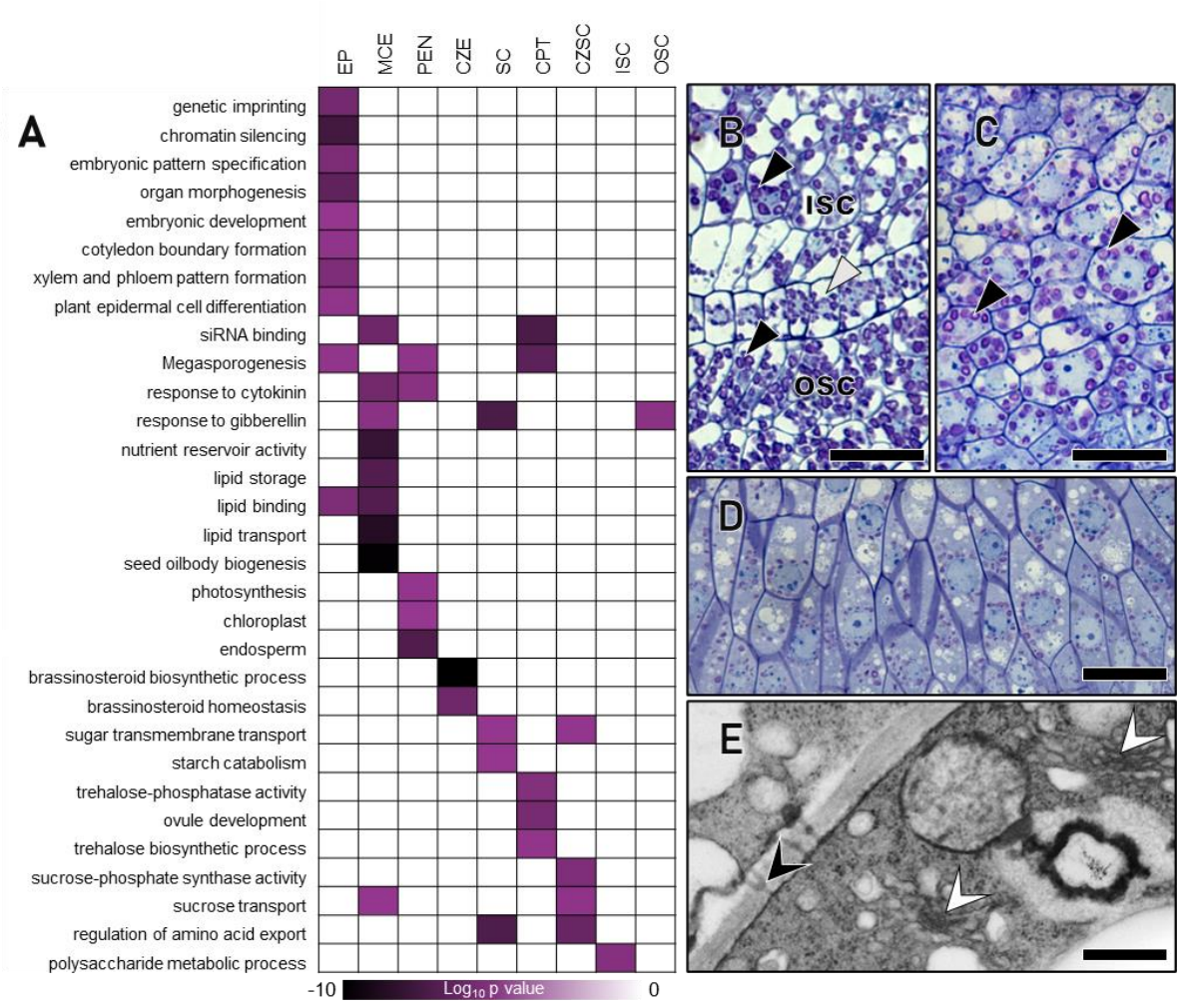


Figure 2.3. Predicted biological processes and anatomical features of the B. napus globular seed (A) Significantly enriched GO terms of subregion-specific gene sets. Darker purple color indicates more statistically enriched GO term. **(B)** Light microscopy showing abundant starch granules (black triangular arrows) in the ISC and OSC separated by the distinct columnar palisade cell layer (white triangular arrows). **(C)** Longitudinal section of the CZSC showing cytoplasmically dense cells and abundant starch granules (black arrowheads). **(D)** Cellular morphology of the CPT showing elongated fusiform cells. **(E)** TEM of CZSC parenchyma cells connected by plasmodesmata (black arrowheads) facilitating transport among the seed coat tissues. Golgi apparatus (white arrowheads) are also abundant in the CZSC. Scale **(B)** = 40 µm, **(C)** = 15 µm, **(D)** = 50 µm, **(E)** = 1 µm**.**

GO term enrichment of maternal subregions showed the CZSC to be enriched for sucrose synthesis (log10p = –4.12) and transport (log10p = –3.41), as well as amino acid transport (log10p = –5.13). The ISC and OSC were significantly enriched for polysaccharide metabolism

(log10p = –3.81) and response to gibberellin (log10p = –3.58), respectively. The CPT, however, was uniquely enriched for trehalose biosynthesis (log10p = –3.26) and shared enrichment of megasporogenesis (log10p = –5.65) with the EP and PEN and siRNA binding (log10p = –6.48). The CPT's genetic profile was uniquely enriched in maternal developmental programs pertaining to early seed development. Maternal subregions were also analyzed using Fuzzy K means clustering (Figure S1, Table S6). GO Terms including ER to Golgi transport (log10p = –9.19), Golgi cisterna membrane (log10p < −5.13), Golgi stack (log10p = –5.88) and vesicle-mediated transport (log10p = –5.88) were all enriched in the maternal region and the plasmodesmata GO term was enriched in all subregions except the CPT. Fuzzy K means clustering analysis output are reported in Tables S6 (gene lists associated with each DP) and S7 (GO summary).

2.3.3 Anatomical features are continuous with enriched GO terms

Histological observations of the maternal subregions of the *B. napus* seed provided additional structural evidence supporting our gene expression analysis. Starch granules were prevalent in the ISC, OSC, and CZSC (Figure 2.3B–C). The presence of these storage reservoirs confirms that all the seed coat subregions contribute to carbon metabolism for the seed. The CPT of the *B. napus* seed is enlarged compared to other documented Brassicaceae species, comprising elongate, cytoplasmically dense, fusiform cells aligned parallel to the chalazal pole. Golgi complexes were ubiquitous throughout the CZSC parenchyma (Figure 2.3E), reinforcing the necessity of carbon and amino acid transport that were enriched in this subregion. While endomembrane system GO terms were not significantly enriched in any specific subregion, Golgi and subcellular endoplasmic reticulum networks were observed across all subregions (Figure S1). To expand this transportation network further, we identified plasmodesmata

connections within the CZSC parenchyma (Figure 2.3D). The plasmodesmata of the CZSC were organized as singular, twinned, and Y-shaped structures (Figure 2.3E).

2.3.4 Predicted regulatory networks underlying the *B. napus* seed

Our network analysis predicted the underlying transcriptional circuitry operative in regions and sub-regions of the *B. napus* globular seed (Figure 2.4). We confirmed the relative expression of select transcription factors and predicted targets identified within GO terms using qPCR (Figure S2). Predicted TF-promoter interactions identified a putative transcriptional circuit in mRNAs accumulating in the maternal region of the seed (Figure 2.4A). MADS box, bHLH, and MYB TFs were predicted to control biological processes including seed coat development, anthocyanin biosynthesis, auxin and brassinosteroid response, and vascular development. Additional analysis of the CZSC subregion revealed a PLATZ TF-promoter interaction and biological processes associated with amino acid and sugar transport, peptide hormone binding, and proanthocyanidin biosynthesis (Figure 2.4B).

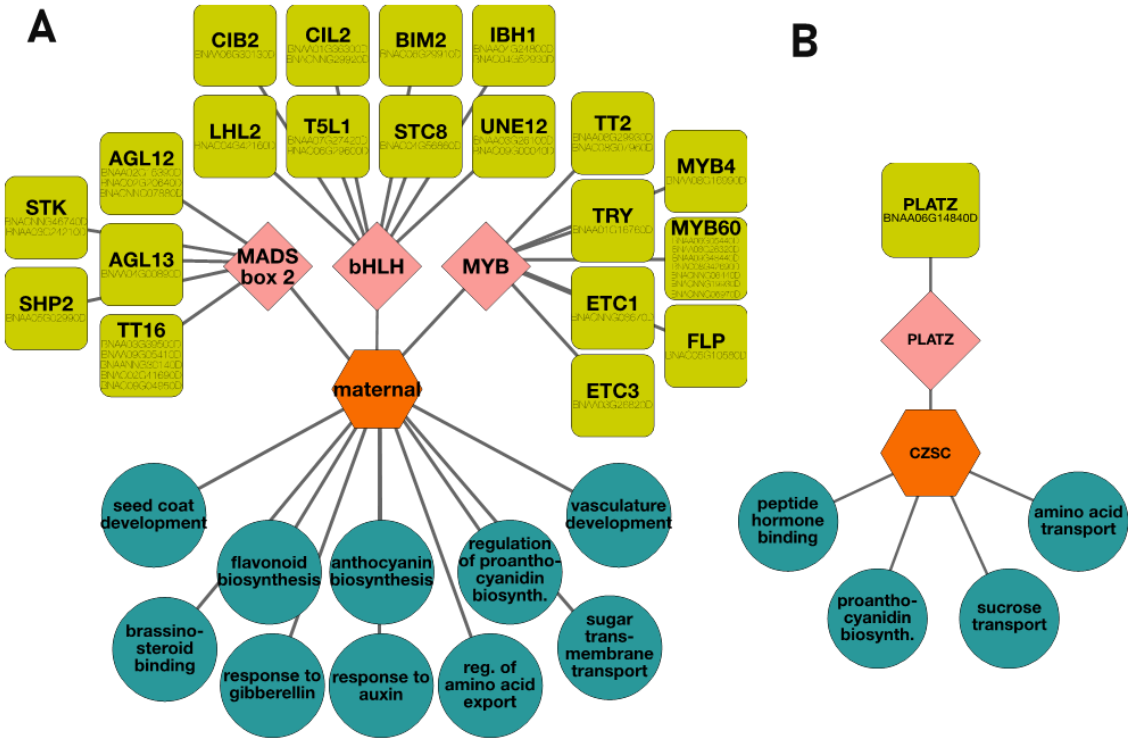


Figure 2.4. Predicted transcription factor modules of the B. napus globular seed (A) A MADS, bHLH, and MYB transcription factor module predicted to be operative within maternal subregions. **(B)** Predicted CZSC-specific PLATZ transcription factor module. Networks predict the regulation of biological processes (GO terms, teal circles) by transcription factors (green rounded squares) that bind to DNA sequence motifs (pink hexagons) within a co-expressed gene set.

In the filial subregions, we identified a putative FUS3 regulatory circuit in the embryo proper (Figure S3A). FUS3 was predicted to regulate cell proliferation, embryonic development/patterning, and cotyledon boundary formation in addition to chromatin silencing and DNA methylation. Two homeologs of the putative seed developmental TF, MINISEED3 showed PEN-preferred activity, binding to the WRKY DNA sequence motif in the regulation of endosperm development, photosynthesis, DNA methylation and in the response to cytokinin. Our network analysis output is ap-pended in Table S8. We validated the subregion-enrichment

of TT2, TT4, TT5, PLATZ, and SWEET15 transcripts, to confirm that our predictive networks can accurately determine transcriptional circuitry (Figure S2).

2.4 Discussion

2.4.1 Subregion-level resolution uncovers greater complexity of the *B. napus* seed transcriptome

Biological processes underlying seed development are compartmentalized into regions and subregions that each contribute to the making of a seed. Here, we present the transcriptome landscape of eight filial and maternal subregions of the *B. napus* globular seed using a combination of LMD and RNA sequencing. Given LMD is a powerful tool to enhance the resolution of gene expression datasets beyond whole-tissue analyses (Schon and Nodine 2017), our data revealed subregion-enriched transcriptomic partitioning of biological processes at the interface between filial and maternal regions. Additionally, the poorly characterized CPT of the maternal region was found to be anatomically and transcriptomically distinct from neighboring CZSC and CZE subregions providing unique insight into the putative function of the CPT in the *B. napus* seed.

2.4.2 The micropylar pole of the *B. napus* seed exhibits early transcriptional activation of fatty acid metabolism

Genes preferentially expressed in the EP were associated with developmental patterning, epigenetic modification, vascular and epidermal development, and fatty acid biosynthesis. Arabidopsis seeds exhibit similar transcriptional profiles of the EP (Belmonte et al. 2013) with significant enrichment in patterning and morphogenesis during the globular stage. Despite this, the MCE in *B. napus* is dissimilar to Arabidopsis globular seeds with significant accumulation of

transcripts associated with lipid storage, transport, and seed oilbody biogenesis. Belmonte et al. (2013) found these GO terms to be significantly enriched in Arabidopsis in the EP, MCE, PEN, and CZE during maturation rather than in morphogenesis. This dissimilarity may imply *B. napus* deposits lipid storage much earlier than in Arabidopsis.

Our data reveal FUS3 (FUSCA3) as a putative regulator in the globular stage embryo of *B. napus*. FUS3 is canonically associated with fatty acid biosynthesis in the seed maturation program of Arabidopsis (Wang et al. 2014) and is expressed highly in the maturing cotyledons of soybean (Zheng and Perry 2014), though our data suggest that the regulatory action of FUS3 is present earlier in *B. napus*. The activation of fatty acid metabolic networks early in the filial micropylar pole contrasts with the regulatory networks active in the maternal subregion, in which regulators that repress fatty acid biosynthesis and favor pigment accumulation, such as TRANSPARENT TESTA2, are dominant (Wang et al. 2014). Thus, early activation of FUS3 may contribute to the high oil phenotype observed in *B. napus* seeds (Rahman 2013).

2.4.3 Hormone biosynthesis and response are compartmentalized in subregions of the *B. napus* seed

The MCE and PEN are identified as hubs of cytokinin response at the RNA level in the *B. napus* seed. MINISEED3 has been shown to be an essential regulator of endosperm development in A. thaliana (Luo and Dennis 2005) and here it is predicted to be a potential regulator of cytokinin response in the PEN of *B. napus* (Figure S3B). Similarly to Arabidopsis, cytokinin biosynthesis genes are expressed primarily in the CZE (Belmonte et al. 2013). The functional significance of separating aspects of cytokinin biosynthesis from response among the endosperm subregions is unclear, but given the position of the CZE at the filial-maternal

interface, it is possible that the CZE acts as a hub for cytokinin signaling that could integrate maternal and filial signals to regulate overall endosperm development.

Our data analysis also provides evidence into the transcriptional separation of brassinosteroid synthesis in the CZE, and response in the maternal subregions. We identified several possible regulators of brassinosteroid response in the maternal subregions, including the BIM2, IBH1, and CIL2 bHLH transcription factors, which are involved in the hierarchical modulation of brassinosteroid response in Arabidopsis vegetative tissues (Ikeda et al. 2012). In Arabidopsis, seed growth and filling are contingent on proper BR signaling (Jiang et al. 2013). Interestingly, Arabidopsis seeds do not show preferential accumulation of transcripts associated BR synthesis or signaling in the CZE (Belmonte et al. 2013), which further individualizes the chalazal pole of the *B. napus* seed. We therefore hypothesize that the chalazal pole could be the interface through which brassinosteroid homeostasis, and subsequently seed growth and filling, is maintained between the maternal and filial generations of the *B. napus* seed.

2.4.4 The chalazal seed coat of *B. napus* is anatomically and transcriptomically specialized for nutrient transport

In our transcriptomic dissection of *B. napus* seed subregions, cells and tissues of the CZSC were active in carbon metabolism, including sucrose biosynthesis and transport. While the accumulation of starch in all maternal subregions provides additional evidence into the role of these cells and tissues in sugar metabolism and nutrient reservoir activity, their distinct anatomical features suggest subregions like the CZSC may preferentially function to exchange nutrients from the vascular tissues of the plant. The abundance of plasmodesmatal connections and extensive endomembrane networks supports the role of the CZSC in regulating nutrient mobilization to surrounding seed subregions while the ISC and OSC may have evolved to

maintain structural integrity around the seed. Previous evidence suggests the seed coat to be integral in sugar transport from the seed coat to the embryo (Chen et al. 2015), and the CZSC to be vital in phosphate mobilization to the embryo (Vogiatzaki et al. 2017). SWEET15 transcripts have been previously shown by Chen et al. (2015) to accumulate in the endosperm of Arabidopsis seeds early in development and is predominant in the seed coat and the micropylar pole later in development. In contrast, our data indicate that SWEET15 transcripts accumulate primarily in the CZSC of the globular seed in *B. napus*. The enrichment of amino acid export, sucrose transport/biosynthesis, and polysaccharide metabolism genes further indicates the importance of the maternal subregions in providing essential nutrients to the *B. napus* seed. Based on our predictive networks, an uncharacterized PLATZ TF represents an ideal target for the modulation and optimization of nutrient transport within the seed.

2.4.5 Recruitment of AGAMOUS-like transcription factors and regulators of epidermal development in the *B. napus* seed coat

Our network analysis predicts STK, SHP2/AGL5, AGL12, and AGL13 as putative regulators of maternal seed development in *B. napus* alongside TT16/ABS. TT16 is a known regulator of proanthocyanidin accumulation in the endothelium of the Arabidopsis seed coat (Debeaujon et al. 2003), and forms a regulatory circuit with STK and SHP2 to regulate fertilization and seed development (Mizzotti et al. 2012; Ehlers et al. 2016). Thus, we can use our network analysis to identify predicted regulators of early maternal seed development in *B. napus*. The roles of AGL12 and AGL13 in seed development are yet to be defined, though they may be involved in growth regulation of the seed coat, as AGL12 has been identified as a regulator of cell proliferation in Arabidopsis roots (Tapia-López et al. 2008). Functional characterization of AGL12 and AGL13 should determine if these transcription factors act alongside other

AGAMOUS-LIKE proteins to regulate proanthocyanidin accumulation and reproductive development, or if they regulate other aspects of growth and the determination of seed size in *B. napus*.

Core aspects of epidermal cell fate specification are conserved in the regulation of pigment accumulation in the *B. napus* seed coat, including potentially new or unique mechanisms governing seed coat development not previously studied in Arabidopsis seeds. TRY, ETC1, and ETC3, which work in concert in Arabidopsis to specify epidermal cell fates (Kirik et al. 2003; Tominaga et al. 2008; Pesch and Hülskamp 2011; Pesch et al. 2014) are predicted regulators of seed coat development and proanthocyanidin biosynthesis in our dataset. Our regulatory network analysis also predicts MYB60, which acts to regulate anthocyanin biosynthesis in lettuce leaves (Park et al. 2008), as a potential regulator of maternal seed development. It is possible that these epidermal regulators could be involved in the regulation of pigment biosynthesis and accumulation in the seed coat. While TRY has been identified as a potential regulator of pigment accumulation in the early Arabidopsis seed coat (Khan et al. 2014), neither ETC1 nor ETC3 are expressed in Arabidopsis seed development (Belmonte et al. 2013; Klepikova et al. 2016). Given the presence of multiple homologues of both ETC1 and ETC3 in *B. napus*, it is possible that some homologues may have diverged and show greater involvement in seed coat development than their Arabidopsis counterparts.

2.4.6 Origin of the chalazal proliferating tissue

The CPT of the *B. napus* globular seed consists of elongate fusiform cells concentrated intramarginally to the CZSC. In Arabidopsis, the CPT is heavily compressed and diminutive, consisting of only a few cells persisting from the degraded nucellus (Nguyen et al. 2000). In contrast, the *B. napus* seed develops a conspicuous and organized CPT that is not immediately

crushed by the proliferating CZE. Additionally, the nucellus is entirely absent in the *B. napus* globular seed, with no histological evidence of the nucellar lysate seen medial to the CZE and CPT in Arabidopsis and Stanleya pinnata (Brown et al. 2004), thus the seeds of *B. napus* are tenuinucellate. In our GO term subregion-enriched analysis, the CPT was enriched for megasporogenesis and megasporocyte differentiation, suggesting common functionality and developmental origin with the chalazal nucellus, consistent with that observed by Schulz and Jensen (1971) in Capsella bursa-pastoris. While the ovule of *B. napus* is campylotropous, post-fertilization the ovule elongates and recurves, placing the micropyle adjacent to the chalaza in an amphitropous position. The embryo follows suit with the micropylar pole, while the chalazal nucellus remains localized to the chalaza. While the embryo (at the micropylar end of the seed) is well into seed development, there appears to be transcriptional holdover of ovule development that persists at the chalazal end of the seed during the globular stage, potentially arising from activity of the chalazal nucellus remnant.

2.4.7 Trehalose biosynthesis is compartmentalized to the CPT

The CPT was enriched for genes associated with trehalose biosynthesis. Trehalose is well documented to function in carbon metabolism to modulate sucrose concentration in plant tissues (Figueroa and Lunn 2016). Trehalose biosynthesis is also partially compartmentalized to the chalazal end of the Arabidopsis seed (Vandesteene et al. 2010), and is known to be essential to seed development where perturbations in trehalose metabolism leads to embryo lethality (Eastmond et al. 2002). The CPT was the only *B. napus* seed subregion exhibiting specific enrichment of genes associated with trehalose metabolism. In contrast, Belmonte et al. (2013) found trehalose biosynthesis to be enriched only in the CZSC in Arabidopsis seeds later in development. The CPT of *B. napus* may adopt a more specialized role in trehalose metabolism in

the seed than observed in Arabidopsis. In conjuction with the CPT's distinctive morphology in *B. napus*, it is possible that the CPT orchestrates trehalose metabolism and plays integral roles in seed development and in defining the source-sink balance between the maternal and filial tissues of the seed.

2.4.8 Conclusion

The transcriptome landscape of the early *B. napus* seed reveals transcriptional regulators of seed development at subregion-level resolution. We provide evidence for subregion-enriched developmental processes, phytohormone signaling, and carbon metabolism between maternal and filial regions. Moreover, the CZE, CPT, and CZSC are strongly compartmentalized in their contributions to early seed development and further exploration of the *B. napus* chalazal pole in relation to other maternal and filial subregions across developmental time should uncover additional biological processes required to make an oilseed.

2.5 Materials and Methods

2.5.1 Plant materials and Growth

Brassica napus cv. Topas plants were grown in growth chambers under long day conditions (16 h light, 8 h dark, 22°C, RH 50%–70%) and flowers tagged and hand-pollinated (Chan and Belmonte, 2013). Globular stage seeds were collected 7 d post fertilization. Flower pollination and silique harvest were performed between 15:00 and 17:00 to minimize time of day effects on gene activity.

2.5.2 Light microscopy

Tissue fixation and embedding were performed as in Chan and Belmonte (2013). Histological sections were cut at 3 µm on a Leica RM2245 rotary microtome with a steel blade,

and sections mounted on glass slides. Slides were stained to with periodic acid-Schiff's, and toluidine blue O. Following staining, coverslips were mounted and sealed using Cytoseal™ 60 (Richard-Allan Scientific™). A Leica DM2500 light microscope equipped with a Leica DFC425 camera using LAS3.7 software was used to visualize the sections.

2.5.3 Transmission electron microscopy

Tissue fixation and processing were performed as in Chan and Belmonte (2013). Sections were cut at 90 nm on a Reichert-Jung Ultracut ultramicrotome using a Diatome diamond knife, and mounted on copper grids. The sections were visualized using a Hitachi H-7000 transmission electron microscope and captured using the AMT Image Capture Engine (version 601.384).

2.5.4 Tissue processing and collection for RNA-seq

To evaluate tissue-enriched gene activity in the *B. napus* seed, we used laser microdissection to capture individual subregions. All tissue processing and collection was performed under RNase-free conditions. Siliques of globular-stage seeds were cut at 50 mm fragments and fixed in 1:3 (v/v) glacial acetic acid: 85% ethanol for 24 h at 4°C. Tissues were then dehydrated in a graded ethanol series, followed by a transition infiltration with xylenes at 4°C and gradually infiltrated with paraffin at 60°C. Samples were incubated for less than 8 h at 60°C prior to embedding in 100% paraffin wax (McCormick Scientific, Paraplast®Plus™).

2.5.5 Sectioning and laser microdissection

Serial sections (7 µm sections, Leica RM2245 rotary microtome) were mounted on polyethylene napthalate (PEN) membrane slides (Leica Microsystems). Slides were

deparaffinized in xylenes for 1 min and seed subregions were isolated using the Leica Laser Microdissection 7000 system.

2.5.6 RNA isolation

Total RNA was extracted from the laser-microdissected subregions using the Ambion® RNaqueous micro kit and treated with Qiagen RNAase-free DNase to remove genomic DNA. RNA quantity was evaluated using the Quant-iT™ RiboGreen™ RNA Assay kit (Invitrogen™) and quality was evaluated using the Agilent 6000 Pico LabChip® and the Agilent 2100 bioanalyzer (Agilent Technologies, USA). A minimum RIN of 3 was required for further cDNA synthesis and library preparation (Belmonte et al. 2013).

2.5.7 Library preparation and sequencing

First and second cDNA synthesis was performed according to a high-throughput RNA-seq protocol (Kumar et al. 2012). Fragmentation was then carried out using the Covaris™ M220 Focused-ultrasonicator™ in Covaris™ M220 microTUBEs. Samples were then brought down to 5 µL via rotary evaporation. Library preparation was performed using the NuGen® Ovation® Ultralow DR Multiplex System. Total library quality and quantity were determined using the Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen), qPCR, and the High Sensitivity DNA Analysis Kit (Agilent) on the Agilent 2100 Bioanalyzer. Libraries were pooled and size-selected (350–600 bp) on the E-Gel® system (Invitrogen™). Samples were then sequenced on the Illumina® HiSeq™ 2000 platform at Genome Québec. Whole seed RNA was extracted from whole globular stage seeds using PureLink® Plant RNA Reagent (Ambion®). Library preparation was performed as in (Kumar et al. 2012) using NEXTFlex ChIP-seq barcodes (Bioo Scientific®).

2.5.8 Data analysis

Libraries were sequenced on the Illumina® HiSeq™ 2000 platform (150 bp SE). Raw sequencing reads are deposited at GEO (GSE120360). Trimming and QC of sequence reads was performed using Trimmomatic 0.36 (Bolger et al. 2014) (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 HEADCROP:9 LEADING:9 TRAILING:9 SLIDINGWINDOW:4:30 MINLEN:20). Surviving reads, aligning and mapping percentages are reported in Table S9. Reads were aligned to the Darmor-bzh genome (Chalhoub et al. 2014) using HISAT2 (Kim et al. 2015). Normalization and differential expression analyses were performed with multi-read correction using the Cufflinks suite (Trapnell et al. 2010). Quality control and global comparisons of transcriptome libraries were performed using the Cumme(R)bund suite (Goff et al. 2012). Detection cut-off for transcripts was FPKM ≥ 1, while high-abundance transcripts representing ribosomal, chloroplast, or mitochondrial genes (FPKM ≥ 10,000 in any sample), and low-abundance transcripts (FPKM < 1 across all samples) were filtered out to facilitate downstream clustering and enrichment analyses.

Subregion-enriched transcripts were identified (FPKM > 25 in the sample, and FPKM < 10 in all other samples) and enrichment analysis was performed on these transcript populations using SeqEnrich (Becker et al. 2017). GO enrichment analysis using SeqEnrich analyzes all parent and child GO terms and determines significance by the proportion of genes in the query list relative to the amounts of genes within the umbrella GO term, which enables the analysis to be used with polyploid plants, such as *B. napus*. Heatmaps and dendrograms were generated in R using ggplot2 (Wilkinson 2011) and heatmaply (Galili et al. 2018). To uncover regulatory transcriptional circuits in co-expressed genesets, we performed a network analysis as outlined in Becker et al. (2017). Enrichment analysis of DNA sequence motifs and

GO terms ($P < 0.001$) are used to predict transcriptional regulatory circuits, with novel and relevant networks highlighted in this work. Networks were analyzed in Cytoscape (Shannon et al. 2003). Transcript accumulation of transcription factors and selected targets were confirmed using qPCR (NEB Luna® Universal qPCR master mix; primers can be found in Table S10). Significant differences in transcript abundance were determined by ANOVA and Student's t-test. Genes were assigned to dominant patterns (DPs) of transcript accumulation using fuzzy K means (FKM) clustering (R; http://cran.r-project.org/web/packages/cluster/cluster.pdf; implication FANNY).

2.6 Supplemental Figures

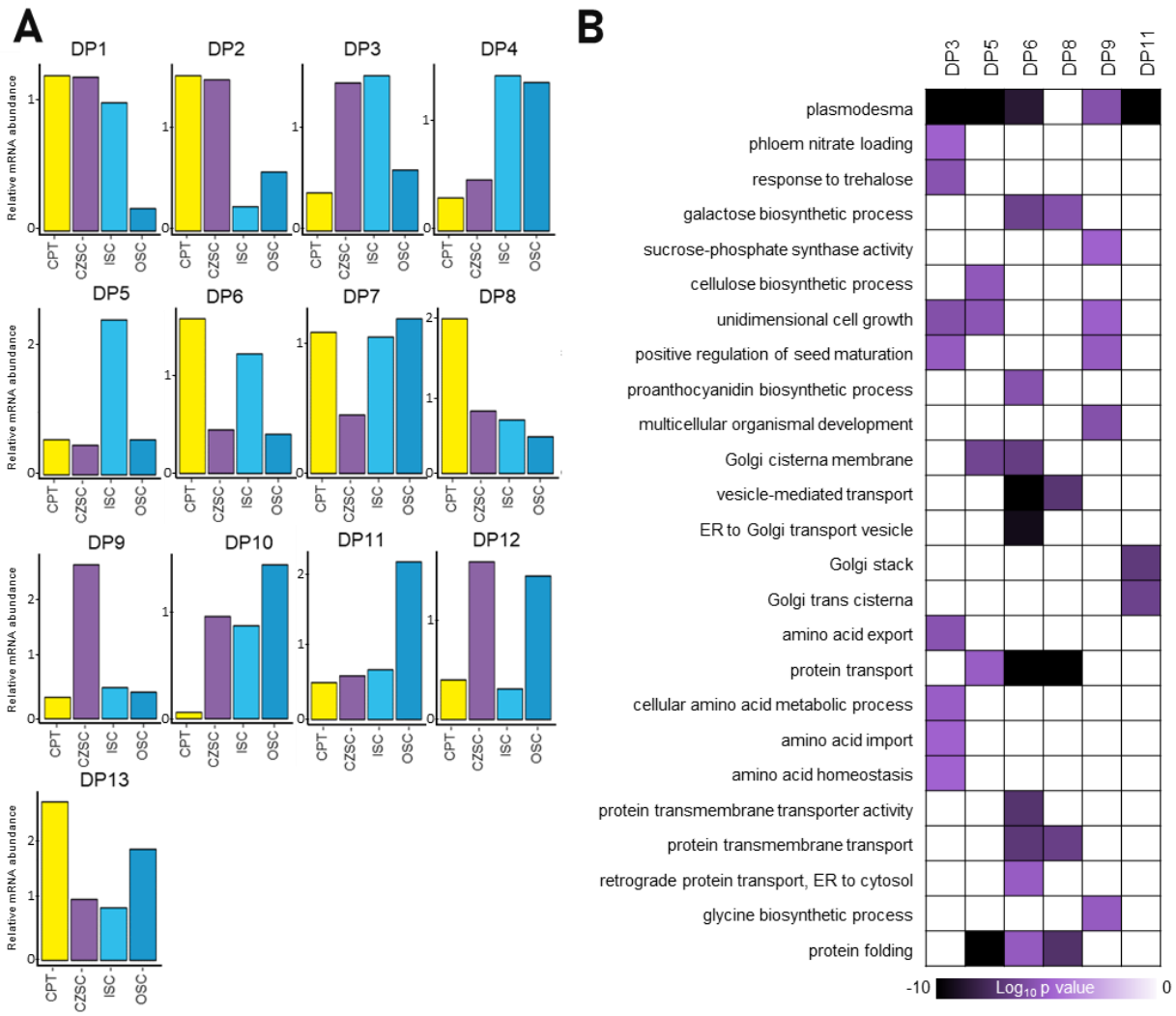
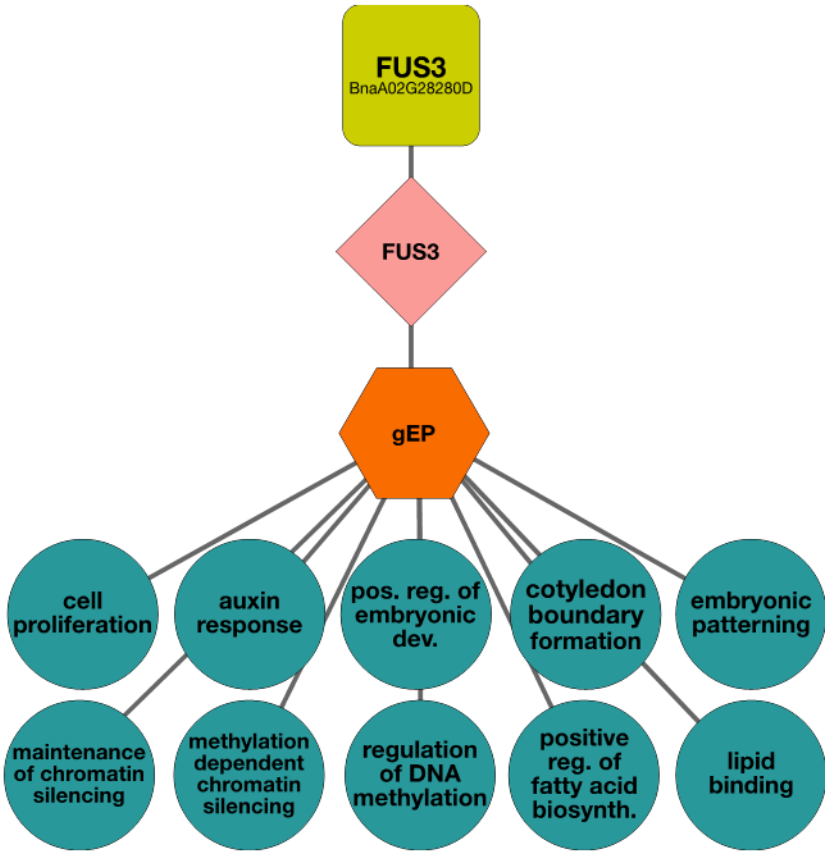


Figure S2.1. Gene expression and GO term analysis of maternal subregions in the B. napus globular seed (A) Fuzzy K means clustering analysis of maternal subregions (data from the Fuzzy K means clustering areantilog2 transformed). **(B)** GO term enrichment of selected fuzzy K means clustering patterns (DPs).

A


FUS3
BnaA02G28280D
FUS3
gEP
cell proliferation
auxin response
pos. reg. of embryonic dev.
cotyledon boundary formation
embryonic patterning
maintenance of chromatin silencing
methylation dependent chromatin silencing
regulation of DNA methylation
positive reg. of fatty acid biosynth.
lipid binding

B

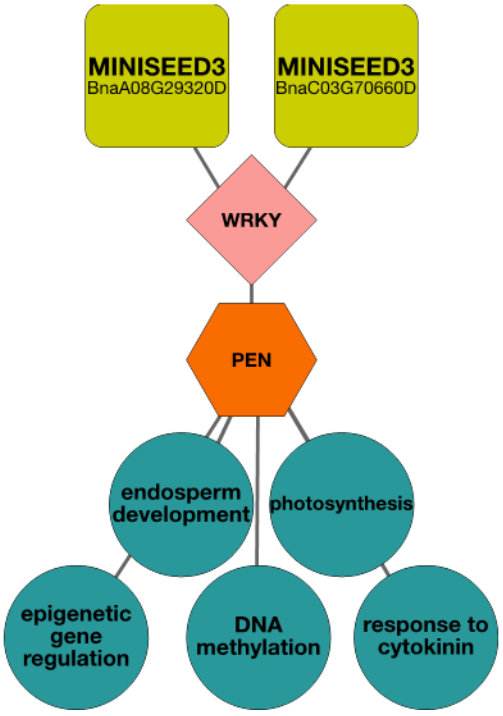

MINISEED3
BnaA08G29320D
MINISEED3
BnaC03G70660D
WRKY
PEN
endosperm development
photosynthesis
epigenetic gene regulation
DNA methylation
response to cytokinin

Figure S2.2. Predictive transcription factor networks of filial subregions of the B. napus seed (A) A predicted EP-specificFUS3network. **(B)** A predicted PEN-specific WRKY transcription factor network.

2.6 Contribution of Authors

J.L.K., D.K., and M.G.B. performed experiments. D.K., D.J.Z., M.G.B. and J.L.K. performed data analysis. D.J.Z., D.K., J.L.K. and M.F.B. wrote the paper. All authors read and approved the paper.

2.7 References

Becker MG, Walker PL, Pulgar-Vidal NC, Belmonte MF (2017) SeqEnrich: A tool to predict transcription factor networks from co-expressed Arabidopsis and *Brassica napus* gene sets. PLoS ONE 12: e0178256

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci USA 110: E435–44

Besnard J, Zhao C, Avice J, Vitha S, Hyodo A, Pilot G, Okumoto S (2018) Arabidopsis UMAMIT24 and 25 are amino acid exporters involved in seed loading. J Exp Bot 69: 5221–5232

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–20

Brown RC, Lemmon BE, Nguyen H (2004) Comparative anatomy of the chalazal endosperm cyst in seeds of the Brassicaceae. Bot J Linn Soc 144: 375–394

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Silva D, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier MC, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee TH, Thi VH, Chalabi S, Hu Q, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CH, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury JM, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C,

Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345: 950–953

Chan A, Belmonte M (2013) Histological and ultrastructural changes in canola (*Brassica napus*) funicular anatomy during the seed lifecycle. Botany 91: 671–679

Chan AC, Khan D, Girard IJ, Becker MG, Millar JL, Sytnik D, Belmonte MF (2016) Tissue-specific laser microdissection of the *Brassica napus* funiculus improves gene discovery and spatial identification of biological processes. J Exp Bot 67: 3561–71

Chen LQ, Lin IW, Qu XQ, Sosso D, McFarlane HE, Londoño A, Samuels AL, Frommer WB (2015) A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. Plant Cell 27: 607–619

Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: Regulation of differentiation and role in seed development. Plant Cell 15: 2514–2531

Eastmond PJ, van Dijken AJH, Spielman M, Kerr A, Tissier AF, Dickinson HG, Jones JDG, Smeekens SC, Graham IA (2002) Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. Plant J 29: 225–235

Ehlers K, Bhide AS, Tekleyohans DG, Wittkop B, Snowdon RJ, Becker A (2016) The MADS box genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in Arabidopsis thaliana. PLoS ONE 11: e0165075

FAOSTAT (2014) Available from: http://www.fao.org/faostat/en/#compare

Figueroa CM, Lunn JE (2016) A tale of two sugars: Trehalose 6-phosphate and sucrose. Plant Physiol 172: 7–27

Fore SR, Porter P, Lazarus W (2011) Net energy balance of small-scale on-farm biodiesel production from canola and soybean. Biomass Bioenerg 35: 2234–2244

Galili T, O'Callaghan A, Sidi J, Sievert C (2018) Heatmaply: An R package for creating interactive cluster heatmaps for online publishing. Bioinformatics 34: 1600–1602

Goff LA, Trapnell C, Kelley D (2012) CummeRbund: Visualization and exploration of Cufflinks high-throughput sequencing data. R Packag. Version 2.2

Ikeda M, Fujiwara S, Mitsuda N, Ohme-Takagi M (2012) A triagnostic basic helix-loop-helix system regulates cell elongation in Arabidopsis. Plant Cell 24: 4483–4497

Jiang WB, Huang HY, Hu YW, Zhu SW, Wang ZY, Lin WH (2013) Brassinosteroid regulates seed size and shape in Arabidopsis. Plant Physiol 162: 1965–77

Khan D, Millar JL, Girard IJ, Belmonte MF (2014) Transcriptional circuitry underlying seed coat development in Arabidopsis. Plant Sci 219–220: 51–60

Khan D, Millar JL, Girard IJ, Chan A, Kirkbride RC, Pelletier JM, Kost S, Becker MG, Yeung EC, Stasolla C, Goldberg RB, Harada JJ, Belmonte MF (2015) Transcriptome atlas of the Arabidopsis funiculus: A study of maternal seed subregions. Plant J 82: 41–53

Kim D, Langmead B, Salzberg SL (2015) HISAT: A fast spliced aligner with low memory requirements. Nat Methods 12: 357–360

Kirik V, Simon M, Huelskamp M, Schiefelbein J (2003) The ENHANCER OF TRY AND CPC1 gene acts redundantly with TRIPTYCHON and CAPRICE in trichome and root hair cell patterning in Arabidopsis. Dev Biol 268: 506–513

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA (2016) A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. Plant J 88: 1058–1070

Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN, Sinha NR (2012) A high-throughput method for Illumina RNA-seq library preparation. Front Plant Sci 3: 1–10

Luo M, Dennis E (2005) MINISEED3 (MIN3), a WRKY family gene, and HAIKU2 (IKU2), a Leucine-Rich Repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proc Natl Acad Sci USA 102: 17531–17536

Millar JL, Khan D, Becker MG, Chan A, Dufresne A, Sumner M, Belmonte MF (2015) Chalazal seed coat development in *Brassica napus*. Plant Sci 241: 45–54

Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L (2012) The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. Plant J 70: 409–420

Nguyen H, Brown RC, Lemmon BE (2000) The specialized chalazal endosperm in Arabidopsis thaliana and Lepidium virginicum (Brassicaceae). Protoplasma 212: 99–110

Park JS, Kim JB, Cho KJ, Cheon CI, Sung MK, Choung MG, Roh KH (2008) Arabidopsis R2R3-MYB transcription factor AtMYB60 functions as a transcriptional repressor of anthocyanin biosynthesis in lettuce (Lactuca sativa). Plant Cell Rep 27: 985–994

Pesch M, Dartan B, Birkenbihl R, Somssich IE, Hülskamp M (2014) Arabidopsis TTG2 regulates TRY expression through enhancement of activator complex-triggered activation. Plant Cell 26: 4067–4083

Pesch M, Hülskamp M (2011) Role of TRIPTYCHON in trichome patterning in Arabidopsis. BMC Plant Biol 11: 130

Rahman H (2013) Review: Breeding spring canola (*Brassica napus* L.) by the use of exotic germplasm. Can J Plant Sci 93: 363–373

Schon MA, Nodine MD (2017) Widespread contamination of Arabidopsis embryo and endosperm transcriptome data sets. Plant Cell 29: 608–617

Schulz P, Jensen WA (1971) Capsella embryogenesis: The chalazal proliferating tissue. J Cell Sci 8: 201–27

Schulz P, Jensen WA, Jensen WA (1974) Capsella embryogenesis: The development of free nuclear endosperm. Protoplasma 80: 183–205

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504

Tapia-López R, García-Ponce B, Dubrovsky JG, Garay-Arroyo A, Pérez-Ruíz R V, Kim S-H, Acevedo F, Pelaz S, Alvarez-Buylla ER (2008) An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in Arabidopsis. Plant Physiol 146: 1182–1192

Tominaga R, Iwata M, Sano R, Inoue K, Okada K, Wada T (2008) Arabidopsis CAPRICE-LIKE MYB 3 (CPL3) controls endoreduplication and flowering development in addition to trichome and root hair formation. Development 135: 1335–1345

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511–515

Vandesteene L, Ramon M, Le Roy K, Van Dijck P, Rolland F (2010) A single active trehalose-6-P synthase (TPS) and a family of putative regulatory TPS-like proteins in Arabidopsis. Mol Plant 3: 406–419

Vogiatzaki E, Baroux C, Jung JY, Poirier Y (2017) PHO1 exports phosphate from the chalazal seed coat to the embryo in developing Arabidopsis seeds. Curr Biol 27: 2893–2900.e3

Wang Z, Chen M, Chen T, Xuan L, Li Z, Du X, Zhou L, Zhang G, Jiang L (2014) TRANSPARENT TESTA2 regulates embryonic fatty acid biosynthesis by targeting FUSCA3 during the early developmental stage of Arabidopsis seeds. Plant J 77: 757–769

Wilkinson L (2011) ggplot2: Elegant graphics for data analysis by WICKHAM, H. Biometrics 67: 678–679

Zheng Q, Perry SE (2014) Alterations in the transcriptome of soybean in response to enhanced somatic embryogenesis promoted by orthologs of AGAMOUS-Like15 and AGAMOUS-Like18. Plant Physiol 164: 1365–77

3 Genomic asymmetry of the *Brassica napus* seed: epigenetic contributions of DNA methylation and small RNAs to subgenome bias

Dylan J. Ziegler[1], Deirdre Khan[1], Nadège Pulgar-Vidal[1], Isobel A. P. Parkin[2], Stephen J. Robinson[2], Mark F. Belmonte[1]*

[1]University of Manitoba, Winnipeg, Manitoba

[2]Agriculture and Agri-Food Canada, Saskatoon, Saskatchewan

This chapter is published in Plant Journal as:

Ziegler DJ, Khan D, Pulgar-Vidal N, Parkin IAP, Robinson SJ, Belmonte MF. 2023. Genomic asymmetry of the *Brassica napus* seed: Epigenetic contributions of DNA methylation and small RNAs to subgenome bias. Plant J. DOI:10.1111/tpj.16254

3.1 Abstract

Polyploidy is a persistent phenomenon in angiosperm genome evolution that is hypothesized to have contributed to the diversity of extant flowering plants. *Brassica napus*, one of the world's most important angiosperm oilseed species, originated from the interspecific hybridization of *Brassica rapa* (An) and *Brassica oleracea* (Cn). While the trends of genome dominance in transcriptomics are beginning to emerge, less is known about the epigenetic and small RNA landscapes in polyploids during reproductive development. The seed is the pivotal developmental transition into the new sporophytic generation, and experiences substantial epigenetic modifications over time. Here, we investigated the prevalence of bias in the contexts of DNA methylation and small interfering (si)RNA profiles in both subgenomes (An and Cn), as well as the ancestral fractionated genomes across *B. napus* seed development. We report ubiquitous Cn subgenome bias of siRNA expression and cytosine methylation, with DNA methylation being particularly abundant on gene promoters in the Cn subgenome. Further, we provide evidence that siRNA transcriptional patterns were conserved within the ancestral

triplicated subgenomes of *B. napus*, but not across the An and Cn subgenomes. We discuss how methylation patterns in the *B. napus* seed relate to genes, promoter regions, siRNA loci and transposable elements through the lens of genome fractionation and polyploidization. Taken together we provide evidence for epigenetic regulation selectively silencing the Cn subgenome during seed development, and explore the impact of genome fractionation on the epigenetic components of the *B. napus* seed.

3.2 Introduction

Polyploidy is an integral driving force in land plant evolution, occurring before the divergence of both monocot and dicot lineages and multiple times therein (Soltis & Soltis, 2016; Vekemans et al., 2012). Polyploidy can either arise from aberrant meiosis/mitosis, whole-genome duplications (WGD) or through hybridization events (Tayalé & Parisod, 2013). Early genomic research in Arabidopsis thaliana (L.) Heynh. (Brassicaceae) demonstrated three ancient WGDs in the history of the species as of eudicot divergence (Blanc et al., 2003). Furthermore, the Brassiceae tribe, containing the economically important Brassica genus, is defined by a whole-genome triplication event (WGT) that distinguishes it from Arabidopsis (Camelineae tribe). This complicated genetic history is in fact the orthodox trajectory of plant genome evolution, with WGDs and polyploidization events leading to the speciation and consequent divergence of most angiosperm lineages.

Brassica napus L. is a nascent amphidiploid species in the Brassiceae tribe formed from the interspecific hybridization of *Brassica rapa* L. (ArAr) and *Brassica oleracea* L. (CoCo) approximately 7500–12 500 years ago (Chalhoub et al., 2014). Consequently, *B. napus* is an allopolyploid with two distinct diploid subgenomes therein (AnAnCnCn), both of which are products of the WGT within the Brassiceae. Additionally, genome fractionation and dominance

may change the genomic landscape of a polyploid species with extensive deletions within the submissive genome while the dominant genome remains prominent (Cheng et al., 2018). Many crop plants like Zea mays, Glycine max, Oryza sativa have polyploid ancestry but their extant genomes are heavily fractionated, with only fragments of their progenitors persisting (Messing, 2009; Schmutz et al., 2010; Zhang et al., 2005). In other cases, such as Triticum aestivum, an allohexaploid, one progenitor species is extinct (Matsuoka, 2011). Other recently formed polyploids, as in Tragopogon, have extensive research utility in understanding how allopolyploidization affects genome evolution (Leitch et al., 2004). In the case of Brassica, the ancestral WGT has resulted in the dominance of one least fractionated genome (LF) and two fractionated, submissive genomes herein called most fractionated 1 (MF1) and MF2 that persist in extant species (Liu et al., 2014; Parkin et al., 2014). Few crop plants exist that have undergone recent allopolyploidization with distinguishable subgenomes, and even fewer with ancestrally fractionated genomes within their subgenomes. As a nascent allotetraploid and one of the world's most important oilseeds, *B. napus* is an excellent system to explore the effects of polyploidization on the genomic landscape and, more specifically, how epigenetic factors such as DNA methylation and small RNA (sRNA) signatures differ between subgenomes during reproduction.

It is well documented that extensive epigenetic changes characterize phases of seed development in angiosperms (Grover et al., 2020; Rajkumar et al., 2020). Seed development of *B. napus* (canola) is a complex yet elegant chapter of the plant lifecycle that starts with the unfertilized ovule (OV) and ends with formation of the dry seed (DS; Figure 3.1a). The first phase of seed development results from a double fertilization event that leads to the establishment of the embryo and the endosperm surrounded and protected by the maternally

derived seed coat (Dresselhaus & Franklin-Tong, 2013; Guignard, 1902). Early in the morphogenesis phase, the globular (GLOB) embryo begins to differentiate through programmed cell divisions and establish tissue identities (Smith & Long, 2010; Tykarska, 1976). By the heart (HEART) stage, the shoot and root apical meristems organize and will eventually give rise to the shoot and root systems of the next sporophytic generation of the plant (Tykarska, 1979; also reviewed in the model Arabidopsis in Jenik et al., 2007). The onset of seed maturation begins with substantial cellular proliferation and growth of the embryo to fill the central cavity of the seed, followed by an arrest of proliferation and the accumulation of storage reserves, leading to the mature green (MG) stage of seed development (Leviczky et al., 2019; Tykarska, 1980). Maturation concludes with the rapid desiccation of the seed in preparation for dormancy at the DS stage. In the current literature, subgenome bias in DNA methylation and small interfering (si)RNA expression is documented in *B. napus* vegetative tissues, with emerging research revealing biases also occurring during seed development. Shen et al. (2015) noted siRNA expression biased towards the Cn subgenome in vegetative tissues, but details of subgenome biases in reproductive tissues of Brassica are still emerging. Further, Lu et al. (2012) uncovered that siRNAs originating from transposable elements (TEs) contribute to genome balance in polypoid tissues such as the endosperm. Edger et al. (2017) also demonstrated that DNA methylation exhibits subgenome dominance in early generation polyploids in vegetative tissue and in infertile floral whorls.

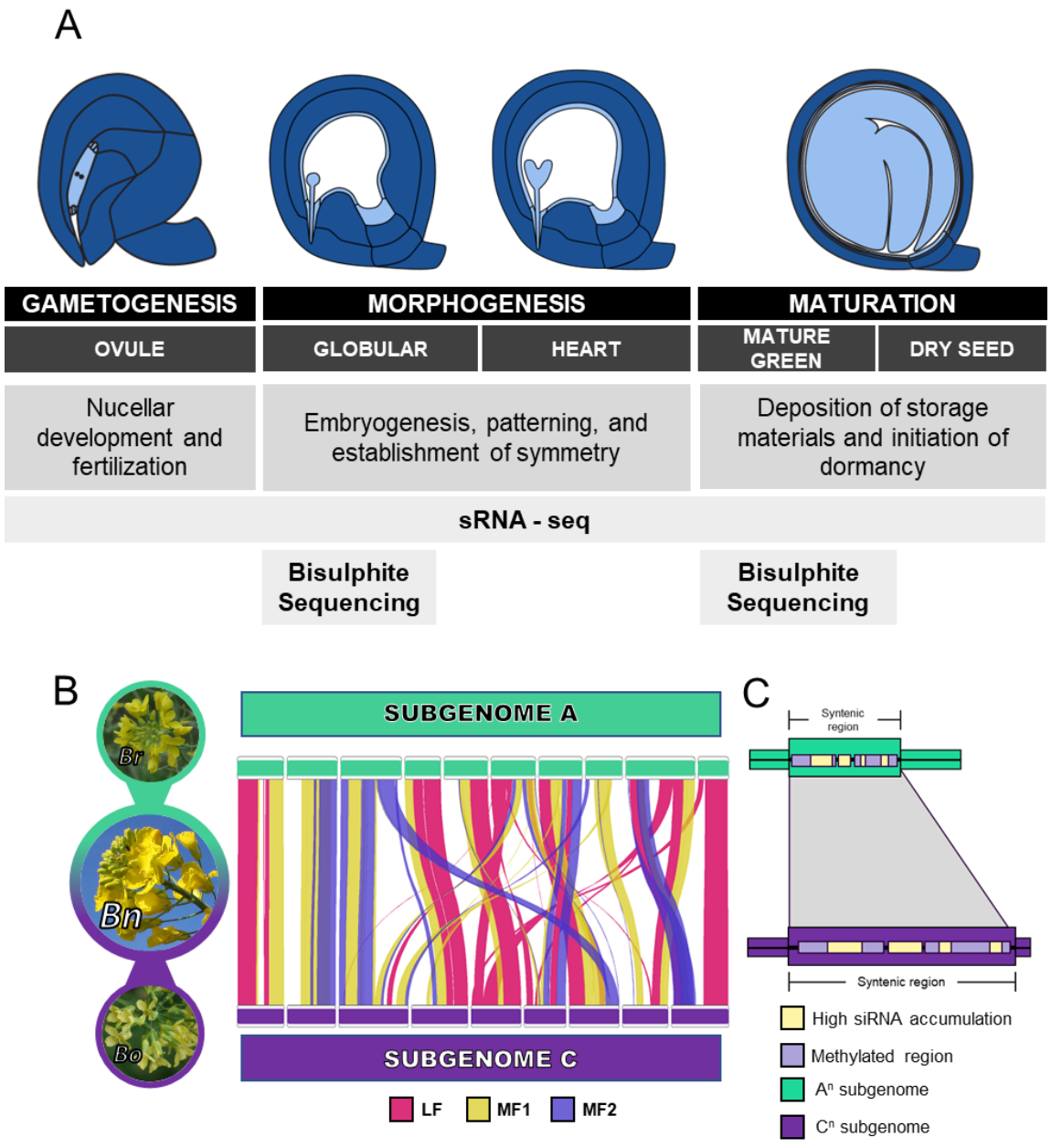


Figure 3.1. Hypothetical model of epigenetics underpinning seed development in the *B. napus* seed. (A) Seed development can be divided into five discrete stages from the beginning of gametogenesis to the end of maturation (ovule (OV), globular (GLOB), heart (HEART), mature green (MG), and dry seed (DS)). We completed sRNA-seq for all five of these stages and bisulphite sequencing for the GLOB and MG stages, representing pivotal developmental transitions in the initiation of morphogenesis and maturation respectively. **(B)** Synteny map showing the largest continuous syntenic block of each chromosome within the amphidipoid *B. napus* genome, progenitor species being *Brassica rapa* (Br) contributing the A^n subgenome, and *Brassica oleracea* (Bo) contributing the C^n subgenome. **(C)** Hypothetical model of epigenetic conservation – broad syntenic regions of the genome indicate high conservation between the

two subgenomes. Conservation of these regions may lead to similar epigenetic architecture in reproduction.

DNA methylation in plants occurs in every cytosine context (CG, CHG, CHH, where H = A, T or C), with each context exhibiting notably different degrees of methylation and pathways of establishment and maintenance (Matzke & Mosher, 2014). Previous work has profiled the methylation signatures of DNA during seed development in Arabidopsis (Bouyer et al., 2017; Lin et al., 2017; Narsai, Gouil, et al., 2017), maize (Li et al., 2014; Lu et al., 2015), Ricinus communis (Xu et al., 2016), O. sativa (Xing et al., 2015) and Glycine (Davis-Richardson et al., 2016; Lin et al., 2017), and OV development in Gossypium (Osabe et al., 2014). However, studies of DNA methylation in allopolyploid plants such as Gossypium hirsutum, Glycine dolichocarpa, Triticum and Erythranthe perigrinus have primarily focused on vegetative tissues (Bird et al., 2021; Chalhoub et al., 2014; Coate et al., 2014; Edger et al., 2017; Gardiner et al., 2015; Li et al., 2014; Li et al., 2016; Li et al., 2017; Song et al., 2015), and thus our understanding of how DNA methylation dynamics compare between progenitor subgenomes in allopolyploids during reproduction and seed development is still relatively unknown. Furthermore, no current studies have endeavored to study DNA methylation patterns of seed development in the nascent allotetraploid *B. napus*.

The sRNAs are important regulators of the genome and epigenome in plants. The most studied sRNAs, micro-RNAs (miRNAs), are involved in the regulation of gene expression via post-transcriptional silencing. miRNAs arise from a single-stranded precursor that forms a secondary hairpin structure after transcription (Wang et al., 2019). Unlike miRNAs, siRNAs do not act on mRNA alone and can induce changes within the epigenome itself. siRNAs differ from miRNA in their mode of biogenesis, wherein siRNAs are transcribed and a

complementary strand is synthesized by RNA-dependent RNA polymerase to form a double-stranded siRNA precursor (Zhou & Law, 2015). siRNAs can be encoded by genic, intergenic, intragenic and intronic regions, but can also originate from TEs and heterochromatic regions (Chen, Zeng, et al., 2018). siRNAs are capable of post-transcriptional silencing like miRNAs via natural anti-sense siRNA, but can exert more versatile functions in genomic regulation. Repeat associated siRNA, referred to in some literature as heterochromatic siRNA, are typically encoded by transposons and are most frequently associated with the silencing of TEs via RNA-directed DNA methylation (RdDM; Wang & Axtell, 2017). Our knowledge on how siRNAs accumulate in subgenomes of amphidiploids and the presence of subgenome bias is currently limited to vegetative tissues, with minimal information on these dynamics in reproduction (Shen et al., 2015). It is also unknown if certain classes of siRNAs accumulate asymmetrically or if these siRNAs are ancestrally conserved within subgenomes of a polyploid species. Literature detailing the siRNAs transcribed during seed development in *B. napus* are lacking, especially in the context of its complicated amphidiploid genome.

The orchestration of seed development in amphidiploids requires the epigenetic balance of progenitor genomes. While DNA methylation and RNA interference are critical processes underpinning development, much of the details associating them to seed development are still emerging, especially in allopolyploids (Ario et al., 2017; Bouyer et al., 2017; Kirkbride et al., 2019; Narsai, Gouil, et al., 2017). To address these questions, we profiled both the DNA methylation and siRNA landscape across seed development of *B. napus*. Mapping of collinear regions established horizontal comparisons of homology between the progenitor An and Cn subgenomes (Figure 3.1b,c). Together, we find strong evidence for Cn subgenome bias in the context of both DNA methylation and siRNA accumulation, and find that the An subgenome

exerts epigenetic dominance over the submissive Cn subgenome in *B. napus* seed development. Further, differences in DNA methylation and siRNA expression in the context of the subgenomes and fractionated genomes provide evidence to how genome fractionation and allopolyploidy affect the epigenetic structuring of the seed across development.

3.3 Results

3.3.1 Single-base-pair resolution profiling of the *Brassica napus* seed methylome reveals Cn subgenome bias

Genome-wide cytosine methylation was calculated over 1-kB windows for both GLOB and MG seed stages in all cytosine contexts, and was compared with available cytosine methylation data for leaves (Table 3.1; Dataset S3.1). We used the plastid genome as a proxy for unmethylated DNA, and estimated the bisulfite conversion rate to be ~98% in both the GLOB and MG samples (Figure 3.2a). Relative to leaves, seed development was characterized by significantly lower mCG and mCHG levels (Table 3.1; $P < 0.001$ Mann–Whitney-Wilcoxon). However, mCHH levels were nearly threefold higher at the MG stage than observed in GLOB seeds or in leaves (Dataset S1; Table 1; $P < 0.001$ Mann–Whitney-Wilcoxon). This increase in mCHH contributed to a 2.5% increase in overall cytosine methylation of MG compared with GLOB seeds. We further identified shifts in DNA methylation between leaves, GLOB and MG seeds in addition to significant differences in the methylation of the An and Cn subgenomes (these genomes arising from the interspecific hybridization are referred to hereafter collectively as 'subgenomes' throughout) of *B. napus*. In leaves, and GLOB and MG seeds, mCG and mCHG levels were 1.4–1.5 × higher in the Cn than in the An subgenome (Table 3.1; Mann–Whitney-Wilcoxon, $P < 0.001$). Significant subgenome bias was also observed in CHH methylation, though mCHH levels were 1.1–1.2 × higher in the Cn subgenome than in the An subgenome. A

sharp increase in mCHH during seed maturation nearly abolishes mCHH subgenome bias in MG seeds (Table 3.1; Mann–Whitney-Wilcoxon, $P < 0.001$). Overall, while subgenome-wide methylation bias persisted in *B. napus* leaves and seeds, this phenomenon was strongest in leaves and weakened in the seed as it matures (Table 3.1).

Table 3.1. Methylation percentages on 1kB windows over the *B. napus* genome for leaves, globular stage seeds (GLOB), and mature green stage seeds (MG). Mean methylation levels were calculated over 1kB windows from individual cytosine methylation levels in each context (CG, CHG, CHH. Averages were also calculated separately for the A and C subgenome, and the ratio of these averages ($\Delta C^n/A^n$) is also shown. * indicates significance at $p<0.001$ according to a Mann Whitney Wilcoxon test with Bonferroni multiple testing correction.

	Leaf			GLOB			MG		
%	CG	CHG	CHH	CG	CHG	CHH	CG	CHG	CHH
Average	53.44	17.53	3.1	46.73	15.94	3.82	48.95	13.69	9.01
A^n	40.27	13.63	2.68	36.71	12.91	3.37	38.2	11.09	8.96
C^n	61.01	19.82	3.35	52.95	17.83	4.1	55.63	15.31	9.8
$\Delta C^n/A^n$	1.52*	1.45*	1.25*	1.44*	1.38*	1.22*	1.46*	1.38*	1.09*

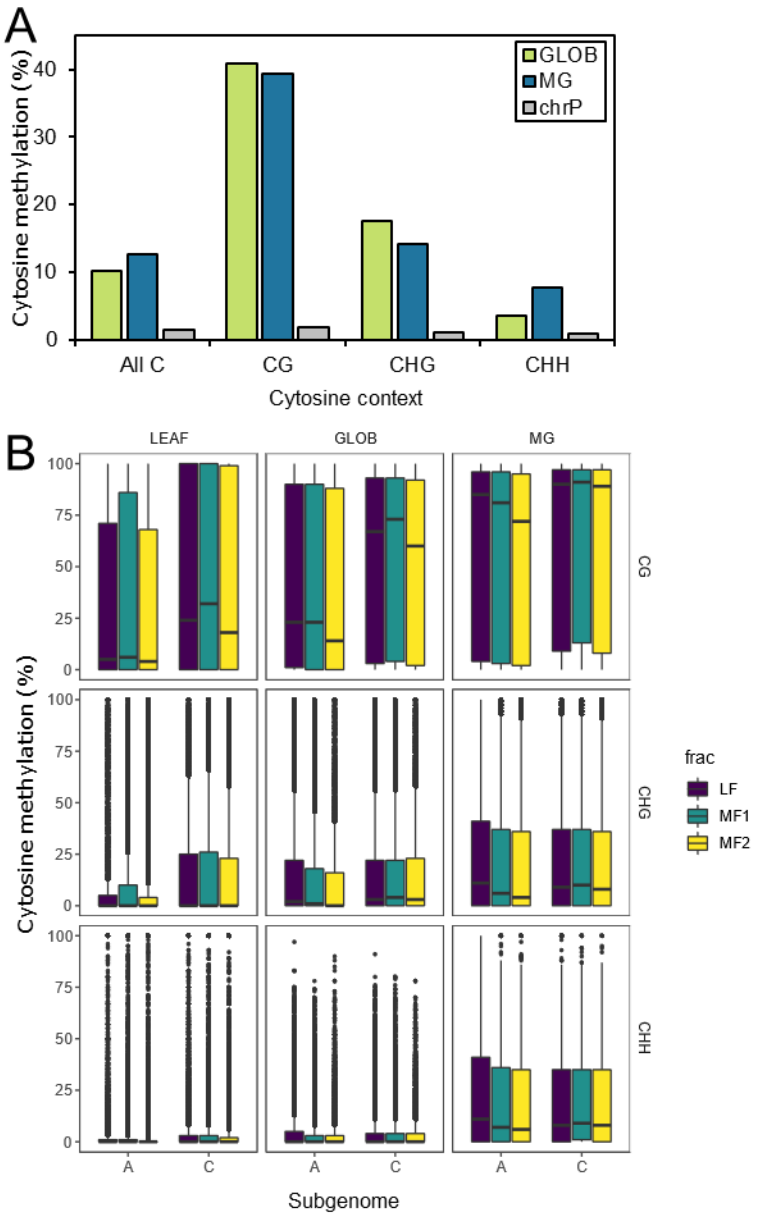


Figure 3.2. The DNA methylation in *Brassica napus* seed development. (A) Average methylation levels by cytosine context. Average methylation for all covered cytosines is shown for the GLOB (green), MG (blue), and plastidial (grey) genomes. (B) Level of methylation fractionated genomes (LF, MF1, and MF2) within the A^n and C^n subgenomes of *B. napus*. Quantile boxplots show level of cytosine methylation as calculated in 1kB windows in the syntenic regions for leaves, and GLOB and MG seeds in the CG, CHG, and CHH contexts.

We further compared the methylation patterns of the LF, MF1 and MF2 genomes (these whole genomes arising from the WGT are referred to hereafter collectively as 'fractionated genomes' throughout) in leaves and the GLOB and MG stages of seed development. We found that the majority of comparisons between the subgenomes, fractionated genomes and developmental stages were significantly different in CG (> 99% of total comparisons), CHG (98%) and CHH (98%) contexts (Figure 3.2b; Dataset S2). Methylation in the same cytosine contexts was significantly different ($P < 0.0001$) across subgenomes in the same fractionated genome (LF-A/C, MF1-A/C and MF2-A/C) and stage in seed development (GLOB, MG). This is true for every possible comparison within the same methylation context aside from one notable exception: methylation in the same cytosine contexts was insignificantly different between the LF and MF2 genomes in the An subgenome specifically. In particular, LF-A and MF2-A were insignificantly different ($P > 0.001$) in MG, GLOB/MG/LEAF and GLOB/MG/LEAF in CG, CHG and CHH contexts, respectively. The fractionated genomes are significantly distinct in their global methylation architecture throughout seed development and within and across subgenomes/fractionated genomes, with the sole exception being the LF and MF2 genomes of the An subgenome.

3.3.2 Differential methylation of promoter elements targets genes involved in carbon metabolism and development in *Brassica napus* seeds

Methylation on promoters is relatively stable during seed development. CG methylation does not significantly change between the GLOB and MG seed stages (leaf mCG = 71.8, GLOB mCG = 64.0, MG mCG = 63.8; Figure 3.3a; Table 2; $P < 0.001$ Mann–Whitney-Wilcoxon). When GLOB and MG seeds were compared with leaves, 84.6% of CG differentially methylated promoters and 77.8% of CHG differential methylation on promoters was shared between the

GLOB and MG seeds (Figures 3.3a and S3.2), suggesting little change in promoter CG methylation once seed development was initiated. Combined, these data suggest the developing seed is distinguished from vegetative leaves by an extensive reduction of CG and CHG methylation on promoters, which is maintained through the development and maturation of the seed. We compared promoter element methylation of seeds with leaves, and found that of the 1976 promoters differentially methylated between GLOB (933) and MG (1043) seeds and leaves (Figures 3.3b and 3.S2), 95% were derived from the Cn subgenome (Dataset S3.3).

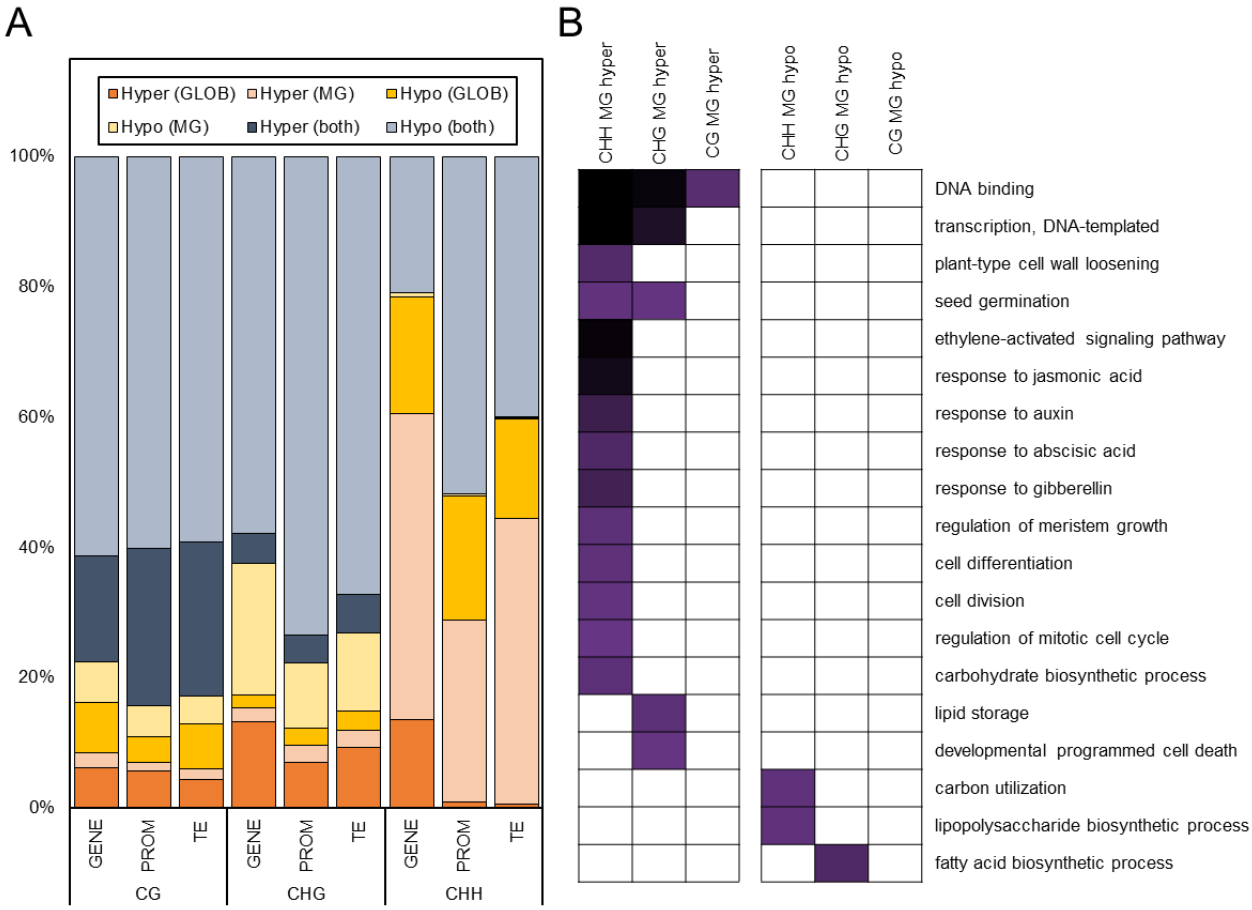


Figure 3.3. Shifts in DNA methylation between different genomic elements throughout seed development. (A) Stacked 100% column showing hyper- and hypo-methylated genes (GENE), 1kB upstream regulatory regions (promoters, PROM), and transposable elements (TE) between seeds (GLOB and MG) and leaves. Data are separated into hyper-methylated in GLOB (orange) or MG seeds (pale orange, hypo-methylated in GLOB (yellow) or MG (pale yellow) seeds, and

hyper-methylated (navy blue) or hypo-methylated (pale blue) in both GLOB and MG stages. (B) GO enrichment of genes with differentially methylated promoters between the GLOB and MG stages of seed development. Highly enriched GO terms are indicated by darker colouring. Gene lists and enrichment output can be found in Dataset 3.S4.

To identify the biological processes affected by developmental shifts in promoter methylation, we examined enrichment of Gene Ontology (GO) terms for genes downstream of promoters differentially methylated between GLOB and MG seeds (Figure 3b; Dataset S4; significance of GO enrichment considered at $P < 0.001$). At seed maturation, processes involved in regulating gene expression were heavily enriched in genes downstream of hypermethylated promoters, such as DNA binding (CG, CHG, CHH hypermethylated) and DNA-templated transcription (CHG and CHH hypermethylated). CHH hypermethylation of promoters in MG seeds was also enriched for response to hormones such as ethylene, jasmonic acid, gibberellin and auxin. We also observed CHH hypermethylation of promoters upstream of genes involved in cell division and differentiation, cell wall loosening, programmed cell death, and seed germination in GLOB seeds. Promoter hypomethylation of MG seeds was enriched for fatty acid biosynthesis (CHG hypomethylation) and carbon utilization (CHH hypomethylation), while promoters of genes involved in carbohydrate biosynthesis were hypermethylated (CHH). Therefore maturation-phase promoter hypermethylation appears to target hormone response, gene activation, and the balance between morphogenesis and quiescence, while promoter hypomethylation in seed maturation is likely associated with the release of genes for mobilizing carbon into lipid biosynthesis (Figure 3b).

3.3.3 Transcription factors (TFs) are hypomethylated relative to other protein-coding genes

We examined the methylation levels of TFs relative to other protein coding genes — data show a significant reduction in both gene body and promoter methylation TF-coding genes relative to other genes (Figure S3; Table 2; $P < 0.001$ Mann–Whitney-Wilcoxon). In leaves, GLOB and MG seeds, TF gene body methylation levels were reduced by 57–78% relative to the levels of other coding gene bodies (Figure S3; Table 2; $P < 0.001$ Mann–Whitney-Wilcoxon). TF promoter CG and CHG methylation is significantly reduced relative to other coding genes, though the reduction is less than observed on gene bodies. While mCG on TFs was significantly lower in seeds than observed in leaves, mCG levels did not significantly change between GLOB and MG seeds nor did they change on the promoters of TFs (Table 3.2; $P < 0.001$ Mann–Whitney-Wilcoxon). Additionally, despite overall lower cytosine methylation levels, subgenome methylation bias was maintained on TF gene bodies and promoter regions (Table 3.2).

Table 2. Percent methylation on repetitive elements (TE), genes bodies, and promoter elements (1kB window upstream of TSS), transcription factor gene bodies and promoters, and siRNA clusters. Mean methylation levels were calculated over features from individual cytosine methylation levels in each context (CG, CHG, CHH). Averages were also calculated separately for the A and C subgenome, and the ratio of these averages ($\Delta C^n/A^n$) is also shown. * indicates significance at $p<0.001$ according to a Mann Whitney Wilcoxon test with Bonferroni multiple testing correction.

	Leaf			GLOB			MG			Feature
%	CG	CHG	CHH	CG	CHG	CHH	CG	CHG	CHH	
Average	91.42	37.85	9.57	80.36	30.81	7.15	84.61	32.35	23.04	TE
A^n	86.56	38.78	10.78	75.89	29.75	7.28	80.06	32.45	24.13	
C^n	93.53	37.46	9.03	83.03	31.45	7.07	87.34	32.29	22.32	
$\Delta C^n/A^n$	1.08	0.97	0.84*	1.09*	1.06*	0.97*	1.09*	0.99	0.92	
Average	71.84	33.96	7.45	63.97	11.43	0.47	63.83	8.09	6.38	Promoter
A^n	62.5	30.25	7.41	54.91	10.17	0.47	54.75	7.45	6.07	

C^n	76.43	35.78	7.47	68.72	12.09	0.46	68.58	8.43	6.54	
$\Delta C^n/A^n$	1.22*	1.18*	1.01*	1.25*	1.19*	0.99*	1.25*	1.13*	1.08*	
Average	22.9	5.82	0.91	21.8	2.16	0.13	21.6	1.18	0.99	Gene
A^n	15.94	3.96	0.66	15.1	1.5	0.11	14.96	0.8	0.73	
C^n	27.02	6.92	1.05	25.72	2.55	0.14	25.58	1.41	1.15	
$\Delta C^n/A^n$	1.69*	1.74*	1.59*	1.7*	1.7*	1.22*	1.71*	1.76*	1.57*	
Average	8.86	1.62	0.4	7.91	0.51	0.05	7.8	0.33	0.45	TF Gene
A^n	7.54	1.31	0.32	6.71	0.45	0.04	6.57	0.26	0.37	
C^n	10.06	1.91	0.47	8.99	0.56	0.05	8.91	0.39	0.52	
$\Delta C^n/A^n$	1.33*	1.46*	1.46*	1.34*	1.25*	1.18*	1.36*	1.48*	1.39*	
Average	61.06	32.08	8.15	51.07	9.66	0.42	50.93	7.2	6.33	TF Promoter
A^n	52.07	26.41	7.19	43	7.8	0.35	42.83	6.02	5.6	
C^n	68.63	36.82	8.98	58.11	11.29	0.49	58	8.23	6.97	
$\Delta C^n/A^n$	1.32*	1.39*	1.25*	1.35*	1.45*	1.4*	1.35*	1.37*	1.24*	
Average	NA			54.40	33.03	12.57	84.61	52.22	44.08	siRNA
A^n				54.11	32.95	12.53	84.55	52.12	44.07	
C^n				58.70	34.15	13.15	85.43	53.58	44.34	
$\Delta C^n/A^n$				1.08	1.04	1.05	1.01	1.03	1.01	

3.3.4 Fractionated genomes exhibit different methylation on genes and promoters

We further compared components of the *B. napus* fractionated genomes across the progenitor An/Cn subgenomes (Figure 4.4). Comparisons between the numerous variables are complicated, and comprise the fractionated genomes (LF, MF1 and MF2), the subgenome (An and Cn), the developmental stage (GLOB and MG) and methylation context (CG, CHG and CHH). Hereafter, statistical comparisons refer to only comparisons made within the same methylation context. We compared the methylation on gene bodies within the fractionated genomes, and in all instances that no significant difference ($P > 0.001$ Mann–Whitney-Wilcoxon) was found, the comparison was within the same subgenome (Figure 4.4a; Dataset S4.3). In fact, this trend is true for all cytosine contexts, where the only insignificant differences were between

fractionated genomes in the same subgenome. Our data suggest that methylation on gene bodies is more similar within different fractionated genomes belonging to the same subgenome than it is in the same fractionated genome in the sister subgenome. Further, this pattern is the same in promoter regions, wherein the only insignificant comparisons were between different fractionated genomes in the same subgenome (Figure 4.4b). Together, it seems methylation of coding regions and their primary regulatory regions (promoters) are dissimilar between the sister subgenomes and more alike within the same subgenome.

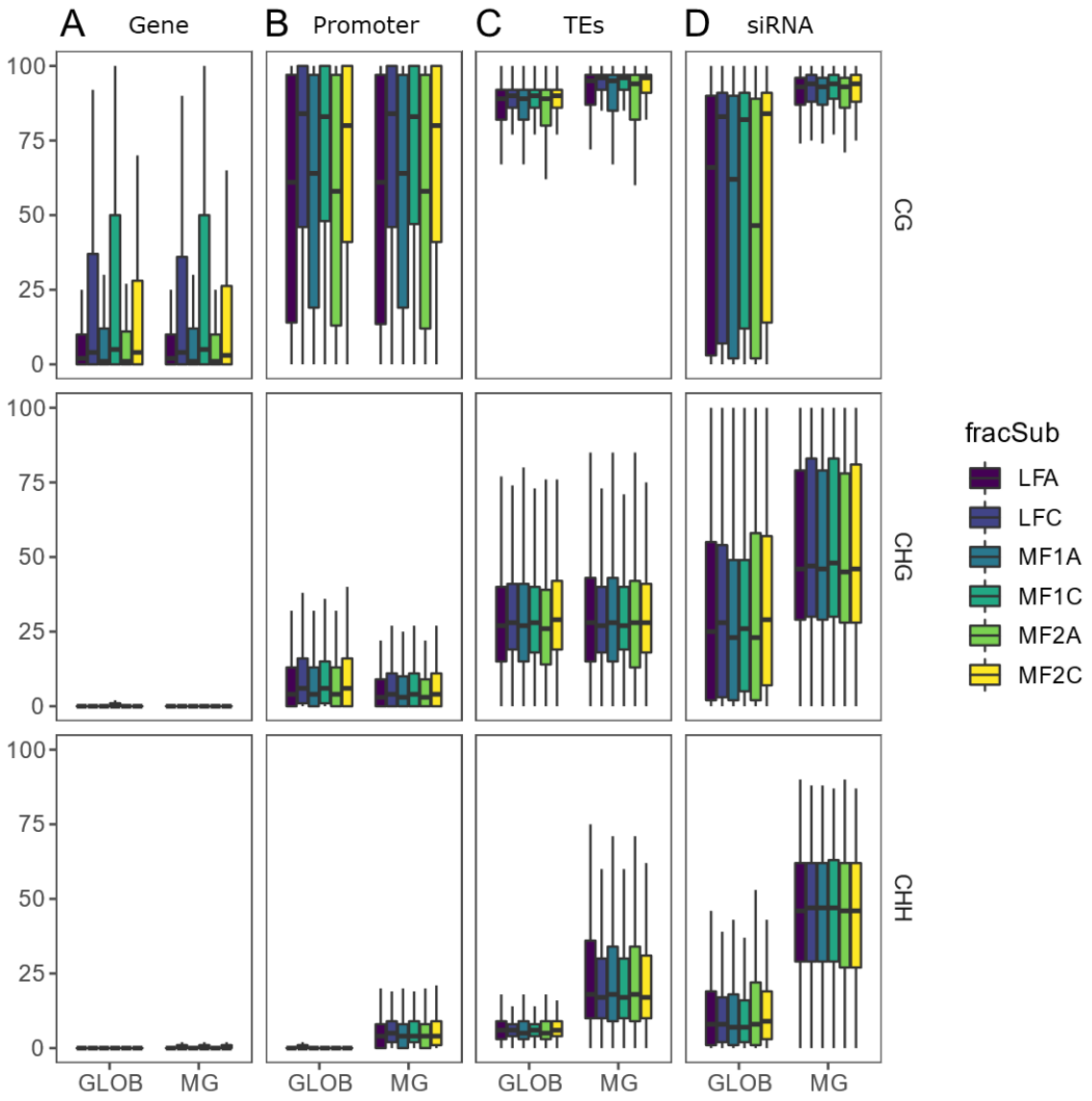


Figure 3.4. The DNA methylation of genome features (genes, promoter regions (prom), transposable elements (TEs), and siRNA loci. Box and whisker plots divided by the fractionated genomes of *B. napus* (Least Fractionated (LF), Most Fractionated 1 (MF1), and Most Fractionated 2 (MF2) and their corresponding subgenome (A, C).

3.3.5 Subgenome bias of DNA methylation on TEs during reproduction is unique from vegetative tissues

In all three cytosine contexts, the vast majority of differential methylation between leaf and seed methylomes occurred on TEs (Figure S4.4; Dataset S4.3). TEs account for

approximately 33.7% of the genome mapped to chromosomes in *B. napus* cv. DH12075, though this is an underestimation of total TE content as scaffolds that could not be confidently mapped to chromosomes were not included in this analysis. Methylation of TEs in the CG and CHG contexts was significantly reduced in seed development relative to leaves (leaf mCG = 91.5%; Table 2). CG methylation on TEs was observed to be lower during seed morphogenesis and increased during seed maturation (GLOB mCG = 80.4%, MG mCG = 84.6%; Table 2; significant $P < 0.001$ Mann–Whitney-Wilcoxon). Overall, a degree of subgenome bias was maintained during these shifts in CG methylation: mCG levels are 1.1 × higher on Cn subgenome TEs (Table 3.2; $P < 0.001$ Mann–Whitney-Wilcoxon).

The CHG methylation on TEs does not change significantly between the GLOB and MG stages of seed development (GLOB mCHG = 30.8%, MG mCHG = 32.4%; Table 2; $P < 0.001$ Mann–Whitney-Wilcoxon; Figure 3.2a). However, there is a unique population of TEs that undergo heavy CHH methylation in MG seeds but not in GLOB seeds or in leaves (Figure 3.S2). The lower CG and CHG methylation observed in seeds relative to leaves is accompanied by an increase in CHH methylation on TEs during seed maturation (leaf mCHH = 9.6%, GLOB mCHH = 7.2%, MG mCHH = 23.0%; Table 2; all significant $P < 0.001$ Mann–Whitney-Wilcoxon) and is observed globally within the genome (Figure 3.4c). Over 40% of mCHH-DM TEs were hypermethylated in the MG stage relative to leaves (Figure S3.2). CHH methylation of An subgenome TEs was 1.2 × higher than Cn subgenome TEs in leaves, though this bias is not observed in the seed (Table 3.2; $P < 0.001$, Mann–Whitney-Wilcoxon). This suggests extensive reprogramming of genome architecture around TEs during seed development, driven by both a reduction in CG and CHG methylation, and subgenome-independent increase of CHH methylation on TEs during seed maturation.

We also compared the mCG, mCHG and mCHH in the TEs of the fractionated genomes. The majority of comparisons between the fractionated genomes in the CG context were significant (82%), though fewer were significant in the CHH (61%) and CHG (24%) context (Figure 3.4c; Dataset 3.S5). The insignificant ($P > 0.001$) comparisons in CG contexts were entirely within the same subgenome (An or Cn) and in the same developmental stage. Both the LF and MF1 genomes were insignificantly different between subgenomes across seed development in the CHG context but not in CG or CHH. The LF genome was significantly different ($P < 0.001$) between subgenomes during the GLOB stage in the CHG context, but not in the MG. In contrast, the MG seed of the MF2 genome had significantly different methylation across subgenomes in both CG and CHG contexts, but was insignificantly different in GLOB seeds.

3.3.6 Methylation of siRNA loci during seed maturation

Cytosine methylation increased one–3.5-fold on siRNA clusters loci from the GLOB to the MG stage of seed development (GLOB mCG = 84.0%, MG mCG = 86.3%; GLOB mCHG = 45.9%, MG mCHG = 55.0%; GLOB mCHH = 12.7%7, MG mCHH = 43.8%; Table 2; $P < 0.001$ Mann–Whitney-Wilcoxon). Most notably, the mCHH levels of siRNA loci at the MG stage are 4.9 × higher than the whole genome average (Table 3.2), and nearly double the levels observed on TEs. We also compared methylation patterns on siRNA loci in the fractionated genomes and found a similar pattern as that observed in TEs where CG methylation of siRNAs is not significantly different across subgenomes but is significantly different within the same subgenome (Figure 3.4d). LF-A/C, MF1-A/C and MF2-A/C are significantly different in GLOB and MG in the CG context, but CHG methylation of LF-A/C and MF2-A/C are not significantly different in MG seeds. CHG methylation of LF-A/C is also similar ($P = 1$; Mann–Whitney-

Wilcoxon) in GLOB seeds. Methylation of expressed siRNA loci in the CHH context exhibit different patterns than seen in genes and promoters, but follow a similar trend observed in TEs, wherein LF-A/C, MF1-A/C and MF2-A/C are insignificantly different (P = 1; Mann–Whitney-Wilcoxon) in both GLOB and MG seeds. Methylation of siRNA loci is divergent amongst fractionated genomes during seed development in CG and CHG contexts, and is remarkably similar across subgenomes in the CHH context.

3.3.7 Seed maturation is characterized by an increase in the number of siRNA loci being transcribed

Next, we profiled sRNA populations, genome wide across *B. napus* seed development. We found that the maturation stages had the greatest number of expressed siRNA loci [reads per million (RPM) > 0]; 36 989 and 30 747 siRNA loci were expressed during the MG and DS stages, respectively (Dataset S6). The OV stage had the fewest siRNA loci with detectable expression at 25 197 in contrast to the GLOB and HEART stages of morphogenesis, which had 29 127 and 28 636 loci being expressed, respectively. Subgenome bias of siRNA expression is persistent throughout seed development – in all stages, ~65% of all detectable siRNA originated from the Cn subgenome, with this bias persisting in total RPM with a minimum of 61% of siRNA reads accumulating from the Cn subgenome in the OV stage, and a maximum of 74% in the GLOB stage (Figure 3.5a). Fewer siRNA loci account for the majority of all RPM in early seed development (Figure 3.4b). In the OV, GLOB and HEART stages, > 99% of all reads map to 21%, 20% and 11% of all siRNA loci in each respective stage. In contrast, 61% and 32% of all reads in the MG and DS stages, respectively, accounted for > 99% of all reads in each stage. These trends were not seen to differ between subgenomes. Fewer siRNA loci contribute to seed development in *B. napus* during morphogenesis than in maturation—siRNAs in morphogenesis

are highly expressed, but fewer loci have detectable expression than in early maturation. A complete list of all siRNA loci across TEs and gene promoter regions encode the majority of transcribed sRNAs in the *Brassica napus* seed.

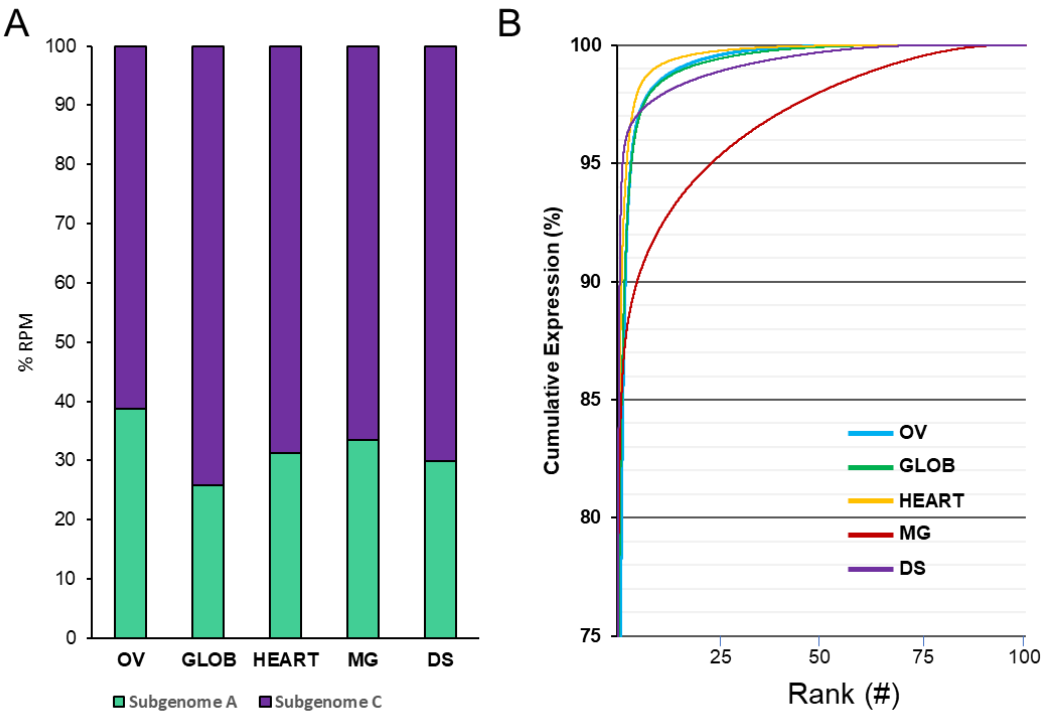


Figure 3.5. siRNA read accumulation throughout seed development in *B. napus*. (A) Proportion of total siRNA RPM in each seed stage library shows consistent C^n subgenome bias. **(B)** Fewer siRNA loci account for a greater proportion of total reads in early seed development, while the proportion of reads is more evenly distributed in the MG stage than in the OV, GLOB, and HEART stages.

We identified sRNAs originating from TEs, gene bodies and promoter regions across seed development, and compared RPM accumulation in each feature. The sRNAs originating from TEs were the most represented in read accumulation and exhibited substantial bias for the Cn subgenome (67%, 79%, 81%, 77% and 91% of all reads in the OV, GLOB, HEART, MG and DS stages, respectively; Figure 3.6a). siRNAs encoded by gene bodies were the least abundant

throughout seed development, but still biased towards the Cn subgenome (70%, 76%, 61%, 65% and 53% of all reads in the OV, GLOB, HEART, MG and DS stages, respectively; Figure 3.6b). Finally, siRNAs originating from promoter regions were more abundant than those derived from gene bodies, and exhibited genomic bias for the Cn subgenome (71%, 83%, 74%, 87% and 65% of all reads in the OV, GLOB, HEART, MG and DS stages, respectively; Figure 3.6c). While siRNAs encoded by TEs were of comparable abundance across seed development, siRNAs originating from promoters gradually increased as the seed matured. siRNAs originating from promoters of the An subgenome were 1.1 × as abundant in the MG stage as at the OV stage, while siRNAs from the Cn subgenome were 3.3 × as abundant. siRNA accumulation from promoters rapidly declines in the Cn subgenome at the DS, with 2.5 × less siRNAs than in the MG seeds. Conversely, siRNAs encoded by the An subgenome gradually increased from the beginning to end of seed development, with the DS having 1.4 × the siRNA accumulation relative to the MG seeds.seed development is found in Dataset S6.

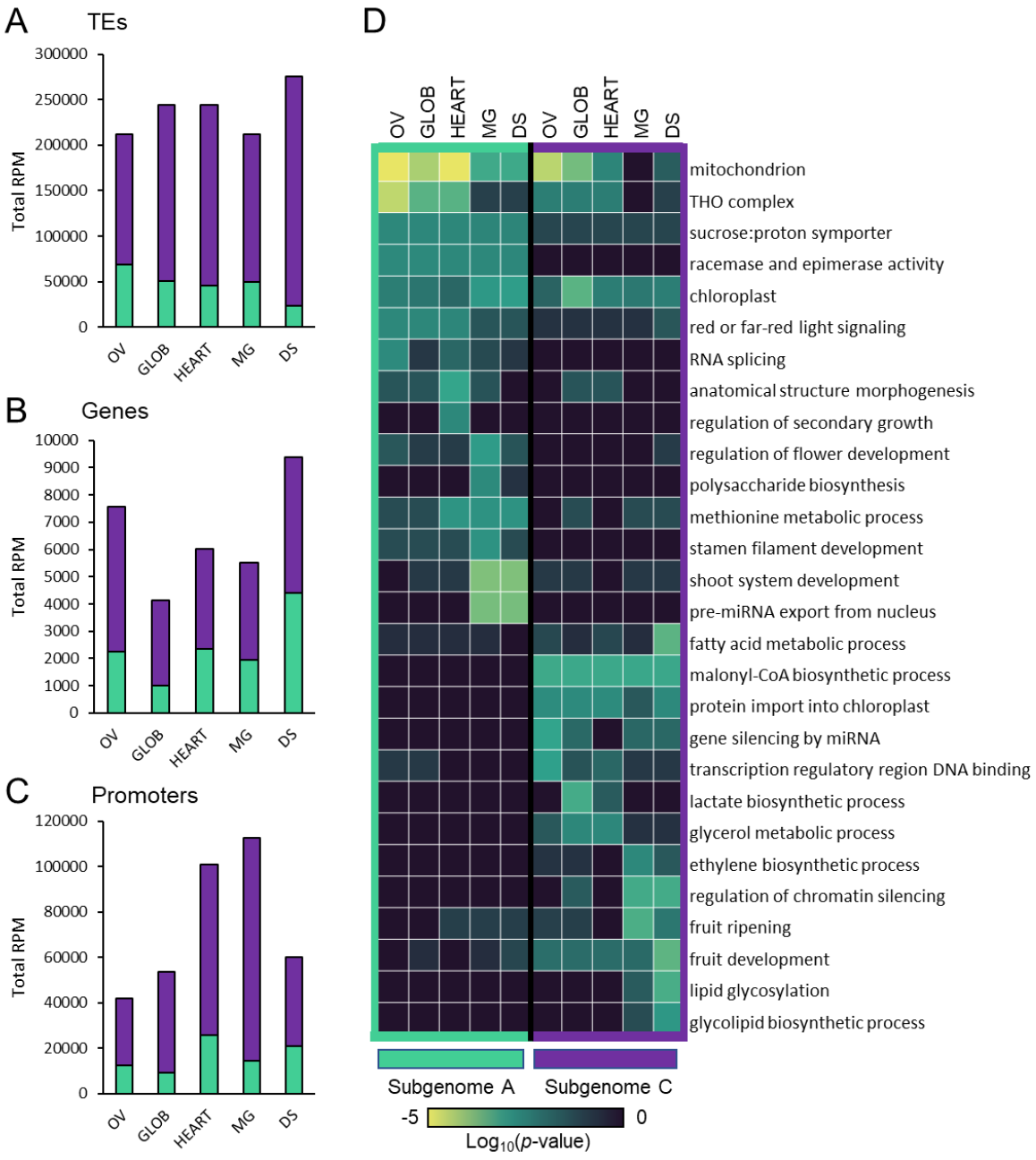


Figure 3.6. Total RPM (Reads per Million) of siRNA clusters encoded by transposable elements (TEs), gene bodies and promoters. (A) siRNAs encoded by TEs accumulate with consistent C^n subgenome bias throughout seed development. (B) siRNAs originating from gene bodies accumulate most strongly in the DS stage and are less numerous than those encoded by promoter regions. (C) siRNAs originating from promoter regions are most abundant during the MG stage, with increasing expression from the OV to MG stages. Promoter regions exhibit stronger subgenome bias than siRNAs encoded by gene bodies. (D) GO analysis of coding genes associated with the promoter region encoding the top 500 most highly expressed siRNA clusters in each seed stage.

3.3.8 Biological processes associated with gene promoters are different across subgenomes

Next, we analyzed the top 500 most highly expressed siRNAs from gene promoter regions in each subgenome and performed GO enrichment analysis on their downstream genes (Figure 3.6d). GO enrichment analysis identified significant subgenomic bias ($P < 0.01$) for many developmental, genomic and biochemical processes throughout seed development for the most abundant siRNAs encoded by gene promoter regions. For example, sucrose transport and polysaccharide synthesis are represented more strongly in the An subgenome ($P < 0.01$) both throughout seed development and in the MG stage, respectively. Additionally, mitochondrial and chloroplast processes are represented across both subgenomes, though some amino acid metabolism processes (racemerase/epimerase activity, methionine metabolism) are asymmetrically biased towards the An subgenome ($P < 0.01$ in An subgenome, $P > 0.1$ in Cn subgenome). Developmental processes associated with anatomical structure morphogenesis ($P < 0.01$ An HEART), secondary growth ($P < 0.01$ An HEART), flower development ($P < 0.01$ An MG), stamen filament ($P < 0.01$ An MG) and shoot system development ($P < 0.001$ An MG/DS) are also biased towards the An subgenome. Interestingly, we find developmental processes associated with fruit development are biased towards the Cn subgenome ($P < 0.05$ Cn OV/GLOB/HEART/MG, $P < 0.001$ Cn DS), implicating siRNAs from the An subgenome intervene in regulation of genes relevant to vegetative and early reproductive development. Instead, siRNAs encoded by the Cn subgenome impact the expression of genes known to be involved in post-fertilization processes. Further, we find that diverse lipid metabolism (glycosylation, biosynthesis) processes are biased significantly towards the Cn subgenome

(P < 0.01 Cn DS). Fatty acid metabolism (P < 0.001 Cn DS) and glycolipid processes (P < 0.001 Cn DS) are represented in the final stages of seed development, only becoming significantly enriched in the maturation phase. Taken together, the most abundant siRNAs encoded by each subgenome do not mirror each other's activity on various biological processes, and that particular biological processes of interest to *B. napus* development including lipid metabolism are likely asymmetrically regulated by the Cn subgenome.

3.3.9 Ancestral genome fractionation has little effect on siRNA accumulation during seed development

The WGT shared amongst the Brassiceae resulted in the LF, MF1 and MF2 genomes in Brassica. Our data suggest that siRNAs transcribed during seed development from each of the three fractionated genomes do not vary considerably from each other (Figure 3.7). Broadly, all three fractionated genomes have fewer siRNA loci expressed during pre-fertilization and morphogenesis than observed during seed maturation. Further, morphogenesis in all three fractionated genomes is characterized by fewer loci accounting for the majority of the total RPM. This is consistent with the whole subgenome comparisons in Figure 3.5. In contrast, there is a global shift in siRNA expression upon the transition to the MG stage, wherein all three fractionated genomes express considerably more siRNA loci but are proportionately lowly expressed. In the LF genome, there are 3.4 × and 4.4 × more siRNA loci expressed in the MG stage than in the HEART stage in the An and Cn subgenomes, respectively. This trend remains consistent in the other two fractionated genomes, wherein 3.5 ×, 4.4 ×, 3.2 × and 3.8 × more siRNA loci are transcribed in the MG stage relative to the HEART stage of the MF1-A, MF1-C, MF2-A and MF2-C genomes, respectively. This dramatic change in siRNA expression seems to disappear once the seed enters dormancy at the DS stage in all three ancestral genomes, with

the distribution appearing to be a holdover from the notable shift that characterizes the inception of maturation.

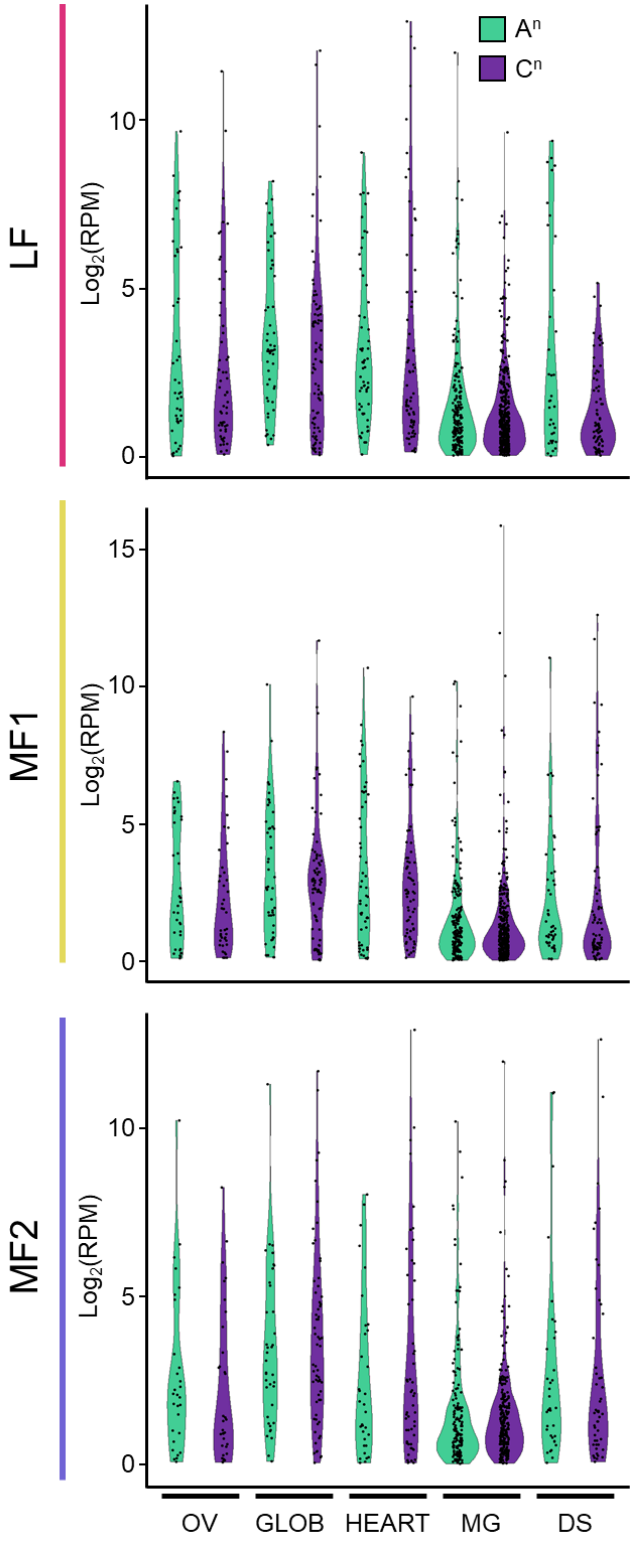


Figure 3.7. Violin plots of siRNA loci with >1 RPM in LF, MF1, and MF2 genomes of OV, GLOB, HEART, MG, and DS stages.

We used a Kolmogorov–Smirnov (K–S) test to determine differences in data distributions between both the stages of seed development and across the ancestral genomes (Tables S3.1–S3.3). Data show the distribution of siRNA expression between LF/MF1, MF1/MF2 and LF/MF2 was not significantly different when compared at the same seed stage. The only significant differences ($P < 0.05$) were identified at the MG stage, which is characterized by a global increase in siRNA loci being expressed, but at much lower levels than seen in the OV and during seed morphogenesis. This phenomenon appears true for all comparisons across both the subgenomes and fractionated genomes. For example, there are no statistical differences between the seed stages and fractionated genomes unless it is between the MG seed and any other stage. While not statistically significant, the siRNA expression patterns were more similar across fractionated genomes than they were across subgenomes. For example, the OV, GLOB, HEART, MG and DS stages within the LF genome were more dissimilar between the An and Cn subgenomes (P-values 0.3641, 0.2084, 0.0508, 0.8434 and 0.0333, respectively) than they were in their matching corresponding subgenomes in the MF1 genome (An P-values 0.3743, 0.7089, 0.7553, 0.9944 and 0.4323, respectively; Cn P-values 0.9829, 0.4134, 0.2445, 0.9998 and 0.2361, respectively).

3.3.10 Genome fractionation has less influence on non-protein coding elements than on gene bodies and promoter regions

The WGT that defined the divergence of the Brassiceae resulted in three fractioned genomes that evolved independently within *B. rapa* and *B. oleracea* until the hybridization and inception of *B. napus*. As these fractionated genomes existed independent of one another in their

progenitors, we compared the significant and insignificant differences across these subgenomes. We found that DNA methylation of every comparison spanning the two subgenomes to be significantly different in the CG context for genes, promoters, TEs and expressed siRNA loci in both GLOB and MG seeds (Figure 3.8). This trend persisted in all cytosine contexts and in every seed stage for gene bodies specifically, wherein these regions of the fractionated genomes were all significantly different ($P < 0.001$). However, the most non-significant comparisons were identified in the methylation of TEs and expressed siRNA loci, which are consistently insignificantly different in both CHG and CHH contexts within the GLOB (Figure 3.8a) and MG (Figure 3.8b) seeds, as well as across GLOB/MG (Figure 3.8c). This trend was most obtuse in the MG stage, which is expected given that the MG stage is characterized by a sweeping increase in CHG and CHH methylation. Despite the substantial shift in DNA methylation from the morphogenesis to maturation phases of seed development, CHG methylation across many GLOB/MG comparisons was insignificantly different in the context of promoters and TEs (Figure 3.8c). Methylation of siRNA loci was more similar across subgenomes when contained within the same developmental stage—that is, methylation on expressed siRNA loci was more dissimilar in the fractionated genomes across seed development than they were during the same developmental stage. Altogether, gene bodies had substantially different methylation profiles within the fractionated genomes, but non-protein coding regions including promoters, TEs and siRNAs loci had more similar architecture by comparison.

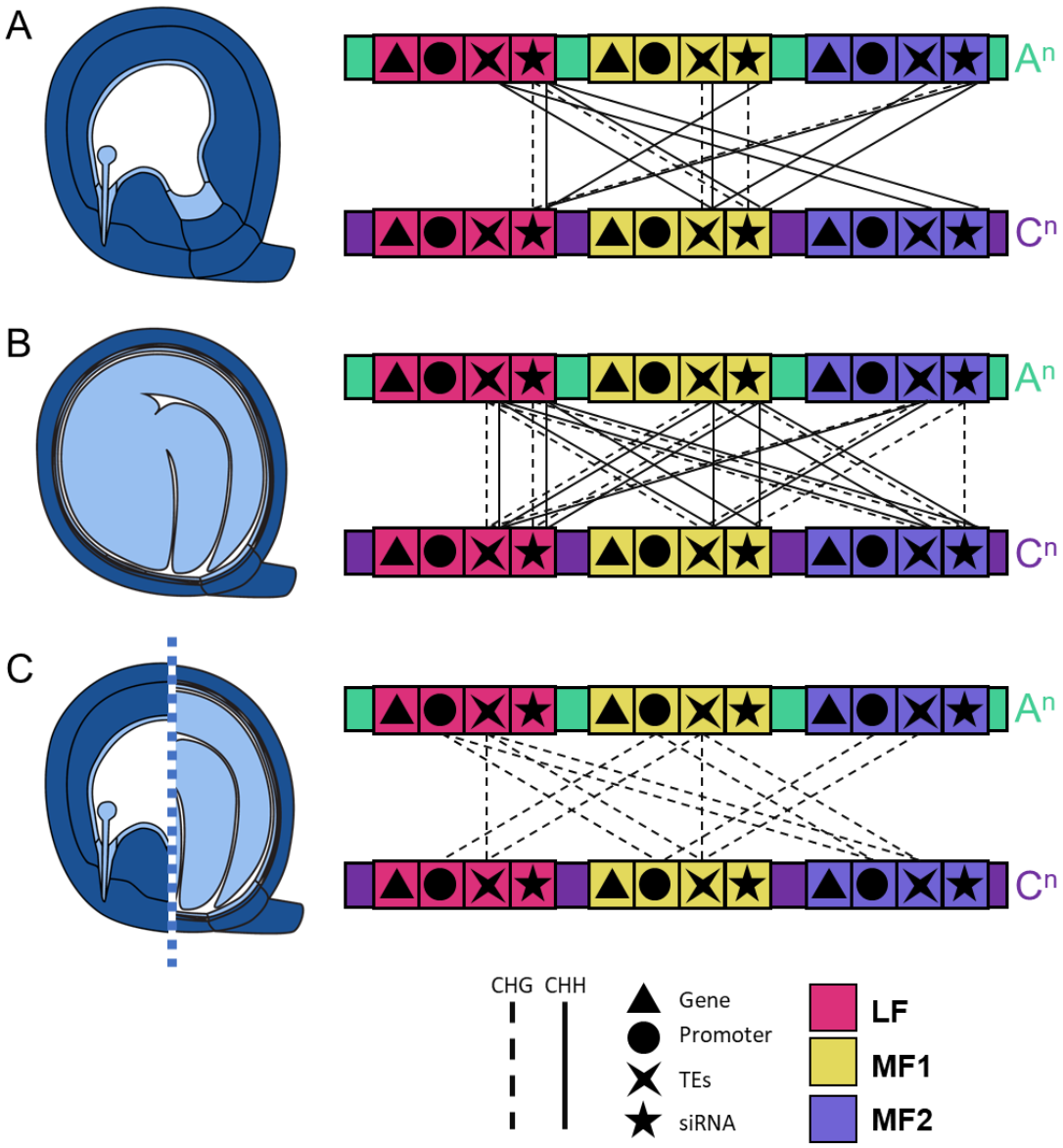


Figure 3.8. Highly insignificant comparisons ($p > 0.1$; Mann-Whitney-Wilcoxon) of Gene, Promoter, TEs, and siRNA loci in CHG and CHH methylation contexts. Comparisons of (A) GLOB, (B) MG, and (C) GLOB-MG are shown, with full dataset presented in Dataset S5, S6. Only comparisons that are A^n and C^n subgenomes are shown.

3.4 Discussion

3.4.1 Epigenetic bias in the Cn subgenome of *Brassica napus*

In allotetraploids like *B. napus*, interspecific hybridization can cause immediate silencing of genomic components and lead to morphological dominance of one progenitor species over the other, suggesting subgenome bias has profound impacts on plant development (Alexander-Webber et al., 2016; Comai et al., 2000; Wang et al., 2006). In the current study, we discovered epigenomic bias across seed development in *B. napus*. Specifically, consistent bias of cytosine methylation and siRNA accumulation in the Cn subgenome implies the selective silencing of its components.

Patterns of cytosine methylation and diversity of siRNA expression derived from TEs undergo substantial modifications upon allopolyploidization events (Kenan-Eichler et al., 2011). Amongst these, siRNAs are known to become deregulated upon allopolyploidization and cause an influx in TE activity in addition to generating instability within the genome. In this study, we find cytosine methylation and siRNA accumulation are universally biased towards the Cn subgenome in the *B. napus* seed, and are indicative of repression of its relative contribution to seed development and by extension are submissive to the An subgenome. These siRNA loci were also subject to profound CHH hypermethylation. This phenomenon of Cn subgenome bias persists throughout seed development, with siRNAs encoded by genes, promoter regions and TEs all accumulated disproportionately from the Cn subgenome. Consistent with previous work, more genes are transcribed from the Cn subgenome but their mean expression values are lower, suggesting the Cn subgenome is likely being silenced throughout development (Gao et al., 2022). The accumulation of non-coding elements such as siRNAs originating from TEs in the Cn subgenome may contribute to a greater overall increase in mCHH of the same Cn

subgenome as the seed develops. Together, the Cn subgenome appears to accumulate silencing elements that persist across all stages of the seed lifecycle. The asymmetric bias of the Cn epigenome indicates that the An subgenome of *B. napus* exhibits incipient genomic dominance during reproduction. Acute subgenome bias defines the epigenetic dynamics at play during seed development in this amphidiploid species, but it is unclear to what extent the components of the Cn subgenome will become fractionated in the future.

3.4.2 TFs are comparably hypomethylated in seeds

The TF genes have been documented to exist within DNA methylation valleys (DMVs), which are regions that have constitutively low or even undetectable DNA methylation levels. These account for 46% of TF genes in Glycine, which is indicative of either erasure or prevention of DNA methylation during seed development of genes pertaining to transcriptional machinery (Chen, Lin, et al., 2018). *Brassica napus* seeds seem to elicit the same phenomenon, with substantially reduced DNA methylation on TF gene bodies. It is possible that these TFs exist in DMVs in Brassica as well, but it would be conducive in future research to analyze the methylation levels of all TF genes throughout all stages of seed development. Interestingly, while TF gene bodies in *B. napus* are comparably demethylated in all cytosine contexts, their promoter regions are substantially more methylated than other gene bodies. It is thus possible that the bulk of transcriptional regulation of TF genes comes not from mCG on gene bodies, but instead increased methylation of promoters.

3.4.3 The contribution of TEs to subgenome bias in *Brassica napus* seed development

Seed maturation is characterized by the rapid deposition of protein and lipid energy stores for the developing seed, and coincides with an influx of both siRNA clusters and DNA

methylation, especially in the CHH context. While siRNAs encoded by TEs were most diverse in the MG stage, the accumulation of these siRNAs was prevalent throughout seed development. Thus, it is possible the increasing number of expressed siRNA loci originating from TEs promotes global genomic silencing as the seed enters maturation. This transition is also accompanied by an influx of siRNA reads from the Cn subgenome likely due to the Cn subgenome having more repetitive regions. siRNAs derived from TEs increase in genomic coverage as seed development progresses where CHH methylation exhibits a similar pattern once maturation is initiated.

Repetitive elements (REs) can be targets of epigenetic modification especially in allopolyploid plants, like *B. napus* (Parisod et al., 2009; Parisod et al., 2010). For example, a higher density and abundance of TEs in one progenitor genome may lead to downstream epigenetic modification in a hybrid polyploid (Wendel et al., 2018). Here, we observe that subgenome bias was most dramatic on TEs in *B. napus* seeds and leaves. TEs are more highly methylated in the Cn subgenome of *B. napus*, and while this phenomenon has been observed in vegetative tissues (Chalhoub et al., 2014) and microspores (Li et al., 2016), our data suggest epigenetic modifications go beyond what is observed in vegetative leaves via CHH hypermethylation on TEs. Moreover, DNA methylation may provide an additional layer of genetic regulation that serves to modulate variable or novel regions of the genome. Thus, trends in subgenome bias favoring higher methylation of the Cn subgenome may be a consequence of its higher TE load.

In *B. napus*, our data provide evidence of a global loss of methylation on TEs, especially from the An subgenome in the CG and CHG contexts relative to leaves. Thus, a higher level of methylation is still maintained on the TEs in the Cn subgenome. Currently, we cannot detect

whether this loss of methylation on TEs is via a passive or active process. For example, it is possible that loss of mCG and mCHG in seeds is a passive process that occurs over the numerous cell divisions and DNA replication cycles that establish the seed. Whether a loss of methylation relative to leaves is actively or passively achieved, it appears that Cn subgenome bias is a feature that persists over the transition of the plant to the next sporophytic generation.

3.4.4 RdDM as a modulator of developmental transitions in seed development

Active demethylation of TEs in ephemeral reproductive structures such as vegetative nuclei in pollen and the endosperm may lead to the activation of siRNA machinery to modify DNA methylation in other cells (Gehring et al., 2009; Slotkin et al., 2009). It is known that DNA hypomethylation in Arabidopsis gametophytes causes alterations in seed size, and that interference in DNA methylation machinery causes loss of seed viability (Xiao, Brown, et al., 2006; Xiao, Custard, et al., 2006; Zhang et al., 2010). Interestingly, our data indicate more siRNA loci are expressed as the seed approaches maturation, with the OVs expressing the majority of siRNAs from fewer loci, consistent with previous findings in the Brassicaceae (Grover et al., 2020). Given the predominant role siRNAs play in managing DNA methylation via the canonical RdDM pathway, the progressive increase in transcribed siRNA loci through seed development may be indicative of increasing epigenetic silencing as the seed grows. The most highly expressed siRNAs originating from gene promoter regions exhibit subgenomic biases with developmental and biochemical processes being asymmetrically regulated in each subgenome. siRNAs from the An subgenome accumulate from promoters of genes associated with cellular morphogenesis and shoot development. In late seed development, promoters of genes involved in cell division and auxin/gibberellin/cytokinin are targeted for CHH hypermethylation. These targets reflect the developmental phenotypes seen in Arabidopsis

lacking DNA methylation machinery, wherein improper cellular divisions are frequently observed (Kim et al., 2008; Xiao, Brown, et al., 2006; Xiao, Custard, et al., 2006).

The RdDM in maternal tissues is essential for early *B. rapa* seed development (Grover et al., 2018). As the maternal seed coat makes up a substantial portion of the *B. napus* seed at this stage, it is possible that some patterns of early RdDM activation are maintained from its progenitor *B. rapa*. The role of RdDM in seed maturation is still unclear, though it appears to be characteristic of seed maturation but viably non-essential in the model system Arabidopsis (Bouyer et al., 2017; Lin et al., 2017). In the current study, we found that siRNA loci were heavily methylated in mature *B. napus* seeds, where the embryo comprises the vast majority of the seed body. A spatiotemporal analysis of DNA methylation and sRNA accumulation in the embryo, endosperm and seed coat would provide insight into the RdDM in the maternal and zygotic tissues throughout seed development, and whether the embryo, endosperm and seed coat are all subject to the same regimes of subgenome bias observed in whole seeds.

3.4.5 siRNA accumulation correlates with global increases in CHH methylation

Many angiosperm genomes are typically comprised of transposable, repetitive sequences that accumulate through evolutionary time. It is accepted that these transposable sequences can cause withstanding genomic changes via mobilization and insertion into either coding genes or regulatory regions of the genome, and as such they are frequently silenced and regulated by the RdDM pathway. The various modes of TE insertion and their effects on gene expression are reviewed extensively in Hirsch and Springer (2017). It is also theorized that the RdDM pathway originally evolved in plants to combat the potential deleterious consequences from transposon mobilization as the pathway acted as a damage control method by having TE transcripts processed into siRNAs, which guide DNA methylation, and consequently silence the

TEs and prevent genome damage (Bräutigam & Cronk, 2018). In nascent polyploids, it stands to reason this method would be of importance to manage the influx of genomic content and attenuate the effects of transposon insertions through extensive silencing. In *B. napus* seeds, TEs account for the majority of all siRNA expression at all stages of seed development. Early in seed development, it is possible that certain TE regions are constitutively silenced via high siRNA transcription that transitions into global silencing of TEs in seed maturation. Both the global increase in CHH methylation and increase in number of siRNAs invoked from TEs in maturing seeds may be indicative of this phenomenon. As *B. napus* is a recent amphidiploid, this ubiquitous silencing effect throughout both subgenomes implies the necessity to control these genome elements in early sporophytic development. It is this same polyploid complexity, however, that has been known to contribute to genomic plasticity, and is a critical component of artificial selection in crop plants (Dubkovsky & Dvorak, 2007). Therefore, global TE silencing via RdDM in seeds may mitigate regulatory errors during embryo morphogenesis and maturation, but is removed or lost promptly upon germination, as is observed in Arabidopsis (Narsai, Gouil, et al., 2017), Capsicum (Portis et al., 2004), Triticum (Meng et al., 2012) and Oryza (Narsai, Secco, et al., 2017).

3.4.6 Fractionated ancestral genomes are more dissimilar in protein-coding regions than in TEs and siRNAs

It is known that even in extensively polyploid clades that collinear blocks are preserved throughout generations of ploidy changes and genomic stress (Hardigan et al., 2020). However, while stability of collinear regions may be preserved over evolutionary time in polyploids, it is less understood whether the non-protein coding elements of collinear regions in allopolyploid progenitor genomes experience the same retention. While the LF, MF1 and MF2 genomes are

generated by referencing collinear genes in Arabidopsis, the similarities may be overestimated. Nevertheless, the epigenome is critically important to the management of genome stability, and provides a provisional solution to times of genomic shock (Ha et al., 2009; Springer et al., 2015). Furthermore, while studies exist on the consequences of polyploidization on genome evolution, there remain inconsistencies amongst plant taxa regarding their evolutionary trajectory post-polyploidization. The genome balance hypothesis is one of the withstanding ideas in polyploid evolution that has endured scientific inquiry (Blanc & Wolfe, 2004).

Genomic balance describes the phenomenon of duplicated genes that settle into singletons over time via fractionation and deletions, whilst regulatory components like TFs, signal transductors and others are more frequently retained in their duplicated state. By extension, this means that coding elements of the genome undergo more dramatic changes over evolutionary time (fractionation) than what is expected to occur within non-coding regions. This is also seen in the confounding nature of angiosperm genomes that are known to vary 2400-fold in size with relatively little changes in the amount of coding genes (Dodsworth et al., 2015). In our study, we demonstrate that DNA methylation on genes and promoter regions are divergent across subgenomes even when accounting for ancestral fractionation where the same fractionated genomes exhibit different methylation patterns throughout development depending on the subgenome of origin. This is consistent with the idea of genome balance where genes undergo more robust alterations post-polyploidization than other components of the genome, and implies these modifications also entail shifts in methylation architecture. Given that methylation of genes and promoters in the fractionated genomes were more similar in the same subgenome, it is likely this divergence occurred post-WGT and precedes the allopolyploidzation event. Furthermore, methylation of TEs and siRNA loci showed distinct

subgenomic bias but were similar in their methylation patterns in the same fractionated genome across the subgenomes. For example, CHG/CHH methylation on siRNA loci and TEs of LF-A and LF-C is more similar than global comparisons between the An and Cn subgenomes. This implies that the TEs and siRNA loci of the genome have retained epigenetic patterning from the ancient triplication in CHG and CHH contexts specifically. Genes and promoters have likely incurred enough genomic modification throughout evolutionary time to be highly divergent in all cytosine contexts throughout seed development. In contrast, TEs and siRNA loci are more similar in GLOB, MG and GLOB-MG comparisons even in different fractionated genomes and different subgenomes. The insignificant comparisons in this work suggest that DNA methylation in fractionated genomes is likely to hold a withstanding landscape in non-protein coding genome elements like TEs and siRNAs than in protein-coding gene regions and their promoters. Research on other species within the Brassiceae that have not experienced recent amphidiploidization (such as *B. rapa*) may be more ideal candidates to study the specific outcome of genome fractionation as they do not have additional variables from another subgenome complicating their epigenome. Nevertheless, *B. napus* presents a unique opportunity to study fractionated genomes from within an interspecific hybrid.

Together, this work provides evidence for both epigenetic bias of the Cn subgenome within the *B. napus* seed, and the retention of epigenetic landscapes across non-protein coding elements of the genome. Moreover, our data suggest that genome fractionation has not substantially changed the methylation profile of both the transposons and the siRNAs that are invoked during seed development. Taken together, the epigenetic architecture of the seed may be an important factor in determining the evolutionary trajectory of *B. napus* and is especially relevant to polyploid crop plants.

3.5 Experimental Procedures

The DH12075 genome sequence and annotation were provided by Drs Isobel Parkin and Steve Robinson via the Canadian Canola Genome Sequencing industry consortium project.

3.5.1 Plant growth conditions and sample collection

Methylome and siRNA analyses were performed in *B. napus* cv. DH12075. Plants were grown in growth chambers under long-day conditions at 22°C. Flowers were hand-pollinated and collected at 0 days post-fertilization (DPF; OV), 7 DPF (GLOB), 10 DPF (HEART), 28 DPF (MG) and 35 DPF (DS) for sRNA-seq, and at GLOB and MG stages for methylome analysis. Seeds were manually harvested from siliques and ground in liquid nitrogen for nucleic acid extractions. DNA was isolated from seeds using the Qiagen Plant DNeasy Mini Kit. RNA was isolated from seeds using the Qiagen miRNeasy mini kit. Three biological replicates were used for WGBS and sRNA-seq.

3.5.2 Collinearity map and TE annotation of the DH12075 genome

Broad collinear blocks are observed in both An and Cn subgenomes of *B. napus* spanning the LF, MF1 and MF2 genomes (Figure 3.1b). Here, we annotated collinear blocks using the ancestral Brassica karyotype detailed in Liu et al. (2014), which is generated using alignments to A. thaliana to infer ancestral collinear blocks. For the purposes of this analysis, we include inversions as valid collinear regions as inversions primarily affect recombination. The coordinates for all collinear regions within the LF, MF1 and MF2 genomes are recorded in Dataset S7. Further, we used RepeatMasker to broadly identify low-complexity REs within the DH12075 genome, which are annotated here as TEs and include both retrotransposons and DNA transposons. RepeatMasker is not sensitive to discovery of novel transposons with low copy number, and as such this analysis focuses on TEs with conspicuous repetition.

3.5.3 Methylome library preparation and sequencing

Library preparation for bisulfite sequencing was performed at Genome Quebec. DNA was sheared via sonication on the Covaris platform prior adaptor ligation and bisulfite conversion using the EZ-96 DNA Methylation-Lightning Kit (Zymo®, Irvine, CA, USA). Libraries were then amplified using the DNA UltraII Kit (NEB®, Ipswich, MA, USA). Libraries were sequenced on the HiSeq X platform (2 lanes, 150 bp PE). Bisulfite-sequencing was performed on three replicates for both stages. A total of 1.24 billion 150-bp paired-end reads were generated from the GLOB and MG libraries. Chloroplast genome cytosine methylation was used as a proxy to estimate bisulfite non-conversion rate – conversion efficiency is estimated to be ~98%. Bisulfite-sequencing libraries (> 97% bisulfite conversion rate; Dataset S8) for both samples. Approximately 67% of paired reads aligned to the *B. napus* cv. DH12075 genome (Dataset S8). Differential methylation was determined over 400-bp windows using MethylKit (Akalin et al., 2012). Promoter regions were calculated as 1000-bp regions upstream of the transcription start site (TSS) of a gene. To identify genes with upstream regulatory regions containing differentially methylated regions (DMRs), overlap of the 400-bp DMR regions and the 1000-bp upstream promoter regions was calculated using BEDtools. Homologous genes in the Darmor Bzh annotation were identified, and enrichment of transcript accumulation patterns was performed using SeqEnrich (Becker et al., 2017).

3.5.4 Bisulfite sequencing analysis

Quality trimming and adaptor removal were performed using Trimmomatic (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Successfully paired reads were aligned to the DH12075 genome using BSMap2, and cytosine methylation levels were called using the methration script (Xi & Li, 2009). Individual cytosines had to be covered by at least 4

reads to be considered for analysis. Average DNA methylation over 1-kB regions, and average methylation ratios of genes, TEs and promoters were calculated using BEDtools (Quinlan & Hall, 2010). We also examined the effect of bin size on determining subgenome bias in average methylation and found that the C subgenome was detected as more highly methylated regardless of bin size (Dataset S8). Differential methylation was determined over 400-bp windows using MethylKit (Akalin et al., 2012). Promoter regions were calculated as 1000-bp regions upstream of the TSS of a gene. We did not attempt to normalize any differences between the subgenomes, be it chromosome length, TE composition, gene number or siRNA loci, as we believe these differences are important and would diminish the biological reality of this complex polyploid.

3.5.5 sRNA library preparation and sequencing

Total RNA collected from seeds was used for library prep with the NEBNext® Small RNA Library Prep Set for Illumina® (San Diego, CA, USA) (Multiplex compatible). Libraries were pooled and sequenced on the Illumina HiSeq2500 (2 lanes, 50 bp SE). Quality trimming and adaptor removal were completed with Trimmomatic (Dataset S8). Surviving reads were aligned to the DH12075 genome using Bowtie on default settings, with an average of 87% alignment. Aligned reads were then filtered for rRNA, tRNA, snRNA and snoRNA species by aligning them to known records in NCBI under all Viridiplantae.

3.5.6 siRNA identification and analysis

The siRNA prediction was done using Short Stack on default settings (Axtell, 2013). siRNA locus identification followed the protocol used in Medina et al. (2018). Briefly, all replicates were analyzed using Short Stack on default settings to generate candidate siRNA loci. siRNA loci were merged across all samples when separated by ≤ 1 bp. The merged loci were

then re-input into Short Stack in count mode to quantify RPM at the merged loci. Promoter regions in the genome were annotated as 1 kbp upstream from coding genes. Homologous genes in the Darmor Bzh annotation were identified, and GO enrichment on genes downstream of siRNA loci localized to the corresponding promoter was performed using SeqEnrich (Becker et al., 2017).

3.6 Supplemental Figures

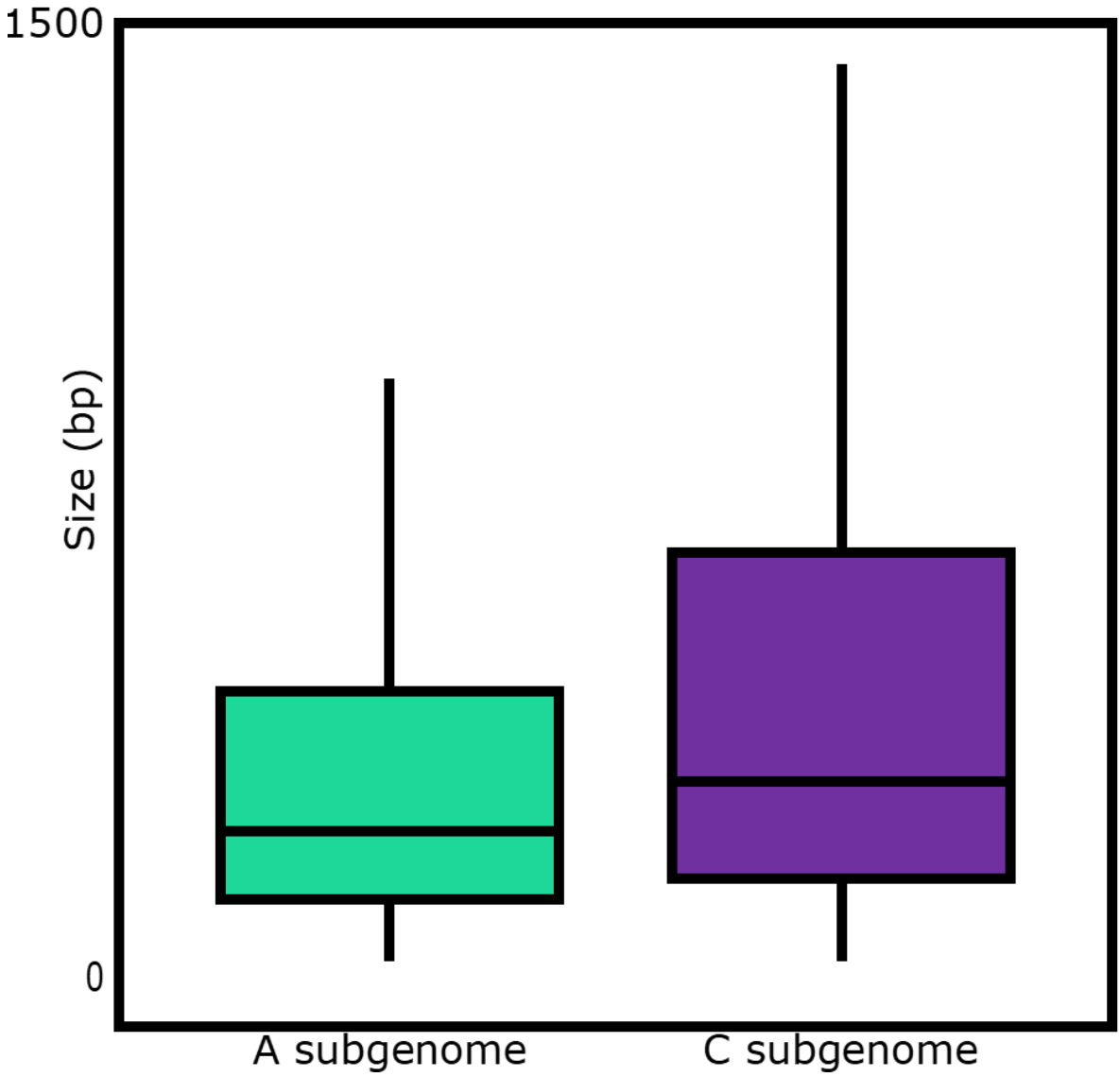


Figure S3.1. Quantile boxplots showing the distribution of size of transposable elements (TEs) in the A^n (teal) and C^n (purple) subgenomes of *B. napus* cv. DH12075.

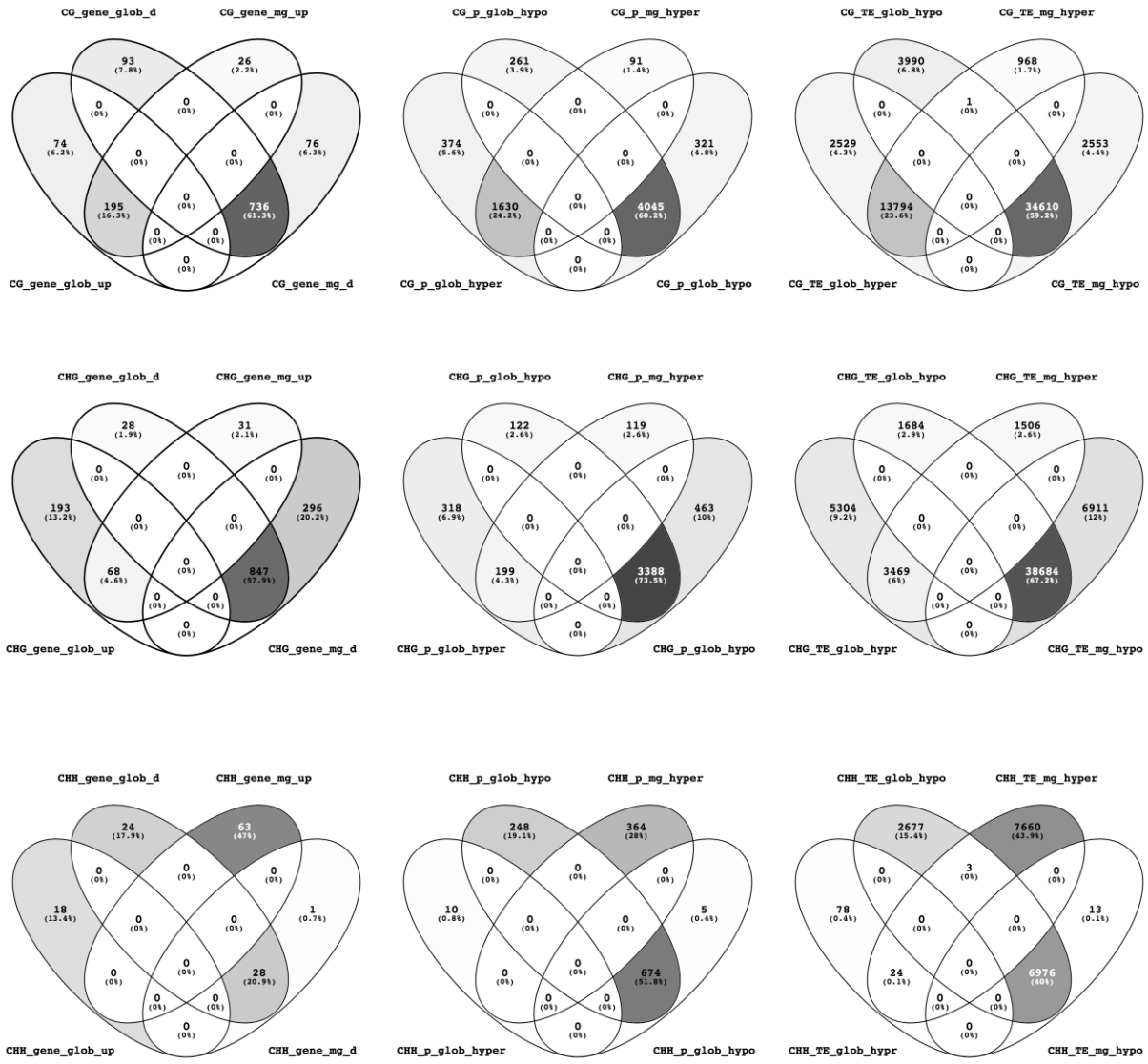


Figure S3.2. Differential methylation between leaves and seeds. Venn diagrams showing hyper and hypo methylated genes, promoters and repetitive elements in GLOB and MG seeds relative to leaves. Differential methylation was determined using MethylKit over 400bp windows of the genome. Intersection with transposable elements was determined using BEDtools.

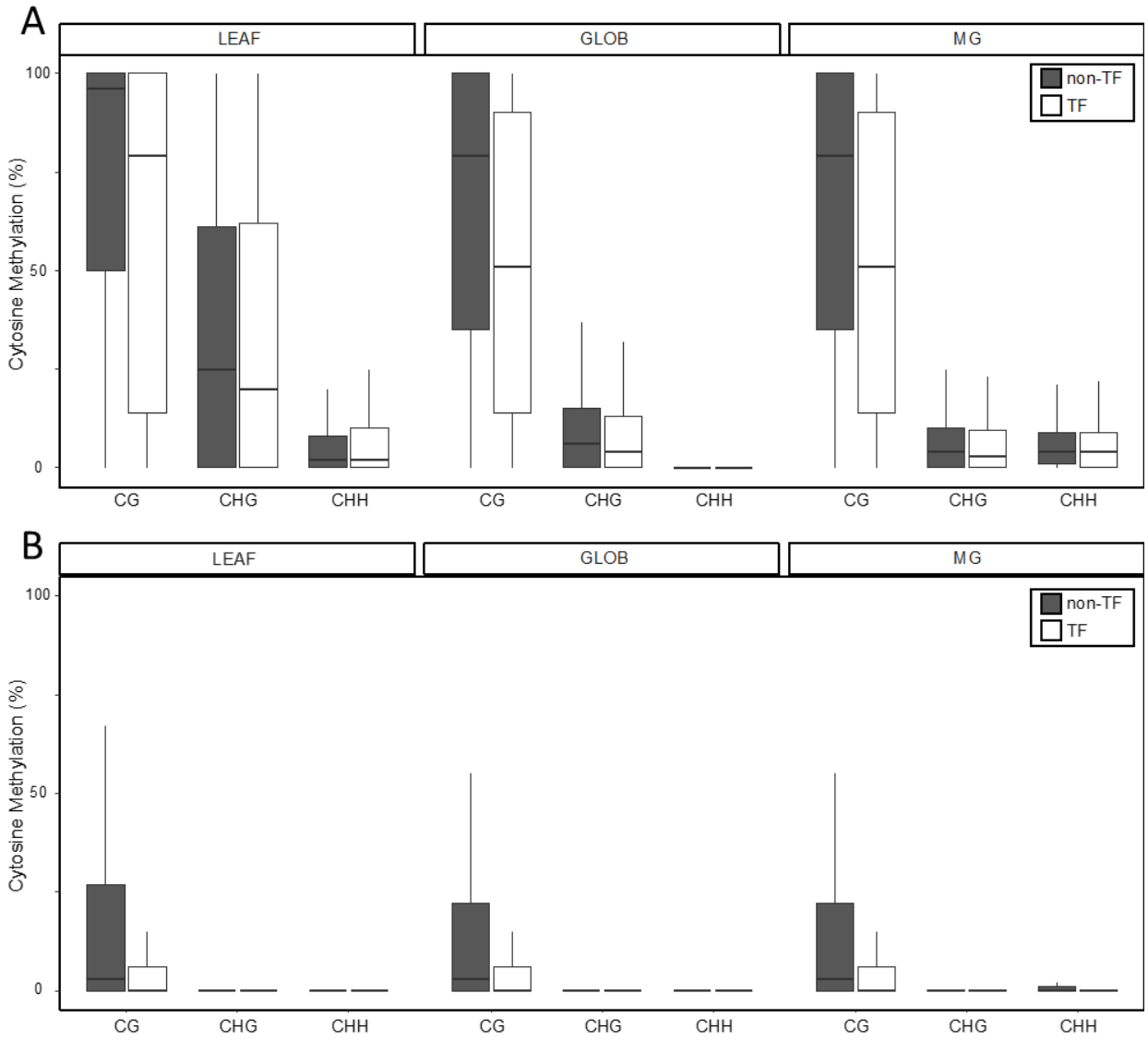


Figure S3.3. Methylation levels of transcription factors. Quantile boxplots illustrating the methylation levels of gene bodies (A) and promoters (B; 1kb upstream of TSS) of non-transcription factor genes (grey) and transcription factors (white) in the CG, CHG, and CHH contexts in leaves, and globular (GLOB) and mature green (MG) seeds).

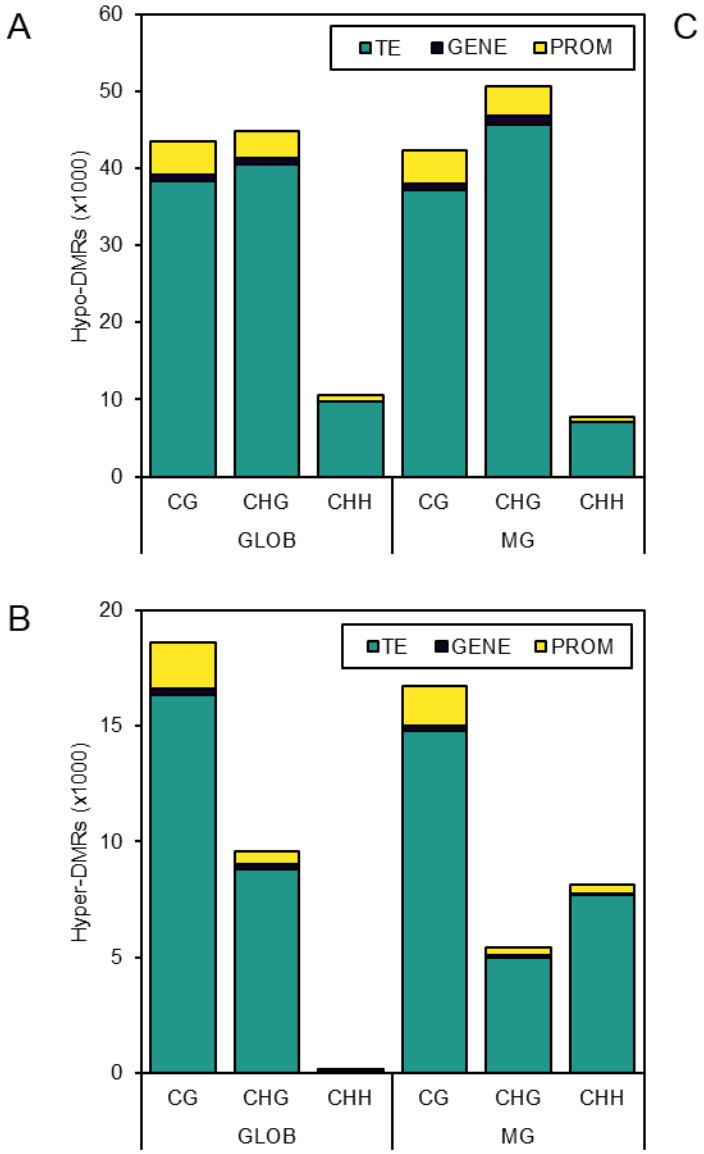


Figure S3.4. Shifts in DNA methylation between in development. (A-B) Differential cytosine methylation in seeds relative to leaves. The GLOB and MG seed methylomes were compared to the leaf methylome (Parkin et al., unpublished). Differential methylation was calculated for all three cytosine contexts (CG, CHG, CHH) over genes, promoters and repetitive elements (REs). Number of differentially methylated TEs (teal), gene bodies (purple) and promoters (yellow) that are (A) hypomethylated in the GLOB or MG seed relative to leaves or (B) hypermethylated in the GLOB or MG seed relative to leaves.

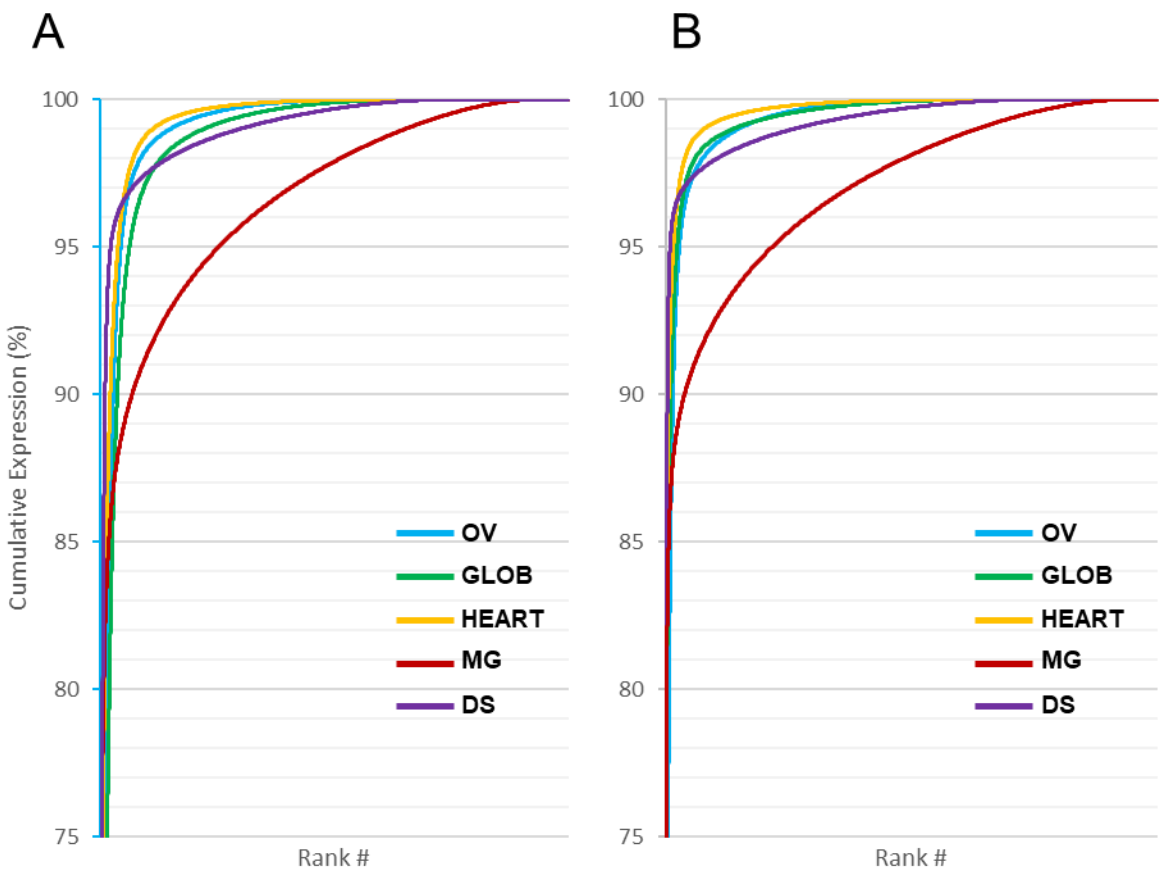


Figure S3.5. siRNA read accumulation throughout seed development in the *B. napus* subgenomes. (A) An subgenome and (B) Cn subgenome.

3.7 Contribution of Authors

DJZ, DK, NPV performed experiments. IAP and SR provided the DH12075 genome and annotation, and provided feedback on the manuscript. DJZ, DK and MFB prepared the manuscript.

3.8 References

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F., Figueroa, M.E., Melnick, A. et al. (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. BMC Genomics, 13, R87.

Alexander-Webber, D., Abbott, R.J. & Chapman, M.A. (2016) Morphological convergence between an allopolyploid and one of its parental species

correlates with biased gene expression and DNA loss. The Journal of Heredity, 107, 445–454.

Ario, M.D., Griffiths-Jones, S. & Kim, M. (2017) Small RNAs: Big Impact on Plant Development. Trends in Plant Science, 22, 1056–1068. Available from: https://doi.org/10.1016/j.tplants.2017.09.009

Axtell, M. (2013) ShortStack: comprehensive annotation and quantification of small RNA genes. RNA, 19(6), 740–751.

Becker, M.G., Walker, P.L., Pulgar-Vidal, N.C. & Belmonte, M.F. (2017) SeqEnrich: a tool to predict transcription factor networks from co-expressed Arabidopsis and *Brassica napus* gene sets. PLoS One, 12, e0178256.

Bird, K., Niederhuth, C., Ou, S., Gehan, M., Pires, J.C., Xiong, Z. et al. (2021) Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. New Phytol, 230, 354–371.

Blanc, G., Hokamp, K. & Wolfe, K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Research, 13, 137–144.

Blanc, G. & Wolfe, K.H. (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell, 16, 1679–1691.

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F. et al. (2017) DNA methylation dynamics during early plant life. Genome Biology, 18, 179. Available from: https://doi.org/10.1186/s13059-017-1313-0

Bräutigam, K. & Cronk, Q. (2018) DNA methylation and the evolutionary complexity of plants. Frontiers in Plant Science, 9, 1447.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P.P., Tang, H., Wang, X. et al. (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science, 345, 950–953.

Chen, C., Zeng, Z., Liu, Z. & Xia, R. (2018) Small RNAs, emerging regulators critical for the development of horticultural traits. Horticulture Research, 5, 63. Available from: https://doi.org/10.1038/s41438-018-0072-8

Chen, M., Lin, J.Y., Hur, J., Pelletier, J.M., Baden, R., Pellegrini, M. et al. (2018) Seed genome hypomethylated regions are enriched in transcription factor genes. Proceedings of the National Academy of Sciences of the United States of America, 115, E8315–E8322.

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M. & Wang, X. (2018) Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants, 4, 258–268.

Coate, J.E., Bar, H. & Doyle, J.J. (2014) Extensive translational regulation of gene expression in an allopolyploid (Glycine dolichocarpa). Plant Cell, 26, 136–150.

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.

Davis-Richardson, A.G., Russell, J.T., Dias, R., Mckinlay, A.J., Canepa, R., Fagen, J.R. et al. (2016) Integrating DNA methylation and gene expression data in the development of the soybean-bradyrhizobium N2-fixing symbiosis. Frontiers in Microbiology, 7, 952.

Dodsworth, S., Leitch, A.R. & Leitch, I.J. (2015) Genome size diversity in angiosperms and its influence on gene space. Current Opinion in Genetics & Development, 35, 73–78.

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

Dubkovsky, J. & Dvorak, J. (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science, 316, 1862–1866.

Edger, P.P., Smith, R., McKain, M.R., Cooley, A.M., Vallejo-Marin, M., Yuan, Y. et al. (2017) Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell, 29, 2150–2167.

Gao, P., Quilichini, T.D., Yang, H., Li, Q., Nilsen, K.T., Qin, L. et al. (2022) Evolutionary divergence in embryo and seed coat development of U's triangle brassica species illustrated by a spatiotemporal transcriptome atlas. The New Phytologist, 233, 30–51. Available from: https://doi.org/10.1111/nph.17759

Gardiner, L.-J., Quinton-Tulloch, M., Olohan, L., Price, J., Hall, N. & Hall, A. (2015) A genome-wide survey of DNA methylation in hexaploid wheat. Genome Biology, 16, 273.

Grover, J.W., Burgess, D., Kendall, T., Baten, A., Pokhrel, S., King, G.J. et al. (2020) Abundant expression of maternal siRNAs is a conserved feature of seed

development. Proceedings of the National Academy of Sciences of the United States of America, 117, 15305–15315.

Grover, J.W., Kendall, T., Baten, A., Burgess, D., Freeing, M., King, G.J. et al. (2018) Maternal components of RNA-directed DNA methylation are required for seed development in Brassica rapa. The Plant Journal, 94, 575–582.

Guignard, L.M. (1902) La double fécondation chez les Crucifères. Journal de Botanique, 16, 361–368.

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K.D., Chapman, E.J. et al. (2009) Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. Proceedings of the National Academy of Sciences of the United States of America, 106, 17835–17840.

Hardigan, M.A., Feldmann, M.J., Lorant, A., Bird, K.A., Famula, R., Acharya, C. et al. (2020) Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. Frontiers in Plant Science, 10, 1.

Hirsch, C.D. & Springer, N.M. (2017) Transposable element influences on gene expression in plants. Biochimica et Biophysica Acta, Gene Regulatory Mechanisms, 1860, 157–165.

Jenik, P.D., Gillmor, C.S. & Lukowitz, W. (2007) Embryonic patterning in Arabidopsis thaliana. Annual Review of Cell and Developmental Biology, 23, 207–236.

Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M. et al. (2011) Wheat hybridization and polyploidization results in deregulation of small RNAs. Genetics, 188, 263–272.

Kim, M., Ohr, H., Lee, J.W., Hyun, Y., Fischer, R.L. & Choi, Y. (2008) Temporal and spatial downregulation of Arabidopsis MET1 activity results in global DNA hypomethylation and developmental defects. Molecules and Cells, 26, 611–615.

Kirkbride, R.C., Lu, J., Zhang, C., Mosher, R.A., Baulcombe, D.C. & Chen, Z.J. (2019) Correction for “maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in Arabidopsis”. Proceedings of the National Academy of Sciences of the United States of America, 116, 2761–2766. Available from: https://doi.org/10.1073/pnas.1807621116

Leitch, A.R., Soltis, D.E., Soltis, P.S., Leitch, I.J., Pires, J.C., Soltis, D.E. et al. (2004) Recent and recurrent polyploidy in Tragopogon (Asteraceae): cytogenetic, genomic and genetic comparisons. Biological Journal of the Linnean Society, 82, 485–501.

Leviczky, T., Molnár, E., Papdi, C., Őszi, E., Horváth, G.V., Vizler, C. et al. (2019) E2FA and E2FB transcription factors coordinate cell proliferation with seed maturation. Development, 146, dev179333. Available from: https://doi.org/10.1242/dev.179333

Li, A., Liu, D., Wu, J., Zhao, X., Hao, M., Geng, S. et al. (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. Plant Cell, 26, 1–24.

Li, J., Huang, Q., Sun, M., Zhang, T., Li, H., Chen, B. et al. (2016) Global DNA methylation variations after short-term heat shock treatment in cultured microspores of *Brassica napus* cv. Topas. Scientific Reports, 6, 38401. Available from: https://doi.org/10.1038/srep38401

Li, Q., Rana, K., Xiong, Z., Mei, J., Song, H. & Qian, W. (2017) Genetics and epigenetic alterations in hybrid and derived hexaploid between *Brassica napus* and *B. oleracea* revealed by SSR and MSAP analysis. Acta Physiol Plant, 41, 61.

Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F. et al. (2017) Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. Proceedings of the National Academy of Sciences of the United States of America, 114, E9730–E9739.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A.P. et al. (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nature Communications, 51(5), 1–11.

Lu, X., Wang, W., Ren, W., Chai, Z., Guo, W., Chen, R. et al. (2015) Genome-wide epigenetic regulation of gene transcription in maize seeds. PLoS One, 10, 1–20.

Lu, J., Zhang, C., Baulcombe, D.C. & Chen, Z.J. (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. Proceedings of the National Academy of Sciences of the United States of America, 109, 5529–5534.

Matsuoka, Y. (2011) Evolution of polyploid Triticum wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. Plant & Cell Physiology, 52, 750–764.

Matzke, M.A. & Mosher, R.A. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews. Genetics, 15, 394–408.

Medina, C., da Rocha, M., Magliano, M., Raptopoulo, A., Marteu, N., Lebrigand, K. et al. (2018) Characterization of siRNAs clusters in Arabidopsis thaliana galls induced by the root-knot nematode Meloidogyne incognita. BMC Genomics, 19, 943.

Meng, F.R., Li, Y.C., Yin, J., Liu, H., Chen, X.J., Ni, Z.F. et al. (2012) Analysis of DNA methylation during the germination of wheat seeds. Biologia Plantarum, 56, 269–275.

Messing, J. (2009) The polyploid origin of maize. In: Handbook of Maize. Genet Genomics (J.F. Bennetzen and S. Hake, eds). Berlin, Germany: Springer Science, Business Media LLC, pp. 221–238.

Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y.V., Liew, L.C. et al. (2017) Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. Genome Biology, 18, 172. Available from: https://doi.org/10.1186/s13059-017-1302-3

Narsai, R., Secco, D., Schultz, M.D., Ecker, J.R., Lister, R. & Whelan, J. (2017) Dynamic and rapid changes in the transcriptome and epigenome during germination and in developing rice (Oryza sativa) coleoptiles under anoxia and re-oxygenation. The Plant Journal, 89, 805–824.

Osabe, K., Clement, J.D., Bedon, F., Pettolino, F.A., Ziolkowski, L., Llewellyn, D.J. et al. (2014) Genetic and DNA methylation changes in cotton (Gossypium) genotypes and tissues. PLoS One, 9, e86049.

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C. et al. (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. The New Phytologist, 186, 37–45.

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M. & Ainouche, M. (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in spartina. The New Phytologist, 184, 1003–1015.

Parkin, I.A.P., Koh, C., Tang, H., Robinson, S.J., Kagale, S., Clarke, W.E. et al. (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biology, 15, 1–18.

Portis, E., Acquadro, A., Comino, C. & Lanteri, S. (2004) Analysis of DNA methylation during germination of pepper (Capsicum annuum L.) seeds using methylation-sensitive amplification polymorphism (MSAP). Plant Science, 166, 169–178.

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841–842.

Rajkumar, M.S., Gupta, K., Khemka, N.K., Garg, R. & Jain, M. (2020) DNA methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. Communications Biology, 31(3), 1–13.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W. et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature, 463(7278), 178–183.

Shen, E., Zou, J., Hubertus Behrens, F., Chen, L., Ye, C., Dai, S. et al. (2015) Identification, evolution, and expression partitioning of miRNAs in allopolyploid *Brassica napus*. Journal of Experimental Botany, 66, 7241–7253.

Smith, Z.R. & Long, J.A. (2010) Control of Arabidopsis apical-basal embryo polarity by antagonistic transcription factors. Nature, 464, 423–426.

Soltis, P.S. & Soltis, D.E. (2016) Ancient WGD events as drivers of key innovations in angiosperms. Current Opinion in Plant Biology, 30, 159–165.

Song, Q., Guan, X. & Chen, Z.J. (2015) Dynamic roles for small RNAs and DNA methylation during ovule and fiber development in allotetraploid cotton. PLoS Genetics, 11, e1005724.

Springer, N.M., Lisch, D. & Li, Q. (2015) Creating order from chaos: epigenome dynamics in plants with complex genomes. Plant Cell, 28, 314–325.

Tayalé, A. & Parisod, C. (2013) Natural pathways to polyploidy in plants and consequences for genome reorganization. Cytogenetic and Genome Research, 140, 79–96.

Tykarska, T. (1976) Rape embryogenesis I. the proembryo development. Acta Societatis Botanicorum Poloniae, 45, 3–16.

Tykarska, T. (1979) Rape embryogenesis. II. Development of embryo proper. Acta Societatis Botanicorum Poloniae, 48, 391–421.

Tykarska, T. (1980) Rape embryogenesis. III. Embryo development in time. Acta Societatis Botanicorum Poloniae, 49, 369–385.

Vekemans, D., Proost, S., Vanneste, K., Coenen, H., Viaene, T., Ruelens, P. et al. (2012) Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. Molecular Biology and Evolution, 29, 3793–3806.

Wang, F. & Axtell, M.J. (2017) AGO4 is specifically required for heterochromatic siRNA accumulation at pol V-dependent loci in Arabidopsis thaliana. The Plant Journal, 90, 37–47.

Wang, J., Mei, J. & Ren, G. (2019) Plant microRNAs: biogenesis, homeostasis, and degradation. Frontiers in Plant Science, 10, 360.

Wang, J., Tian, L., Lee, H.S., Wei, N.E., Jiang, H., Watson, B. et al. (2006) Genomewide nonadditive gene regulation in arabidopsis allotetraploids. Genetics, 172, 507–517.

Wendel, J.F., Lisch, D., Hu, G. & Mason, A.S. (2018) The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Current Opinion in Genetics & Development, 49, 1–7.

Xi, Y. & Li, W. (2009) BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics, 10, 232.

Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B. & Fischer, R.L. (2006) Regulation of seed size by hypomethylation of maternal and paternal genomes. Plant Physiology, 142, 1160–1168.

Xiao, W., Custard, R.D., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B. et al. (2006) DNA methylation is critical for Arabidopsis embroyogenesis and seed viability. Plant Cell, 18, 805–814.

Xing, M.Q., Zhang, Y.J., Zhou, S.R., Hu, W.Y., Wu, X.X.T., Ye, Y.J. et al. (2015) Global analysis reveals the crucial roles of DNA methylation during rice seed development. Plant Physiology, 168, 1417–1432.

Xu, W., Yang, T., Dong, X., Li, D. & Liu, A. (2016) Genomic DNA methylation analyses reveal the distinct profiles in castor bean seeds with persistent endosperms. Plant Physiology, 171, 1242–1258.

Zhang, M., Kimatu, J.N., Xu, K. & Liu, B. (2010) DNA cytosine methylation in plant development. Journal of Genetics and Genomics, 37, 1–12.

Zhang, Y., Xu, G.H., Guo, X.Y. & Fan, L.J. (2005) Two ancient rounds of polyploidy in rice genome. Journal of Zhejiang University. Science. B, 6, 87–90.

Zhou, M. & Law, J.A. (2015) RNA pol IV and V in gene silencing: rebel polymerases evolving away from pol II's rules. Current Opinion in Plant Biology, 27, 154–164.

4 Gene expression landscape of the *Brassica napus* seed reveals subgenome bias in both space and time

Dylan J. Ziegler[1], Deirdre Khan[1], Jenna L. Kalichuk[1], Michael G. Becker[1], Mark F. Belmonte[1]*
[1]University of Manitoba, Winnipeg, MB

This chapter is being prepared for publication within Plant Physiology

4.1 Abstract

The vast majority of angiosperms are thought to have originated from some form of polyploidization. The amphidiploid *Brassica napus* ($A^nA^nC^nC^n$) contains both complete diploid genomes from its progenitors *B. rapa* (A^n) and *B. oleracea* (C^n). Despite growing knowledge into the gene expression landscape of the *B. napus* seed, little is known about subgenome bias underpinning the development of specific cells and tissues across the seed lifecycle. Here, we present a large-scale transcriptome atlas of the *B. napus* seed, including both maternal seed coat and filial embryo and endosperm subregions. We report on extensive, global C^n subgenome bias throughout development, and use homologous gene pairs to describe how subgenomic bias differs across subregions. We find that subgenome bias is most prominent during early development, and that the maternal subregions experience far more asymmetric transcript accumulation in favour of the C^n subgenome. In particular, the unexpectedly distinct

transcriptome profile of the chalazal pole indicates the unique developmental processes involved within the chalaza. Further, we report that genes integral to seed storage comprise a large portion of the transcriptome of mature seeds, especially within the embryo and that gene pairs previously documented to be instrumental in seed development exhibit low transcriptional bias. Together, this work represents an important synthesis of polyploid transcriptomics in seed biology and provides a comprehensive overview of the *B. napus* gene expression landscape in both space and time.

4.2 Introduction

Seeds are complex yet elegant reproductive structures that provide evolutionary advantages to higher plants. In flowers, the seed is generally comprised of the embryo that goes on to produce the next sporophytic generation of the plant, the endosperm that supports early embryo growth, and the seed coat that helps to protect the seed (Ziegler et al., 2019). Furthermore, these three constituent parts are of different genetic identities. For example, the seed coat originates from the maternal parent while the triploid endosperm, and the diploid embryo, from as a result of a unique double fertilization event characteristic of Angiosperm seeds (Bleckmann et al., 2014). Seed development must therefore manage these genetic identities through a finely orchestrated process to ensure that each constituent part develops simultaneously in tight coordination. This balance has been the subject of many research endeavours in the past which describe the elongation and development of the seed coat(Radchuk and Borisjuk, 2014), the divisions of the embryo (reviewed in (Boscá et al., 2011b), the cellularization of the endosperm (Batista et al., 2019b; Xu et al., 2023), and the cross-talk between the constituent parts (Belmonte et al., 2013b; Figueiredo et al., 2015; Xu et al., 2015; Figueiredo et al., 2016; Doll et al., 2020). These works have contextualized the importance of this

strict coordination amongst seed regions and have advanced our understanding of seed development. However, the genetic complexity underpinning seed development is complicated by the extensive polyploid histories of angiosperms. Polyploidy is ubiquitous in land plant evolution, constituting an estimated 50-70% or more of all extant flowering plants (Masterson, 1994; Otto and Whitton, 2000a), and while autopolyploidy happens in angiosperms, polyploidy more often occurs as a consequence of interspecific hybridization which forms complex amphidiploid genomes. *Brassica napus* L. ($A^nA^nC^nC^n$), is one such species that formed from the interspecific hybridization of *Brassica rapa* L. (A^rA^r) and *Brassica oleracea* L. (C^oC^o) approximately 7,500 – 12,500 years ago as a nascent amphidiploid (Chalhoub et al., 2014c)

As one of the world's most important oilseeds, *B. napus* is of great interest to the scientific community. The Brassicaceae are well known for robust oil reserves in mature seeds, and are used for protein meal and biodiesel applications (DeMarini et al., 2019; So and Duncan, 2021). While our knowledge of *B. napus* seed development is becoming clearer, the underlying gene expression landscape required to make a canola seed is poorly understood especially from the earliest stages of seed development. *B. napus* ovules are comprised of two integuments that encircle the nucellus and female gametophyte. Seed development begins with the ovule (OV), and once fertilized transitions into the globular (GLOB) and then heart (HEART) stages which constitute the morphogenesis phase (Figure 4.1A). Seed storage reservoirs accumulate thereafter, characteristic of mature green (MG) seeds during the maturation phase. As the *B. napus* seed develops, it must coordinate the development of the maternal and filial subregions to orchestrate seed development. This development is further complicated by the presence of two distinct diploid progenitor genomes within *B. napus.*

Our previous work showed that subgenome bias was apparent in transcription factors during seed development and that this bias was dominated by the A^n subgenome in the whole seed and in seed subregions early in development (Khan et al., 2022). Further, subgenome bias of non-protein coding genome elements like DNA methylation and siRNAs were more frequently biased towards the C^n subgenome (Ziegler et al., 2023). Despite the growing body of evidence supporting subgenome bias within the whole *B. napus* seed, we have yet to explore the possibility of subgenome bias in specific cells and tissues across all of seed development at the mRNA level genome-wide. Questions remain however, about the biological processes that occur in a specific cell or tissue of the seed in both space and time. Some recent studies have reported gene expression profiling of *B. napus* seed subregions like the embryo, endosperm, and seed coat (Hong et al., 2017; Liao et al., 2019; Tan et al., 2019; Ziegler et al., 2019; Rong et al., 2021; Gao et al., 2022; Wei et al., 2022). However, these studies do not enable a complete understanding of subgenome bias since none reported on the potential contribution of each subgenome to *B. napus* seed development from fertilization to maturation.

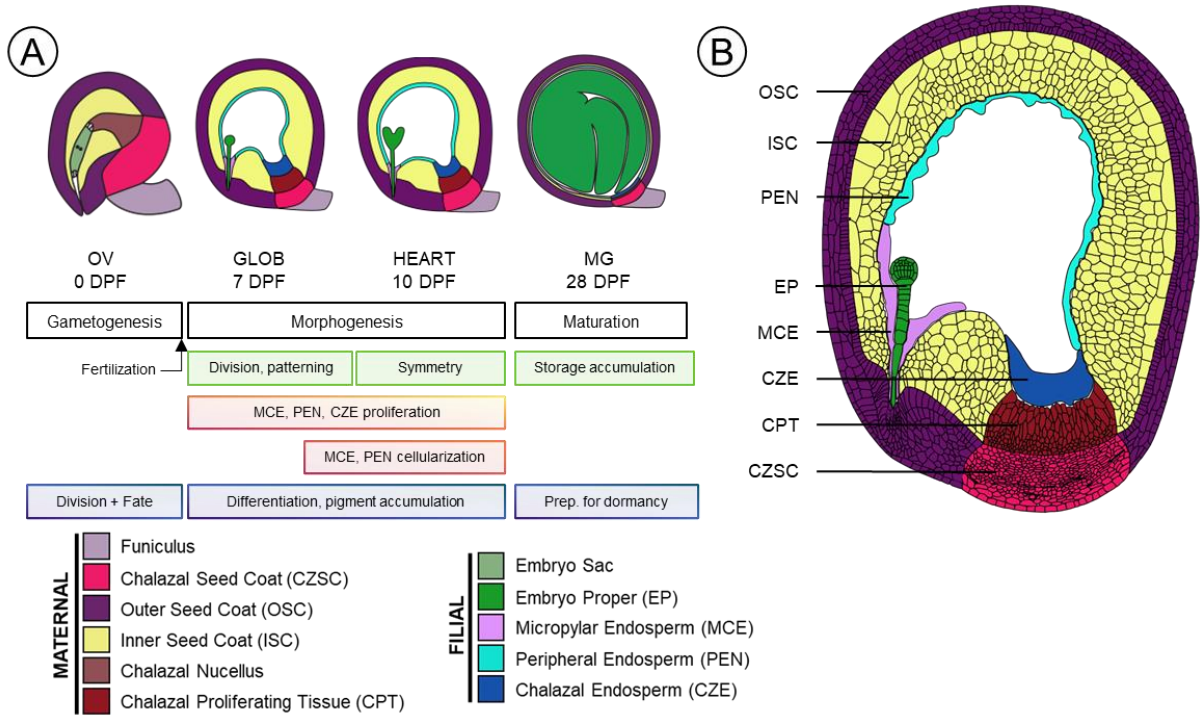


Figure 4.1. (A) Tissue schematic of stages of seed development examined in *Brassica napus*. Biological processes that occur in different subregions are shown in boxes below. The globular stage seed is expanded **(B)**, showing the outer distal seed coat (OSC, indigo), inner distal seed coat (ISC, blue) chalazal seed coat (CZSC, dark teal), chalazal proliferating tissue (CPT, light teal), chalazal endosperm (CZE, yellow), peripheral endosperm (PEN, orange), micropylar endosperm (MCE, pink), embryo (EP, green), and funiculus (FUN, grey).

Our dissection of the *B. napus* gene expression landscape and histological characterization of the seed reveals the integrated biological processes that characterize seed development. We found that each subregion within the seed exhibits distinct transcriptional profiles in the context of subgenome bias. In particular, we find that the chalazal proliferating tissue is distinct from the other subregions, likely due in part to its ontogenetic relationship with the nucellus. We also report that specific gene homologs involved in seed storage accumulation are expressed asymmetrically between the two subgenomes. Analysis of the ancestral subgenomes of this allopolyploid provides seminal evidence into the contributions of

each subgenome that define cell and tissue-specific compartments of both maternal and filial regions that contribute to the making of a seed.

4.3 Results

4.3.1 The *B. napus* seed is a complex yet elegant reproductive structure that can be divided into regions and subregions

We used laser microdissection (LMD) coupled with global RNA sequencing to profile the gene expression landscape of the *B. napus* seed. Cells and tissues were captured at the ovule (ov), globular (g, 7 DPP), heart (h, 10 DPP), and mature green (mg, 28 DPP) stages of development. We profiled gene expression of the maternal seed coat subregions including the inner seed coat (ISC), outer seed coat (OSC), and chalazal seed coat (CZSC), in addition to the chalazal proliferating tissue (CPT). The micropylar endosperm (MCE), peripheral endosperm (PEN), chalazal endosperm (CZE), as well as the embryo proper (EP) were captured from the filial regions of the seed (Figure 4.1A). The EP was further dissected into the embryonic root (ROOT) and cotyledons (COT) in mg seeds. Hereafter, the seed subregion and its corresponding stage are referred to with abbreviations of the stage of development followed by the subregion (ie. ovCPT refers to the CPT of ovules). Further, while ovules are defined as the female gametophyte encircled by elongating integuments which later fuse to form the seed coat, they are referred to as "seed coat" subregions for consistency in-text.

Brassica napus ovules have a distinct chalazal receptacle perpendicular to the micropylar aperture wherein the CZSC is comprised of smaller condensed cells and the cells of the ISC and OSC are comparatively isodiametric with a palisade layer at the interface of the two. Post-

fertilization, the elongation of the fusing seed coat migrates the micropylar pole to be parallel to the chalaza with the MCE encircling the suspensor and is anchored at the fusion point of the micropylar aperture basal to the EP (Figure 4.1B). The endosperm forms a thin syncytium surrounding the inner periphery of the ISC, which is referred to as the PEN. The endosperm additionally forms a syncytial, cyst-like body at the chalazal pole, referred to as the CZE. At the interface between the CZE and CZSC lies the maternally-derived CPT, a maternally-derived subregion. The CPT in *B. napus* is particularly large relative to other plants in the Brassicaceae, proliferating to a large mass of cells at the interstice of the filial CZE and the maternal CZSC post-fertilization (Figure 4.1B).

4.3.2 The C^n subgenome asymmetrically accumulates transcripts across seed development

Collectively, we identified 55,839 genes in our LMD dataset that were not detected in the whole seed thus increasing transcript detection by 44% (Figure S1). Genes belonging to the C^n subgenome were overrepresented compared to the A^n subgenome at both the level of detected transcribed genes (Figure 4.2A) and the total sum of transcript activity as represented by TPM (Figure 4.2B). The seed subregions were universally biased towards the C^n subgenome compared to the whole seed, with both more genes detected and higher transcript accumulation from the C^n subgenome. For example, at the genome level, subgenome bias of total TPM ranged from 48.9% A^n/51.1% C^n in the DS, to 51.1% A^n/48.9% C^n in MG seeds. In contrast, subgenome bias within subregions was most asymmetric in the mgCOT (17.6% An/82.4% Cn) and most equivalent in the gCPT (45.4% A^n/54.5% C^n) (Figure 4.2B). Further, the 500, 1000, 1500, and 5000 most highly transcribed genes across all stages of seed development as well as in all subregions were either not biased towards any subgenome or biased to the C^n subgenome (Figure 4.2C). At

the broad transcriptomic level, C^n subgenome bias was ubiquitous in all seed subregions, both in genes detected as well as in total transcripts.

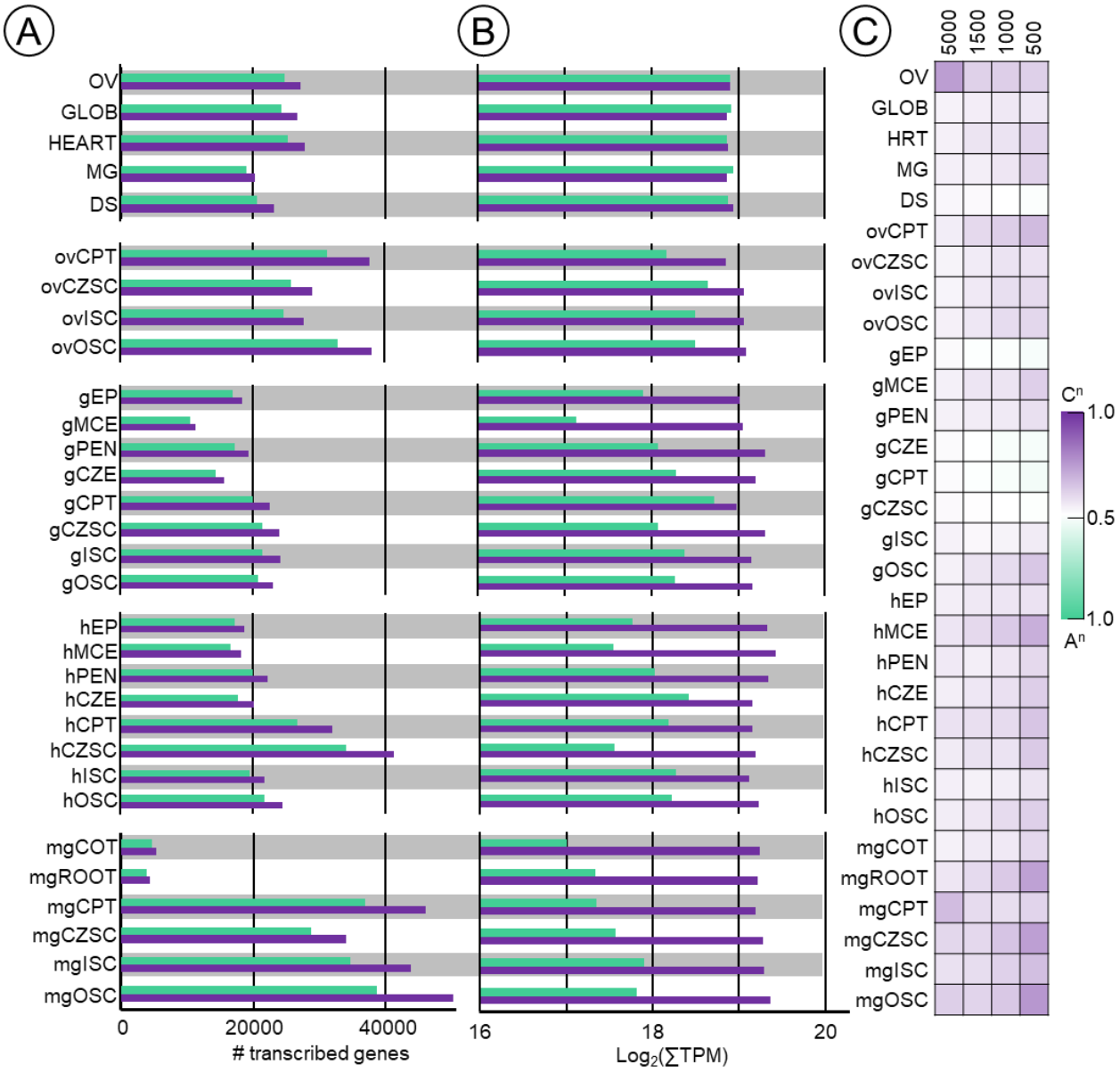


Figure 4.2. Subgenome representation in space and time as **(A)** number of transcribed genes (TPM>1), **(B)** summed TPM (middle), and **(C)** proportion of genes expressed in the top 5000, 1500, 1000, and 500 genes belonging to each respective subgenome (right).

4.3.3 Subgenome bias of homologous gene pairs were most divergent in the ovCPT compared to all other maternal subregions

To further assess subgenome bias in the maternal subregions, we clustered homologous gene pairs between the A^n/C^n subgenomes based on their expression levels relative to each other (Figure 4.3A). The ovCPT was the most distinct subregion, clustering away from all other stages and subregions. The CPT in seeds undergoing morphogenesis clustered together and shared a larger clade with the ISC and OSC subregions. The mgCPT, in contrast, clustered alongside the mgOSC and the mgCZSC. The ovCPT additionally had the most gene homologs with high subgenome bias ($Log_2(A^n/C^n) \geq 2$ or $Log_2(A^n/C^n) \leq -2$) compared to all other seed subregions and developmental stages (Figure 4.3B). Next, we broke down this cluster analysis to examine gene sets with patterns of subgenome bias that may contribute to the conspicuous asymmetry between the two subgenomes (Figure 4.3C and 4.3D). We selected these clusters for downstream analysis based on 1) extensive bias in one or two subregions at certain developmental stages (Figure 4.4A, Clusters 1 and 6); 2) gene lists with homologous pairs in which subregion(s) were visually distinct from the other maternal subregions (Figure 4.4A, Clusters 2, 4-5); or 3) clusters with minimal subgenome bias (Figure 4.4A, Cluster 3). We refer to these gene lists as clusters as they are derived from clustering analysis, but "cluster" in this case does not refer to biological relationships nor genomic location. We then performed Gene Ontology (GO) enrichment analysis on these gene clusters to investigate the biological processes influenced by subgenome bias.

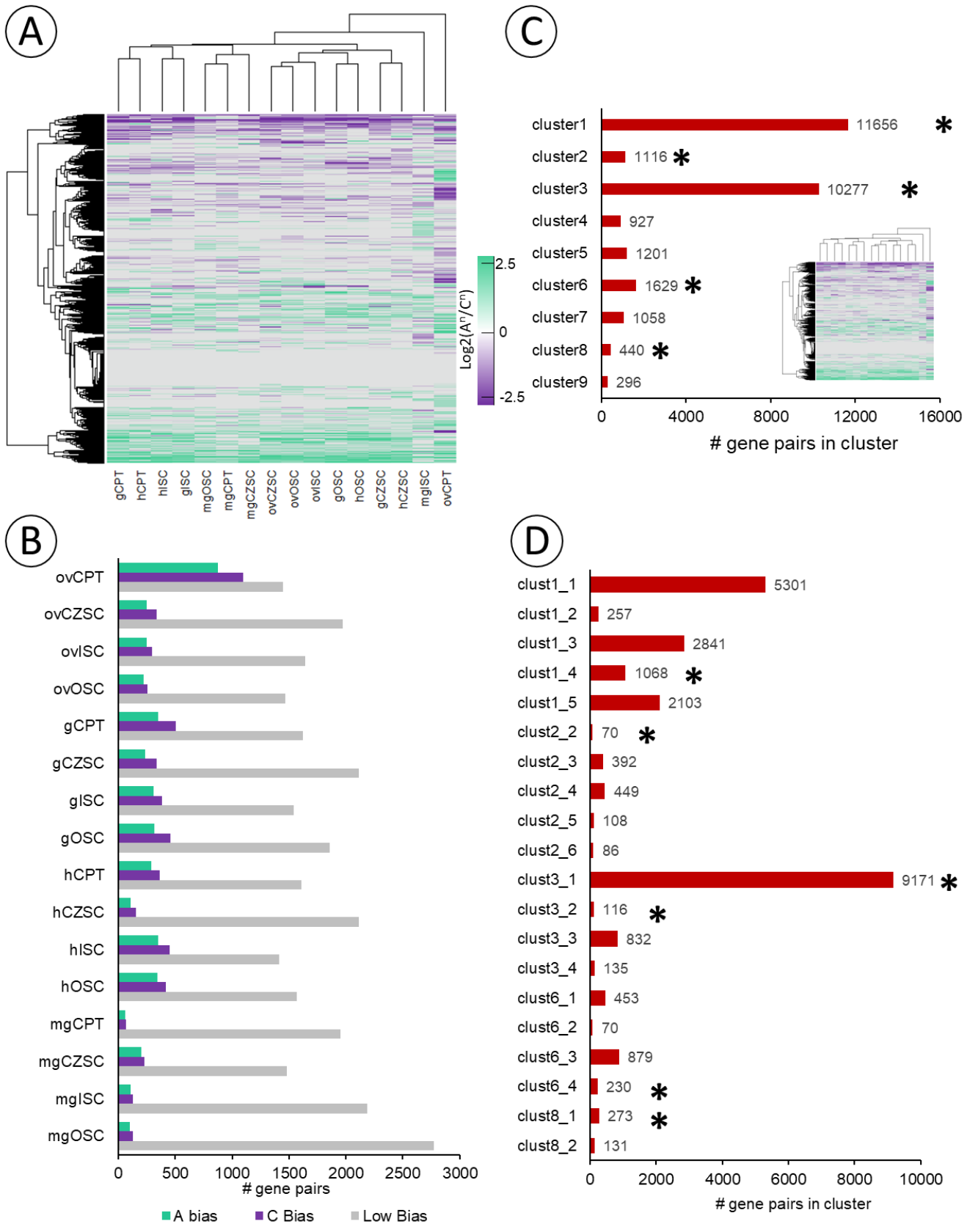

A
gCPT
hCPT
hlSC
glSC
mgOSC
mgCPT
mgCZSC
ovCZSC
ovOSC
ovlSC
gOSC
hOSC
gCZSC
hCZSC
mglSC
ovCPT
Log2(An/Cn)
2.5
0
-2.5
C
cluster1 11656
cluster2 1116
cluster3 10277
cluster4 927
cluster5 1201
cluster6 1629
cluster7 1058
cluster8 440
cluster9 296
0
4000
8000
12000
16000
gene pairs in cluster
B
ovCPT
ovCZSC
ovlSC
ovOSC
gCPT
gCZSC
glSC
gOSC
hCPT
hCZSC
hlSC
hOSC
mgCPT
mgCZSC
mglSC
mgOSC
0
500
1000
1500
2000
2500
3000
gene pairs
A bias
C Bias
Low Bias
D
clust1_1 5301
clust1_2 257
clust1_3 2841
clust1_4 1068
clust1_5 2103
clust2_2 70
clust2_3 392
clust2_4 449
clust2_5 108
clust2_6 86
clust3_1 9171
clust3_2 116
clust3_3 832
clust3_4 135
clust6_1 453
clust6_2 70
clust6_3 879
clust6_4 230
clust8_1 273
clust8_2 131
0
2000
4000
6000
8000
10000
gene pairs in cluster

Figure 4.3. Subgenome bias and clustering analysis of maternal subregions in the *B. napus* seed. (A) Subgenome bias of gene homeologs expressed as $Log_2(A^n/C^n)$ of their respective TPM values. Euclidean distance was used to compute distance between data points. **(B)** Number of gene pairs exhibiting strong genomic bias ($Log_2(A^n/C^n) \geq 2$ ('A Bias') or $Log_2(A^n/C^n) \leq -2$ ('C Bias')), or 'Low Bias' ($0.1 \geq Log_2(A^n/C^n) \geq -0.1$). **(C)** Gene pair list from (A) clustered into k=9 groups, number of gene pairs are tabulated, and heatmaps of all clusters available as Figure SX. Gene pair clusters indicated by an asterisk (*) are carried forward into further analysis in **(D)**. Clusters 1, 2, 3, 6, and 8 were then clustered again to k=5, k=6, k=5, k=4, and k=3 clusters, respectively to parse the gene lists. Clusters in **(D)** with < 60 gene pairs were removed from the bar plot, but included within Dataset SX. Clusters in **(D)** indicated with an asterisk (*) were used in the GO analysis of Figure 4.

4.3.4 Maternal subregions in early seed development are more transcriptomically divergent than that of mature seeds

GO terms associated with cell growth and hormone transport were significantly enriched ($p < 0.001$) in the cluster with ovCPT A^n subgenome bias (Figure 4.4B, Cluster 1). Clusters with ovCPT C^n subgenome bias had enriched GO terms for endomembrane systems and the vacuole in addition to polysaccharide catabolism (arabinan and xylan), karyogamy, and mitochondrial/chloroplast (Figure 4.4B, Clusters 2 and 5). Clusters 4 and 6, which segregated out gCPT and hCPT together as their own clade with subgenome bias distinct from other subregions, were significantly enriched in GO terms relating to epigenetic maintenance, pectin biochemical processing, vacuole organization, autophagy and organ senescence (Figure 4.4B). These clusters were also enriched for biological processes involved in reproduction and reproductive transitions and phase changes. Conversely, the cluster characterized by low subgenome bias was most significantly enriched for processes associated with cell wall organization and loosening, secondary cell wall biogenesis, as well as genes relevant to mitotic/meiotic processes (Figure 4.4B, Cluster 3). At the whole transcriptome level, the CPT was the most diffirent subregion and had the most divergent pattern of subgenome bias

compared to the other subregions at any stage of development. Broadly, the subregions of the MG seeds clustered more closely together than other developmental time points. Conversely, during early seed development, the same subregions more frequently clustered together than the developmental stage. During pre-fertilization and morphogenesis, subgenome bias was most similar in the same subregions. However, these biases were less dramatic once reaching seed maturation and consequently the seed subregions were less transcriptomically distinct in MG seeds at the subgenome level.

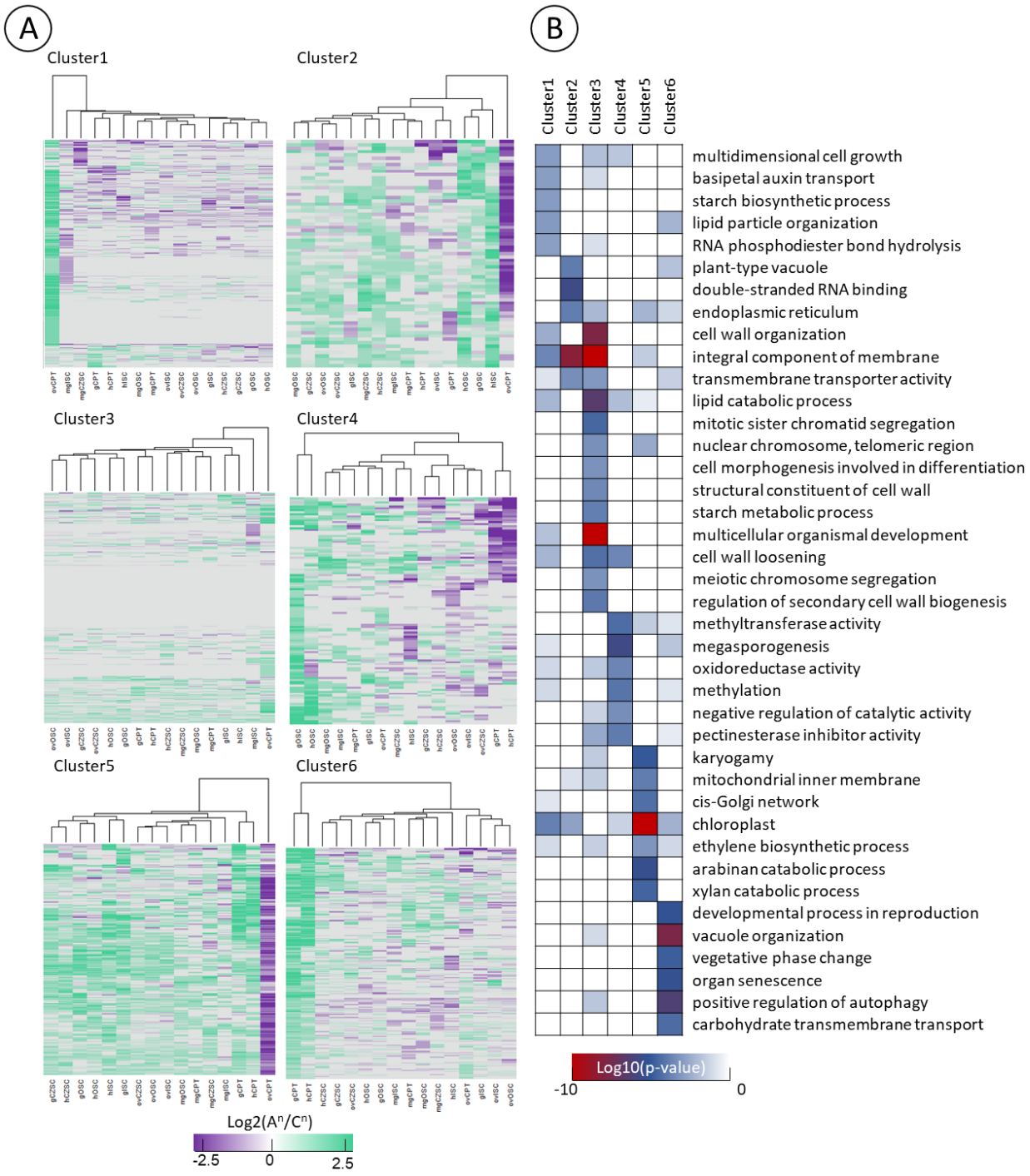


Figure 4.4. (A) Clusters of the data in Figure 3C, selected for subgenome bias in certain stages or lack thereof. **(B)** Gene Ontology (GO) enrichment of the clusters selected in (A), wherein statistically significant GO terms were only selected if >1 homologous gene pair constituted the GO term. Complete GO analysis is documented within Dataset SX.

4.3.5 The CPT is derived from the nucellus and adopts new functions as the seed develops

The CPT is a maternal subregion of unknown function within the Brassicaceae, despite its presence in the model plant *Arabidopsis* and the more modern *Brassica* lineage. Clustering of homologous gene activity implied that the CPT is a distinct subregion of the seed. Next, we examined the gene lists used for GO enrichment in Figure 4.4 and mined the data for genes relevant to seed development, reproduction, and cell wall modification. We found that homologs of *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 27 (XTH27*) were highly expressed early in seed development in the ovCPT with a slight bias for the A^n subgenome, and was also true of another gene in the same endoxyloglucan transferase family, *XTH30* (Figure 4.5A). The methyltransferase *HMT-1* was also transcribed almost exclusively within the ovCPT whie the cysteine protease (*CEP1*) was largely transcribed only within the ISC and CPT during morphogenesis. The *WINDHOSE* (*WIH*) gene homologs were largely transcribed within the CZSC in ovules, followed by a dramatic increase in transcript accumulation of *WIH1* in the ovCPT, and finally with declining transcript abundance of both homologous gene pairs in the CPT and CZSC of HEART and MG stage seeds (Figure 4.5B).

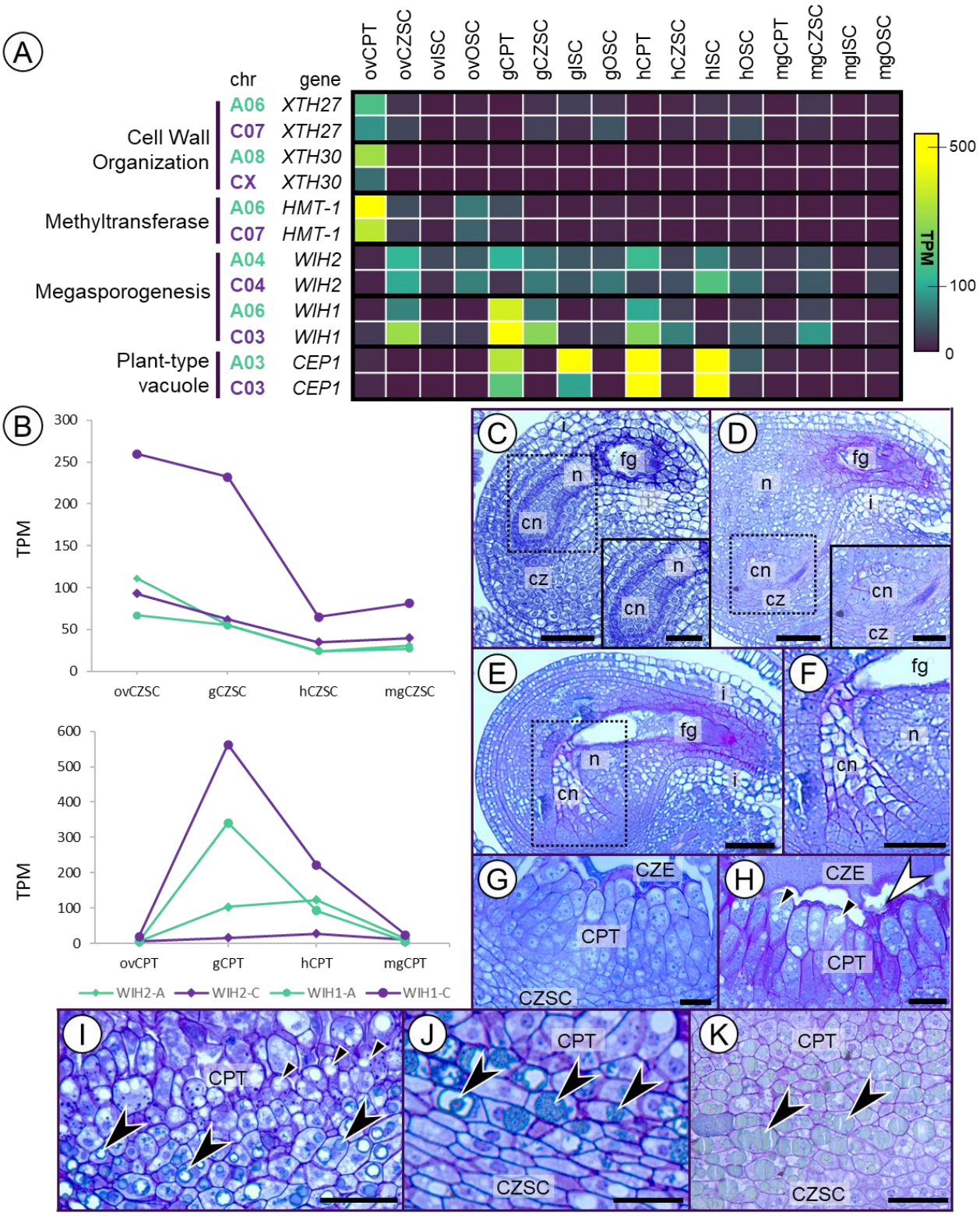

A
ovCPT
ovCZSC
ovISC
ovOSC
gCPT
gCZSC
gISC
gOSC
hCPT
hCZSC
hISC
hOSC
mgCPT
mgCZSC
mgISC
mgOSC
chr
gene
Cell Wall Organization
A06 XTH27
C07 XTH27
A08 XTH30
CX XTH30
Methyltransferase
A06 HMT-1
C07 HMT-1
Megasporogenesis
A04 WIH2
C04 WIH2
A06 WIH1
C03 WIH1
Plant-type vacuole
A03 CEP1
C03 CEP1
500
100
0
TPM
B
TPM
ovCZSC
gCZSC
hCZSC
mgCZSC
ovCPT
gCPT
hCPT
mgCPT
WIH2-A
WIH2-C
WIH1-A
WIH1-C
C
D
E
F
G
H
I
J
K
fg
n
cn
cz
i
CZE
CPT
CZSC

Figure 4.5. (A) Heat map of selected homologous gene pairs belonging to significantly enriched GO terms from the clusters in Figure 4 as they pertain to the developmental shifts within the CPT. **(B)** Expression of *WIH* genes from both A^n and C^n subgenome homeologues in the CZSC and CPT seed subregions spanning OV, GLOB, HEART, and MG seeds. **(C)** Developmental series (stained with toluidine blue and periodic acid and Schiff's reagent) of the early ovule enveloped by both elongating integuments (i) with the tenuinucellate nucellus (n) surrounding the female gametophyte (fg). The nucellus extends from the basal chalazal nucellar body (cn), adjacent to the chalaza (cz). (D) maturing ovule, **(E)** ovule with the cellularized mature fg, **(F)** inset of cn subtending the fg, **(G)** 7 DPP seed chalaza divided into the CPT, chalazal endosperm (CZE), and chalazal seed coat (CZSC), **(H)** 10 DPP seed chalaza with the nucellar lysate (white arrowhead) at the interstice of the CZE and CPT with fragmented vacuoles (black arrow) now populating the cells of the CPT, (I) chalaza at 14 DPP with the incipient deposition of polyphenols in the pigment layer (black arrowheads), **(J)** 21 DPP and **(K)** 28 DPP seed chalaza. Bar = **(C)** 50 μm **(C_{inset})** 25 μm **(D)** 50 μm **(D_{inset})** 25 μm **(E)** 75 μm **(F)** 50 μm **(G)** 50 μm **(H)** 50 μm **(I)** 50 μm **(J)** 30 μm **(K)** 25 μm.

Next, we examined the anatomical development of the CPT at the early ovule stage through to seed maturation. In early ovules the nucellus proliferated as a single cell layer encircling the developing female gametophyte (Figure 4.5C). The nucellar body proliferated shortly afterwards as a robust mass within the ovule, and the chalazal nucellus localized proximally to the developing CZSC with thickened primary walls distinct from the rest of the nucellar body (Figure 4.5D). The chalazal nucellus then elongated and contacted at the antipodal end of the mature female gametophyte (Figure 4.5E and F). At the globular stage, the cells of the chalazal nucellus form conspicuous nuclei, elongate, and increase cytoplasmic density (Figure 4.5G). Seeds at 10 DPP (HEART) accumulate fragmented vacuoles (Figure 4.5H). A distinct cellular lysate of the remnant CPT is seen at the interface between the CPT and the CZE in both GLOB and HEART seeds and is likely derived from the apex of the chalazal nucellus Figure 4.5F. This lysate accumulated as more cell layers of the CPT were lost to the expanding seed. By 14 DPP, the remaining CPT cells were less distinct from the CZSC (Figure 4.5I). The vacuolar fragments accumulated pigment in both the CPT and CZSC once the seed initiated maturation. The cell layers of the CPT proximal to the CZSC accumulated pigment first

by 21 DPP (Figure 4.5J). The remaining layers of the CPT were entirely subsumed into the CZSC anatomically by 28 DPP, filling with pigment bodies and becoming continuous with the seed coat's pigment layer (Figure 4.5K).

4.3.6 The ISC and OSC are genetically and anatomically distinct undergoing different developmental milestones

The seed coat is comprised of the inner and outer integuments that fuse together at the micropylar pole post-fertilization to become the ISC and OSC. The process of seed coat maturation requires extensive cell wall remodeling, cell elongation, and deposition of storage materials. In *B. napus*, we report that many genes responsible for these processes are transcriptomically isolated to specific developmental stages and within specific subregions. Among them, genes implicated in cell wall organization accumulated in the ovule seed coat (*XTH9* endoxyloglucan transferase) and in the seed coat during morphogenesis (*UGE1,* UDP-galactose epimerase) (Figure 4.6A). *XTH9* transcripts were more abundant in the C^n subgenome in all three ovule seed coat subregions and across two sets of gene homologs. *UGE1* transcript accumulation was biased towards the C^n subgenome in the gISC, hOSC, and mgCZSC/OSC. C^n bias persisted in transcripts associated with nuclear fusion (*NUCLEAR FUSION DEFECTIVE 2, NFD2*), starch biosynthesis (*PHOSPHOGLUCAN PHOSPHATASE*, *DSP4* and *FRUCTOKINASE 7, FRK7*), as well as the bidirectional monosaccharide transporter (*SWEET1*) in the mgOSC. A^n subgenome bias was apparent in the homolog pair of *ASPG1* (*ASPARTIC PROTEASE OF GUARD CELL 1*) in the gISC/OSC and hISC/OSC, *SWEET10* in the ovCZSC and hCZSC, as well as in the *ATS3* pair (*EMBRYO-SPECIFIC PROTEIN 3*) at seed maturation.

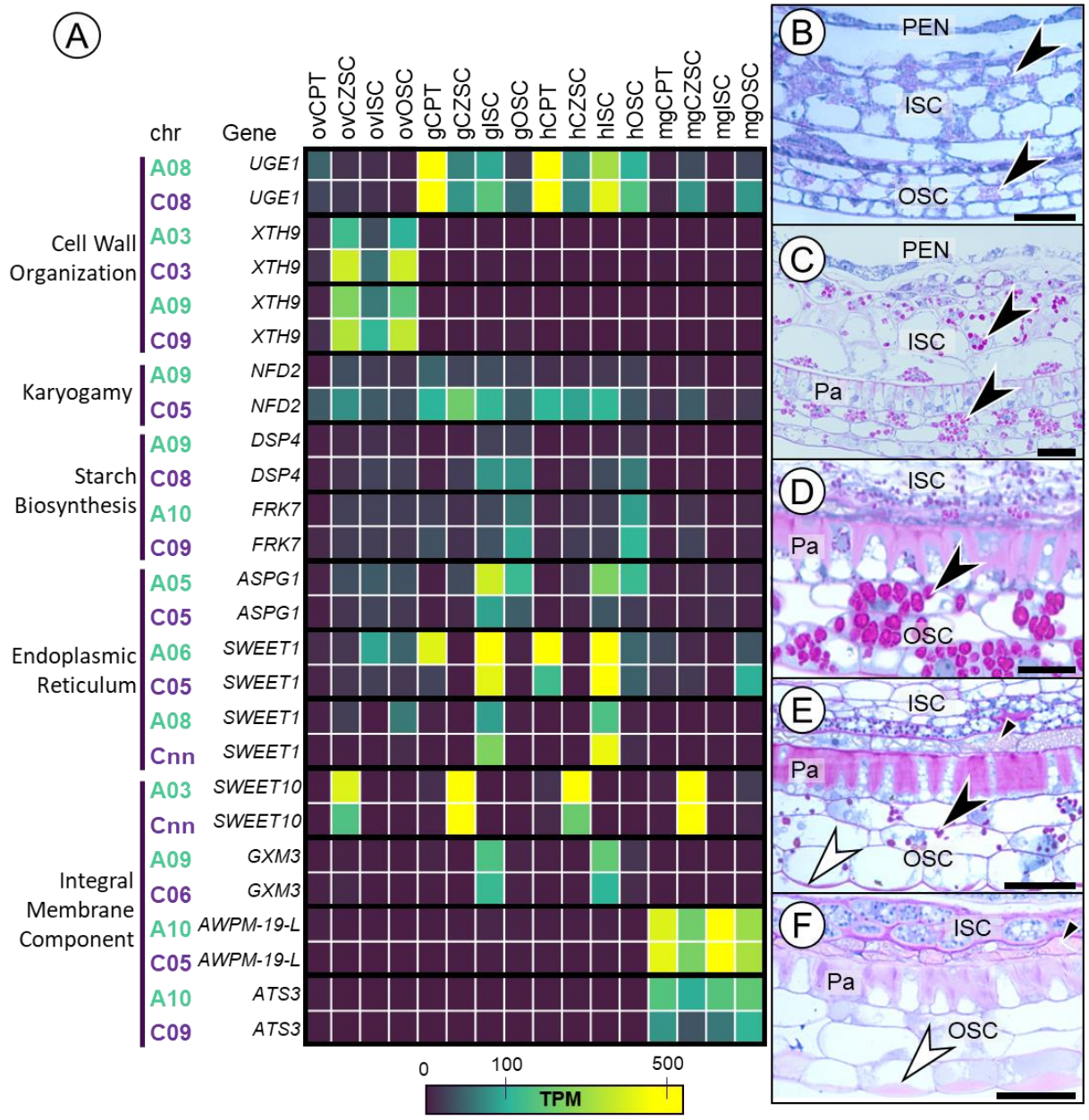


Figure 4.6. (A) Heat map of selected homologous gene pairs belonging to significantly enriched GO terms from the clusters in Figure 4 as they pertain to seed coat development. Longitudinal sections of the developing seed coat (stained with amido black and periodic acid and Schiff's reagent) show the progressive accumulation of starch granules (black arrowheads) in the Inner Seed Coat (ISC) and Outer Seed Coat (OSC) by 7 DPP **(B)**, 10 DPP **(C)**, and 14 DPP **(D)**. Starch reserves dissipate at **(E)** 21 DPP and the inner seed coat accumulates the pigment layer (black arrows) directly beneath the palisade layer of the OSC. Starch reserves are entirely absent by 28 DPP **(F)**. The anticlinal walls of the palisade (Pa) layer of the OSC gradually thickens as the seed develops, and the primary walls of the ISC (black arrowheads) also expand by 28 DPP. Mucilage (white arrowheads) accumulate in the outer periclinal walls of the OSC epidermis. Bar = **(B)** 50 µm **(C)** 40 µm **(D)** 40 µm **(E)** 50 µm **(F)** 50 µm.

Histological analysis on the seed coat revealed cell wall modifications and polysaccharide accumulation throughout seed development. For example, at 7 DPP starch granule accumulation was visible throughout the ISC and within the non-dermal cells of the OSC (Figure 4.6B). Starch granule deposition increased by 10 and 14 DPP while simultaneously the palisade layer of the OSC underwent robust primary wall thickening initially on the internal periclinal wall and following on the anticlinal walls, leaving the outer periclinal wall largely unmodified at the light microscope level (Figure 4.6C and D). Primary wall thickenings also initiated at 14 DPP in the outermost cell layers of the ISC proximal to the palisade layer of the OSC. Cells in this layer accumulated pigment within the vacuole by 21 DPP and the starch granules had been nearly entirely utilized by the developing seed in both the ISC and OSC (Figure 4.6E). Seeds at this stage had begun depositing mucilage within the epidermis of the OSC and the remaining starch granules in the ISC were localized to cells adjacent to inner periphery of the pigment layer. By 28 DPP, these cells contained protein-rich dense cytoplasm with fragmented vacuoles and were surrounded by conspicuously reinforced primary walls, reminiscent of the late stage CPT (Figure 3.6F). The starch reservoirs in both the ISC and OSC were entirely absent at 28 DPP, and mucilage deposits had grown further in the epidermis of the OSC. The palisade layer at 28 DPP had extensive U-shaped wall thickenings with the outer anticlinal wall remaining unmodified. The seed coat layers underwent dramatic anatomical changes throughout seed development. For example, modifications to primary walls, programmed cell death, and rapid consumption of accrued nutrient reservoirs characterized the developmental transitions of the integument cell layers.

4.3.7 Lipid processing exhibits low subgenomic bias in the maternal subregions of the *Brassica napus* seed

While maternal subregions do not accumulate cytoplasmic lipid bodies, lipid homeostasis is important for seed coat waterproofing and maturation. We uncovered gene homologs implicated in lipid homeostasis that showed minimal subgenomic bias (Figure 4.7A). For example, *CUT1,* a gene implicated in very long chain fatty acid synthesis in dermal tissue, was biased towards the A^n subgenome in the gOSC and hOSC. Genes like *EXORDIUM-LIKE 3* (*EXL3*), involved in fatty acid elongation and the *GUARD CELL-ENRICHED GDSL LIPASE* (GGL) implicated in lipid homeostasis both lack subgenome bias. Further, *FAE1* (*FATTY ACID ELONGATION 1*) was transcribed solely in the MG seed in all subregions. Expectedly, we found genes in the oleosin family to also be highly transcribed in all MG seed subregions with minimal subgenome bias across *OLEO2* and *OLEO4* homologs. Despite the high transcript accumulation of oleosins in maternal subregions, lipid droplets are characteristically absent from all maternal seed subregions.

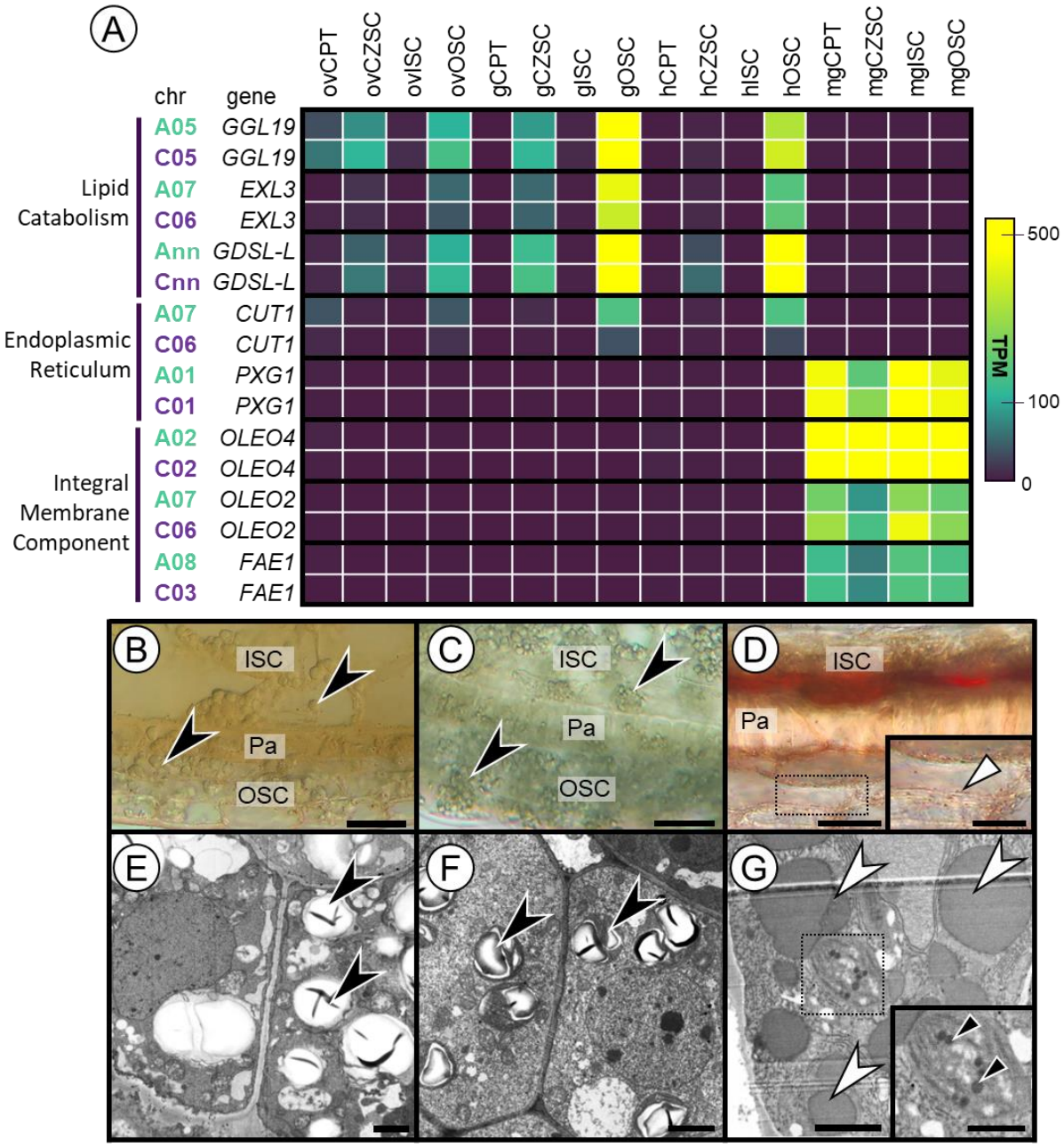


Figure 4.7. (A) Heat map of selected homeologues belonging to significantly enriched GO terms from the clusters in Figure 4 as they pertain to lipid biosynthesis and seed maturation. (B-D) Longitudinal hand sections of the developing seed coat do not stain with Sudan IV at **(B)** 7 DPP and **(C)** 10 DPP. Starch granules (black arrowheads) instead constitute the carbon source during morphogenesis. **(D)** At 28 DPP, lipid accumulation appears in both the Inner Seed Coat (ISC) and Outer Seed Coat (OSC) but is absent in the Palisade (Pa) layer. **(E-G)** TEM micrographs further confirm the presence of starch granules (black arrowheads) throughout the chalazal seed coat. Numerous starch granules fill the CZSC cells of both **(E)** 7 DPP and **(F)** 10 DPP seeds but are eventually used up by **(G)** 28 DPP. Fragmented vacuoles (white arrowheads) gradually fill with proanthocyanidins at 28 DPP. Lipid accumulation occurs in plastids (black arrows,

G_{inset}). Bar = **(B)** 30 μm **(C)** 30 μm **(D)** 40 μm **(D_{inset})** 40 μm **(E)** 2 μm **(F)** 2 μm **(G)** 2 μm **(G_{inset})** 1 μm.

During morphogenesis, the seed coat did not accumulate lipid stores at the level detectable with Sudan IV staining (Figure 3.7B). Instead, lipids accumulated in the palisade layer of the OSC and the outermost layer of the ISC, the pigment layer, at 28 DPP (Figure 3.7C). The lipids in the OSC were primarily localized to elaioplasts. Starch reservoirs populate the CZSC during seed morphogenesis but are eventually consumed and replaced with vacuole fragments filled with proanthocyanidins (Figure 3.7E-G). Lipids appeared late in seed development in all seed coat subregions — the OSC and CZSC with elaioplasts and the ISC with reinforced suberized walls. Much of the machinery characteristic of lipid deposition in seeds does not exhibit substantial subgenome bias.

4.3.8 Filial subregions are transcriptomically less distinct in subgenome bias than the maternal subregions

Data clustering of gene homologs from the filial subregions revealed subgenome bias at the whole transcriptome level was scarce (Figure 4.8A). The mgCOT and mgROOT exhibited the least bias and was likely due in part to the low proportion of genes expressed in the mature embryo (Figure 4.2A). The gEP clustered away from all other filial subregions, while the CZE clustered together throughout morphogenesis. The gEP had the most gene pairs that had a high level of subgenome bias ($Log_2(A^n/C^n) \geq 2$ or $Log_2(A^n/C^n) \leq -2$), at 1715 gene pairs (812 A^n 903 C^n). This made up 11.3% of all identified homologous gene pairs with detectable expression of either gene, while only 8.2% of gene pairs were identified as "low bias" ($0.1 \geq Log2(An/Cn) \geq -0.1$). The mgCOT and mgROOT had the fewest genes with high bias (54 A^n/47 C^n and 58 A^n 45 C^n, respectively), which collectively accounted for 2.3% and 2.8% of the homologs with

detectable expression (Figure 4.8B). All subregions aside from the mgCOT and mgROOT had more gene pairs with high C^n subgenome bias. For example, the gMCE, gPEN, and gCZE had 720, 497, and 502 gene pairs with high C^n bias, compared to 603, 375, and 441 biased toward the A^n subgenome, respectively. The data were then clustered into smaller groups for downstream analysis (Figure 4.8C-D). We identified four clusters of interest based on the same criteria described in Figure 4.3C-D. Clustering analysis revealed bias in the A^n subgenome of the gMCE and C^n bias of the hMCE whereas Cluster 2 formed a clade with distinct Cn subgenome bias of the syncytial endosperm subregions (gPEN, gCZE, hCZE) (Figure 4.9A). Cluster 3 identified A^n bias of the gPEN and a C^n biased clade of the hCZE and gPEN. Cluster 4 had low overall subgenome bias aside from the C^n biased clade that grouped the gMCE, hPEN, and gEP together. Cluster 1 was significantly enriched ($p < 0.01$) for GO terms associated with cell growth, polysaccharide transport and synthesis, Golgi apparatus, and steroid metabolism (Figure 4.9B). Cluster 2 was also significantly enriched for biological processes implicated in Golgi apparatus function, as well as inositol metabolism, cellulose and cell wall processing, and chromatin modification. Cluster 3 was enriched in gene pairs associated with cell division, reproductive development, auxin metabolism, and actin cytoskeleton. Cluster 4 was significantly enriched in vacuolar membrane, vesicle transport, and protein folding GO terms. Notably, while data clustering of the filial regions identified distinct clades, the subregions at large did not exhibit similar genome-wide subgenome bias as seen in the maternal tissues. Further, while many of the data clusters of the maternal subregions identified transcriptomically distinct subregions like the ovCPT, the filial subregions seem to lack these

striking differences.

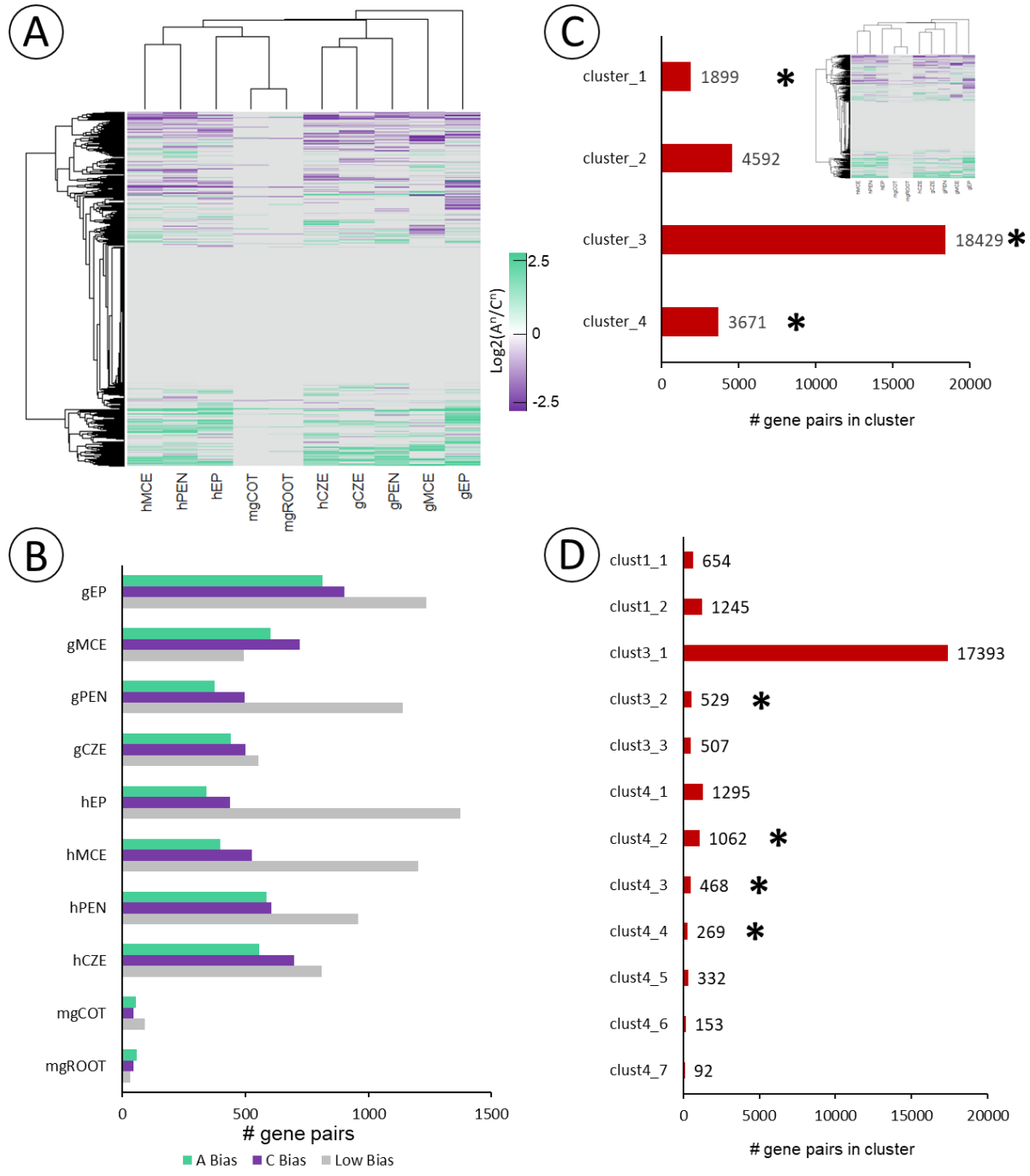


Figure 4.8. Subgenome bias and clustering analysis of filial subregions in the *B. napus* seed. **(A)** Subgenome bias of gene homeologues expressed as $Log_2(A^n/C^n)$ of their respective TPM values. Euclidean distance was used to compute distance between data points. **(B)** Number of gene pairs exhibiting strong genomic bias ($Log_2(A^n/C^n) \geq 2$ ('A Bias') or $Log_2(A^n/C^n) \leq -2$ ('C Bias')), or 'Low Bias' ($0.1 \geq Log_2(A^n/C^n) \geq -0.1$). **(C)** Gene pair list from **(A)** clustered into k=9

groups, number of gene pairs are tabulated, and heatmaps of all clusters available as Figure S4.2. Gene pair clusters indicated by an asterisk (*) are carried forward into further analysis in **(D).** Clusters 1, 2, and 4 were then clustered again to k=2, k=3, and k=7 clusters, respectively to parse the gene lists. Clusters in **(D)** with < 60 gene pairs were removed from the bar plot, but included within Dataset S4.2. Clusters in **(D)** indicated with an asterisk (*) were used in the GO analysis of Figure 9.

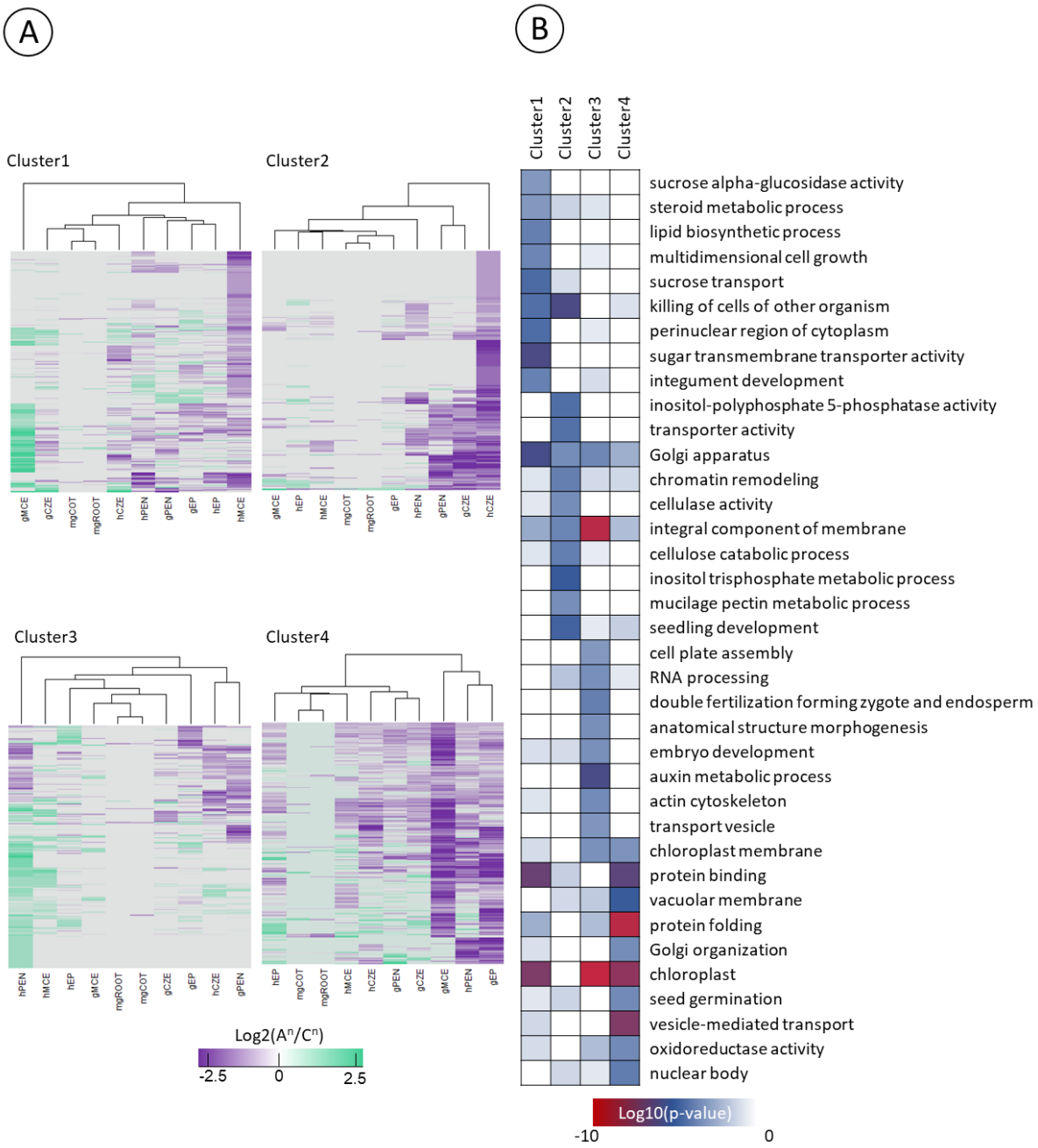


Figure 4.9. (A) Clusters of the data in Figure 8C, selected for exemplifying subgenome bias in certain stages or entire lack thereof. **(B)** Gene Ontology (GO) enrichment of the clusters selected in **(A)**, wherein statistically significant GO terms were only selected if >1 gene homolog pair constituted the GO term. Complete GO analysis is documented within Dataset S4.3.

4.3.9 Endosperm cellularization correlates with Cn subgenome bias of genes involved in cell wall synthesis

The endosperm of *Brassica* forms a contiguous syncytium post-fertilization spanning the micropyle, the periphery, and the chalaza. As in *Arabidopsis*, the endosperm cellularizes, segregating the nuclei via cell plate formation, in a wave beginning at the micropyle and cascading to the chalaza near the end of seed development. From our GO enrichment analysis in Figure 4.9B, we found that some genes implicated in cell wall formation and nutrient reservoir deposition were expressed only in some subregions, and others only in one subgenome (Figure 4.10A). For example, The C^n homolog of *RGP2* (*REVERSIBLY GLYCOSYLATED POLYPEPTIDE 2*) was more abundantly transcribed in the gPEN and hCZE. A similar pattern was observed in the expression pattern of *TBL42* (*TRICHOME BIREFRINGENCE-LIKE 42*), a gene associated with pectin deposition in cell walls where the C^n homolog was highly transcribed in the gPEN, gCZE, and hCZE. In contrast, another gene involved in pectin biosynthesis, *GATL5* (*GALACTURONOSYLTRANSFERASE 5*) was transcribed largely from the C^n homolog in the A^n subgenome. Genes involved in lipid biosynthesis (*FAB2, WRI1*) were slightly more transcribed from the A^n subgenome. *DIM*, a gene involved in brassinosteroid biosynthesis, was observed to have bias in C^n across both homologs in the hEP, hMCE, hPEN, and hCZE.

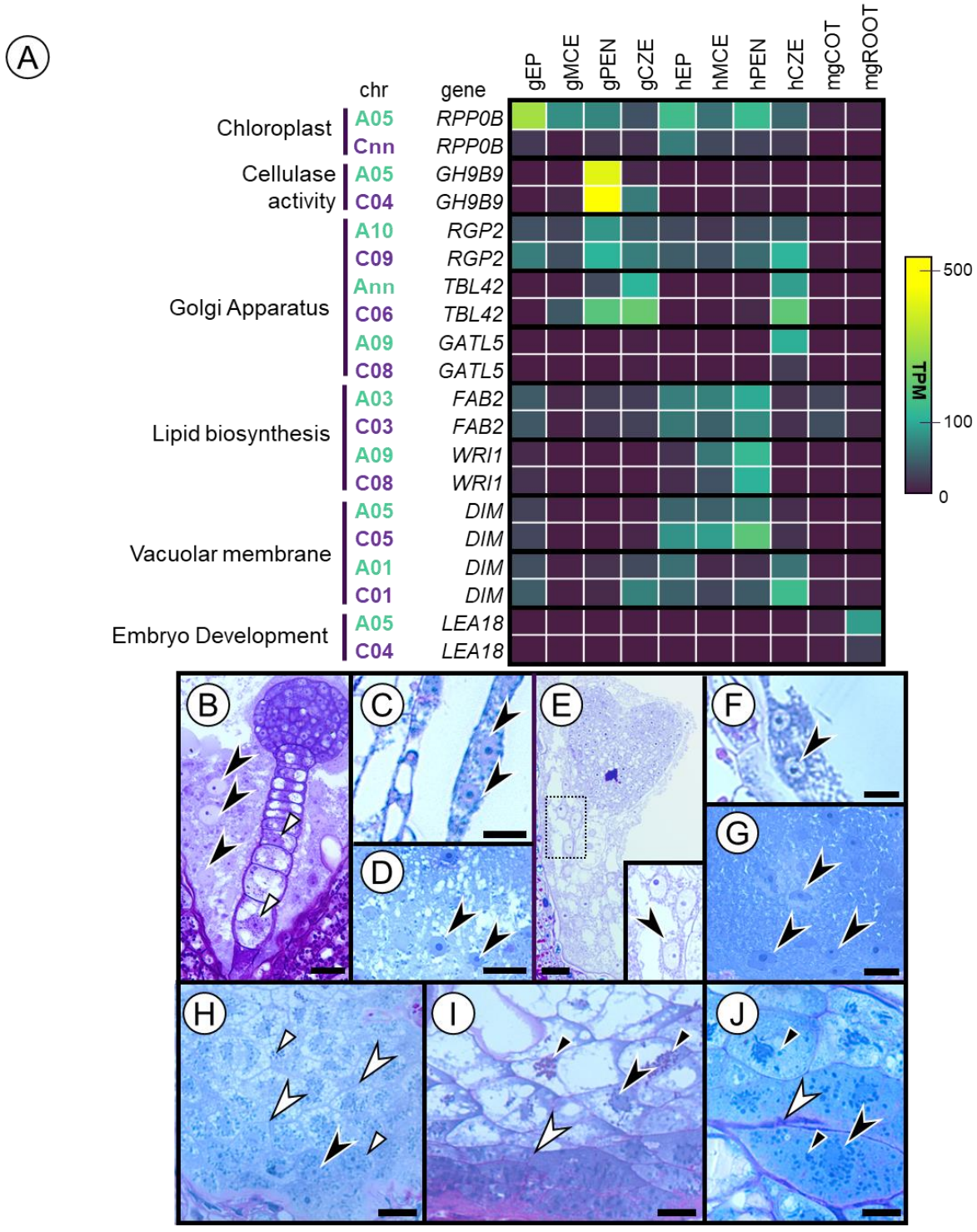


Figure 4.10. (A) Heat map of selected homeologues belonging to significantly enriched GO terms from the clusters in Figure 9. **(B)** 7 DPP seed globular EP atop the suspensor at the micropylar pole, with enlarged nuclei (black arrowheads) in the MCE syncytium and abundant plastids (white arrows) particularly in the perinuclear region. **(C)** the PEN at 7 DPP comprised of free nuclei in the syncytium, continuous with the **(D)** CZE. **(E)** 10 DPP heart EP with enlarged

nuclei in the MCE migrating closer together prior to cellularization. **(F)** PEN 10 DPP with nuclei remaining in a syncytium, as in the **(G)** CZE. **(H)** CZE of D14 seeds initiate cell wall formation (white arrowheads) which isolate the nuclei of the syncytium. Perinuclear protein-rich plastids (white arrows) populate the CZE. **(I)** CZE of D21 seeds establish complete cell walls, terminating the syncytium within the chalaza of the seed. Plastids distal to the CPT convert to starch storage (amyloplasts, black arrows). **(J)** D28 seeds divide the CZE with robust cell primary wall thickenings. Bar = **(B)** 40 μm **(C)** 20 μm **(D)** 40 μm **(E)** 100 μm **(F)** 30 μm **(G)** 30μm **(H)** 30 μm **(I)** 30 μm **(J)** 20 μm

Cellularization of the endosperm in *B. napus* occurred over several weeks after fertilization. At the globular stage of development, the MCE was entirely syncytial, with free nuclei throughout the shared cytoplasm (Figure 4.10B). Conspicuous nuclei were observed in the suspensor and was surrounded by perinuclear plastids. At the same developmental stage, the PEN and CZE at 7 DPP also had discrete nuclei floating in the syncytium (Figure 4.10 C-D). By 10 DPP, the nuclei enlarged and migrated closer together in preparation for cellularization (Figure 4.10E), while the PEN and CZE retained the continuous syncytium (Figure 4.10 F-G). By 14 DPP, cell plates were visible in the CZE with proteoplasts encircling the nuclei (Figure 4.10H). Proteoplasts in the CZE converted to amyloplasts at 21 DPP (Figure 4.10I), and by 28 DPP, thickened primary walls then surround the nuclei of the CZE, finalizing the cellularization of the CZE. Taken together, data analysis revealed subgenome bias to be minimal in the filial tissues compared to the maternal subregions

4.3.10 Paralogous genes involved in seed storage reservoirs exhibit divergent subgenome bias

As a consequence of the ancestral WGT, *B. napus* has many homologous gene pairs that are paralogous to each other. Oftentimes, *B. napus* can have up to six copies of the corresponding gene in *Arabidopsis*; three from the WGT, doubled from the allopolyploidization event. Seed storage proteins (SSPs) are instrumental in seed maturation and are of particular

interest in oilseed crops. We investigated two well known SSP gene families (*SEED STORAGE ALBUMIN* (*SESA*), *CRUCIFERIN* (*CRU*)) and the OLEOSIN (*OLEO*) gene family to determine if subgenome bias was prevalent across not only homologous pairs but also their paralogs. The oleosin gene family is well-documented to be implicated in lipid droplet synthesis in late seed development, and we report that all OLEO genes are highly expressed late in seed development in both maternal and filial subregions and especially within the mgCOT, mgCPT, and mgISC (Figure 4.11). All *OLEO* genes were more highly expressed in the mgCOT than in the mgROOT. Interestingly, a homologous gene pair of *OLEO1* was biased in A^n subgenome transcript activity across all subregions but the CZE in HEART seeds, and in all subregions in MG seeds. The inverse trend was observed in *SESA4*, which was biased towards the C^n subgenome in one homologous pair than the other two. The C^n biased *SESA4* pair was the least expressed pair of *SESA* homologs. Cruciferins are compact SSPs which localize to vacuoles within the embryo to nourish the plant during germination. While most *CRU* genes were highly expressed in all seed subregions during maturation, *CRU2* had a homologous pair with comparatively lower global expression, and another with conspicuous A^n subgenome bias. These data indicate that certain SSPs and oleosins have divergent expression patterns across homologs and paralogs.

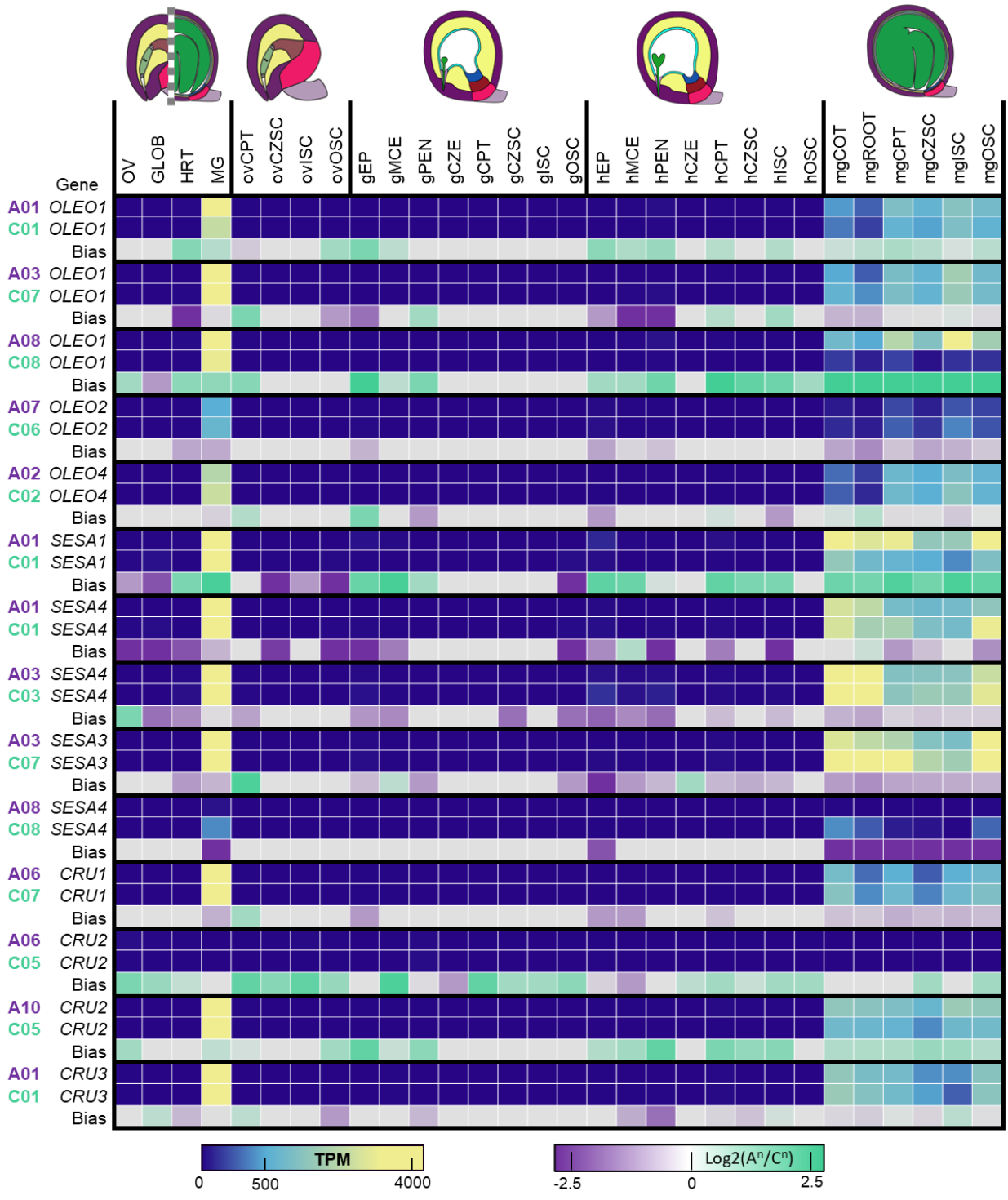


Figure 4.11. Transcript accumulation of *Brassica napus* homologs of known seed storage and seed oil biogenesis genes in *Arabidopsis*. Transcript accumulation represented in Transcripts per Million (TPM) and "bias" represented as ($Log_2(A^n/C^n) \geq 2$ ('A Bias') or $Log_2(A^n/C^n) \leq -2$ ('C Bias').

4.4 Discussion

4.4.1 Global transcriptome of the *Brassica napus* seed is biased towards the C^n subgenome in nearly all subregions

Dynamic changes in gene activity underpins development in *B. napus* and the recent polyploidization of this crop further complicates our understanding of the seed lifecycle. For example, tbe transcriptome of individual cells and tissues of the *B. napus* seed is biased towards the C^n subgenome. The C^n subgenome has ~26.1% more coding genes than the A^n subgenome while the most highly expressed genes were universally biased towards the C^n. These data suggest gene expression bias is unlikely to be solely due to subgenome size in recent polyploids like *B. napus* and that subgenome bias is likely to be more nuanced. Our previous work showed that transcription factors expressed during seed development were biased towards the A^n subgenome (Khan et al., 2022) while our current exploration of the seed revealed many gene homologs are in fact biased towards the A^n subgenome. While C^n subgenome bias pervades throughout development, it is important to consider subgenome bias through many genome contexts — the number of genes expressed from each subgenome, which subgenome generates the most transcripts, and bias between subgenome gene homologs all contribute to the biases whichmay contribute to the making of a seed.

Our previous work showed C^n subgenome bias of DNA methylation and siRNA accumulation in the *B. napus* seed (Ziegler et al., 2023). Resynthesis of *B. napus* through hybridization of *B. rapa* and *B. oleracea* experiments report C^n subgenome bias in vegetative tissues is consistent across independently resynthesized individuals (Bird et al., 2021). Despite these findings, other resynthesis experiments revealed A^n subgenome transcription fluctuates more heavily when compared with the *B. rapa* progenitor (Wu et al., 2018). These studies

suggest the C^n subgenome elicits genome dominance in both reproductive and vegetative tissues, but that the A^n subgenome is potentially more vulnerable to genomic shock following allopolyploidization. Here, we report that subgenomic bias is globally diminished as the seed matures, with the EP having notably fewer unique transcripts. Our previous work demonstrated that the epigenetic transition from morphogenesis to maturation was characterized by a global increase in DNA methylation in tandem with more siRNA loci being transcribed (Ziegler et al., 2023). These data, taken from the whole seed, implied that maturation was characterized by a global silencing effect on the whole genome. While this seems to be true of the whole seed data, when specific compartments of the seed are dissected and mRNAs profiled, the embryo has substantially fewer genes being transcribed, though with comparably similar transcript accumulation to that of other subregions of the seed. While recent research that profiled the transcriptomic bias of the embryo and seed coat reported similarly global C^n subgenome bias in *B. napus*, our work suggests that these biases are exacerbated when dissecting the maternal regions to cell and tissue-specific subregions (Gao et al., 2022). Taken together, the *B. napus* seed undergoes a dramatic shift in transcript activity across the seed lifecycle and the previously documented epigenetic silencing of the Cn subgenome is likely to have a more profound influence on the subgenomic bias of the filial subregions of the seed compared to those subregions of maternal origin.

4.4.2 The chalazal proliferating tissue of *Brassica* presents a unique opportunity to investigate the evolution of oilseeds

The nucellus is the megasporangial tissue of angiosperms and is responsible for the creation of the achesporial cell and megaspore mother cell (Lu and Magnani, 2018a). While the nucellus of gymnosperm seeds are resorbed into the female gametophyte as a nutrient source,

the rapid degeneration of the nucellus has been regarded as a key feature in the evolution of angiosperm seed partitioning (Lu and Magnani, 2018b). The endosperm and nucellus are documented to develop antagonistically to each other, with the nucellus proliferating during ovule development and then promptly resorbed by the proliferating endosperm post-fertilization (Xu et al., 2016). This nutrient transfer between seed subregions is thought to play an important role in the evolution of angiosperm seeds, and the CPT considered to be the remaining vector of nutrient transfer between the maternal and filial subregions in *B. napus* (Millar et al., 2015b; Ziegler et al., 2019).

The seeds of the Brassicaceae have CPTs that undergo varying degrees of degeneration, though few compare to the robust and persistent CPT of *Brassica.* In fact, the CPT of the model plant *Arabidopsis* (Camelineae) is diminutive and ephemeral making it difficult to study. This is true of many clades within the Brassicaceae, including *Lepidium* (Lepidieae), *Capsella* (Camelineae), and *Thlaspi* (Thlaspideae) (Nguyen et al., 2000b; Brown et al., 2004b). *Brassica* (Brassiceae), *Erysimum* (Erysimeae), *Stanleya* (Thelypodieae), and *Descurainia* (Descurainieae) in contrast have a discrete CPT which persists throughout late seed development. Few studies have interrogated the anatomy of the subregions in these groups, and details of gene activity underpinning its development are scant (Brown et al., 2004b; Ziegler et al., 2019). Interestingly, these groups span the entire crucifer family, so it seems the CPT evolved to be persistent or degenerate early many times throughout history. In *B. napus*, the CPT is derived from the chalazal nucellus and adopts new function as the seed matures. It has been established that the maternal-filial interfaces of the seed acts as a central integration hub of developmental signalling and is required for the coordination between the maternal and filial components of the seed (Povilus and Gehring, 2022). The interface of the CPT and CZE in *B. napus* is not

segregated by a cuticle, unlike the junctions between the ISC and the PEN/MCE. Thus, cross-signalling is likely to occur through the chalazal pole, with the CPT acting as the conduit connecting the CZSC and CZE. The CPT then must adopt transcriptomic profiles to match the needs of the coordination between the maternal and filial subregions. Our results show that subgenome bias of homologous gene pairs of the ovCPT are dissimilar to other maternal subregions. The CPT is anatomically one of the most variable subregions of Brassicaceae seeds, and this variation is largely due to the persistence or degeneration of the nucellus, wherein persistent chalazal nucellar tissue differentiates to become distinct, or is entirely lost due to cellular elimination (Ingram and Penfield, 2017). Given our data show that homologous gene pairs in the ovCPT are transcribed very differently than they are in other subregions, it is possible polyploidization affects the nucellar development of crucifer seeds. Further, the Brassicaceae are preceded by a whole genome duplication dating back 24-40 mya. The surviving cells of the CPT in *B. napus* MG seeds are anatomically incorporated into the CZSC, furthering the CPT as a dynamic tissue layer that transitions between ovule, morphogenesis, and maturation phases. Further research documenting subgenome bias in clades with poorly developed CPTs (eg. Lepidieae, Cameliniae, Thlaspideae) in contrast to those of robust CPTs (eg. other Brassiceae species, Descurainieae, Erysmieae, Thelypodieae) may be prudent to understand the relationship between crucifer polyploidization and CPT development. As the robust CPT in *Brassica* likely imparts the ability to accrue substantial nutrient reservoirs, this research could lead to important discoveries in oilseed evolution

Seed coat subregions experience subgenome bias distinct from one another

The integuments serve to protect the developing nucellus by proliferating and elongating to encircle it. The ancestral origin of integuments in seed plants is still a debated

topic, though much of the scientific community agrees that the integuments arose as a form of either dichotomous branches or as lateral organs from the nucellus (Mathews & Kramer, 2012; Meade et al., 2021; Smith, 1964). Despite their ancestral similarity, these structures have become anatomically and transcriptomically divergent with evolutionary time. Our data reveal that the ISC and OSC of *B. napus* exhibit different subgenomic biases across homologous genes, particularly during seed morphogenesis. Our anatomical data show that the ISC accumulates protein and oil bodies accompanied by thickened primary walls during maturation, whereas the OSC consumes much of its starch reservoirs, develops the palisade layer fully with reinforced pectin-rich walls, and deposits mucilage in the dermal cells. While much of these processes are known to occur in *Arabidopsis* (Beeckman et al., 2000; Tsai et al., 2017) our data indicate that these developmental transitions are accompanied by dissimilar subgenomic balance among the two seed coat layers during early seed development. Morphogenesis is marked by an increase in pectin accumulation of the OSC and starch granule acquisition in both layers, but these processes at least in part seem to be asymmetrically controlled by the different subgenomes. The specialization of each seed coat layer is an important facet of seed development (Khan et al., 2014), and the prevalence of subgenome bias demonstrates that the biochemical programs required for these processes are subject to alterations due to polyploidization. In fact, resynthesis experiments of hexaploid *Brassica* have demonstrated that machinery relating to seed coat colour undergo dramatic transcriptomic shifts post-polyploidization (Liu et al., 2020). As seed coat anatomy and structure are critical for the protection and nourishment of the embryo, it is therefore possible that resynthesis experiments could be used to improve seed coat quality or durability for food crops.

4.4.3 Filial subregions experience less subgenomic bias than maternal subregions

B. napus subgenome bias is well documented to exist in many different tissues, including both siliques and whole seeds (Li et al., 2020; Khan et al., 2022; de Jong and Adams, 2023). While we report conspicuous subgenomic bias characterizing the maternal subregions, the filial subregions are not subject to the same subgenomic asymmetry. At a global transcriptome level, large sets of gene pairs were of similar expression across subgenomes, particularly in the mgROOT and mgCOT. It is possible that upon establishment of the new sporophytic generation, the asymmetric silencing of one subgenome to the other is briefly alleviated. The mgROOT/mgCOT both accumulate similar amounts of total transcripts as other subregions but markedly less genes being transcribed altogether. This phenomenon may be due to mass erasure of gene activity in favour of a smaller subset of genes required for maintenance of the matured embryo. In particular, data reveal that some of the most highly expressed genes in the mgROOT/COT encode SSPs and oleosin family proteins at several orders of magnitude higher expression than the vast majority of genes, and in the case of *SESA*, expression as high as 23S ribosomal chloroplast genes. This dramatic reduction in transcript activity is not reflected in mature *Arabidopsis* embryos (Belmonte et al., 2013b). It is likely that the majority of all gene activity in the *B. napus* mgEP is dedicated to seed storage, and the many impertinent genes to the process are reduced to subtle noise at the global transcriptomic level. As a result, the influence of polyploidy in the mature embryo is unlikely to be influential outside of seed storage.

4.4.4 Seed storage reservoirs have divergent expression patterns within homologous gene pairs

Seed maturation is characterized by the acquisition of storage reserves in preparation for dormancy and to nourish the embryo upon germination. This process is essential to the longevity of the seed and the future seedling, and many genes have been identified which enable storage of energy-dense proteins or lipids (Nguyen et al., 2015; Niñoles et al., 2022). As an oilseed, these processes are the crux of *B. napus* seed development. Oleosins are required for the maintenance of lipid droplets, and seed storage albumins and cruciferins are known to be essential seed storage proteins that generate protein bodies within the embryo (Wan et al., 2007; Miquel et al., 2014; Huang, 2018; Zheng et al., 2022; Traver and Bartel, 2023). Despite the high transcript abundance in mature embryos, these genes experience exponential differences in expression across subgenomes. For example, one homolog pair of OLEO1 is expressed a minimum of 6.5x more in the A^n subgenome in the mgROOT, and a maximum of ~39x more highly transcribed from the A^n subgenome in the mgCZSC. Further, these differences occur within homologous gene pairs, including *OLEO1*, *SESA1*, *SESA4*, and *CRU2*. Despite the many copies *B. napus* has of these genes, only certain homologous pairs experience this substantial bias between the two subgenomes. This implies that seed storage genes experience subgenome dominance in certain homologs, but this bias does not occur evenly across all paralogous genes. Future works investigating these gene families should consider examining the gene homolog which exhibits this exponential bias, either to explore the storage reserves in this oilseed or as a model for understanding genome dominance in integral developmental processes.

4.4.5 Conclusions

The influence of polyploidy on plant development, gene expression, and consequent evolution is a well studied topic but one of near endless complexity. Modern research attempts to describe transcriptomic and epigenetic subgenome bias or genome remodelling and fractionation in response to polyploidization (Yuan et al., 2020; Zhang et al., 2022; and reviewed in Cheng et al., 2018). These works have advanced our understanding of the consequences of polyploidy and have allowed the scientific community at large to better understand the genome dynamics of polyploids. However, much of the work done has focused on either vegetative tissues, or whole structures rather than specific cells and tissues found within reproductive organs. Our work provides an analytical foundation for studying polyploid dynamics at a subregion-level in the seed.

Our findings suggest polyploidization affects seed subregions differently, with maternal subregions eliciting a more varied palette of subgenome bias than the comparatively stable filial subregions. Further, we demonstrate that the CPT should be of particular interest to crop development in oilseeds and could be susceptible to developmental changes as a result of polyploidy. Cells and tissues of the seed that are of maternal origin would benefit from more investigation through the lens of polyploidy, both within established oilseeds and their close relatives or progenitors. Together, we find that individual cells, tissues, and organs across the seed lifecycle are influenced by polyploidization in ways distinct from each other yet cooperate in the making of a seed.

4.5 Methods

4.5.1 Plant growth conditions and collection

Brassica napus cv. Topas plants were grown in growth chambers under long day conditions (16 h light, 8 h dark, 22°C, RH 50%–70%) and flowers tagged and hand-pollinated (Chan & Belmonte, 2013). Seeds were collected at 0 (OV), 7 (GLOB), 10 (HEART) and 28 (MG) DAP for laser microdissection (LMD), and 0, 7, 10, 14, 21, and 28 DAP for histological analysis. Flower pollination was performed shortly after anthesis, when the anthers of the flower barely exceeding the corolla. Flower and silique harvest were performed between 15:00 and 17:00 to minimize time of day effects on gene activity.

4.5.2 Laser Microdissection

We used laser microdissection (LMD) to capture individual subregions of the seed. All tissue processing and collection was performed under RNase-free conditions. Siliques of OV, GLOB, HEART, and MG stage seeds were cut at 50 mm fragments and fixed in 1:3 (v/v) glacial acetic acid: 85% ethanol for 24 h at 4°C. Tissues were then dehydrated in a graded ethanol series, followed by a transition infiltration with xylenes at 4°C and gradually infiltrated with paraffin at 60°C. Samples were incubated for less than 8 h at 60°C prior to embedding in 100% paraffin wax (McCormick Scientific, Paraplast®Plus™). Serial sections (7 µm sections, Leica RM2245 rotary microtome) were mounted on polyethylene napthalate (PEN) membrane slides (Leica Microsystems). Slides were deparaffinized in xylenes for 1 min and seed subregions were isolated using the Leica Laser Microdissection 7000 system.

4.5.3 RNA isolation

Total RNA was extracted from the laser-microdissected subregions using the Ambion® RNaqueous micro kit and treated with Qiagen RNAase-free DNase to remove genomic DNA. RNA quantity was evaluated using the Quant-iT™ RiboGreen™ RNA Assay kit (Invitrogen™) and quality was evaluated using the Agilent 6000 Pico LabChip® and the Agilent 2100 bioanalyzer (Agilent Technologies, USA). A minimum RIN of 3 was required for further cDNA synthesis and library preparation (Belmonte et al. 2013).

4.5.4 Library preparation and sequencing

First and second cDNA synthesis was performed according to a high-throughput RNA-seq protocol (Kumar et al. 2012). Fragmentation was then carried out using the Covaris™ M220 Focused-ultrasonicator™ in Covaris™ M220 microTUBEs. Samples were then brought down to 5 µL via rotary evaporation. Library preparation was performed using the NuGen® Ovation® Ultralow DR Multiplex System. Total library quality and quantity were determined using the Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen), qPCR, and the High Sensitivity DNA Analysis Kit (Agilent) on the Agilent 2100 Bioanalyzer. Libraries were pooled and size-selected (350–600 bp) on the E-Gel® system (Invitrogen™). Samples were then sequenced on the Illumina® HiSeq™ 2000 platform at Genome Québec. Whole seed RNA was extracted from whole globular stage seeds using PureLink® Plant RNA Reagent (Ambion®). Library preparation was performed as in (Kumar et al. 2012) using NEXTFlex ChIP-seq barcodes (Bio Scientific®).

4.5.5 RNA-seq analysis

Low-quality reads and adaptors were removed from the RNA-seq dataset with fastp (Chen et al., 2018). Surviving reads were then aligned to a *Brassica napus* cv. Topas reference

genome with hisat2 (Kim et al., 2019). Quantification of uniquely mapped reads was performed using featurecounts, native to the subread package (Liao et al., 2014). Counts were then normalized to transcripts per million (TPM). Data clustering and heatmap visuals were completed using heatmap(r), and Euclidean distance was used to compute distance between nodes during clustering. Gene homologs were identified as annotated within (Gao et al., 2022).

4.5.6 Light microscopy analysis

Samples were fixed in 1.6% Glutaraldehyde 4% Paraformaldehyde solution overnight, then transferred to methylcellosolve for 6-18 hours (older seeds require longer time and more thorough dehydration). Samples were then transferred into ethanol three consecutive times, with each wash lasting 24 hours. Tissue was then infiltrated with activated Leica HistoResin over three days (1/3 ethanol to HistoResin, followed by 2/3 and then 100% HistoResin). Samples were then embedded onto pucks and sectioned at 2.5 µm thickness using a Leica RM2245 rotary microtome with a steel blade and mounted on glass slides. All slides were stained with periodic acid-Schiff's reagent, then followed with either toluidine blue O or amido black. Slides were sealed with Cytoseal™ 60 (Richard-Allan Scientific™). Leica DM2500 light microscope equipped with a Leica DFC425 camera using LAS3.7 software was used to visualize the sections.Transmission electron microscopy analysis

4.5.7 Hand Sectioning

Lipid staining was performed on live tissue sectioned by hand using double edge razors. GLOB, HEART, and MG seeds were sectioned longitudinally and stained with Sudan IV for at least 25 minutes. Sections were mounted in glycerol to visualize.

4.5.8 Transmission Electron Microscopy

Transmission electron microscopy was performed as described within Chan & Belmonte, 2013 and Millar et al., 2015). Ovules and seeds were fixed overnight in 3% glutaraldehyde in 0.025 M cacodylate buffer supplemented with 5 mM calcium chloride (pH 7.0). Plant material was rinsed with cacodylate buffer and post-fixed with 2% osmium tetroxide in 0.8% KFe(CN)6 in cacodylate buffer. After post-fixation, seeds were rinsed with distilled water and stained overnight with a 0.5% aqueous uranyl acetate solution to increase contrast.

Processed seeds were rinsed in distilled water and dehydrated in a graded ethanol series. Seeds were then further dehydrated in 1:1 absolute ethanol to propylene oxide (v:v) and then in 100% propylene oxide. Finally, they were gradually infiltrated and embedded in Spurr's epoxy resin at 70 °C [28]. All methods prior to embedding were performed at 4 °C. Using a Reichert–Jung Ultracut ultramicrotome, sections were cut (90 nm thickness) with a Diatome diamond knife and mounted on copper grids. The sections were visualized with a Hitachi H-7000 transmission electron microscope at 75 kV and pictures were taken using AMT Image Capture Engine version 601.384.

4.6 Supplemental Figures

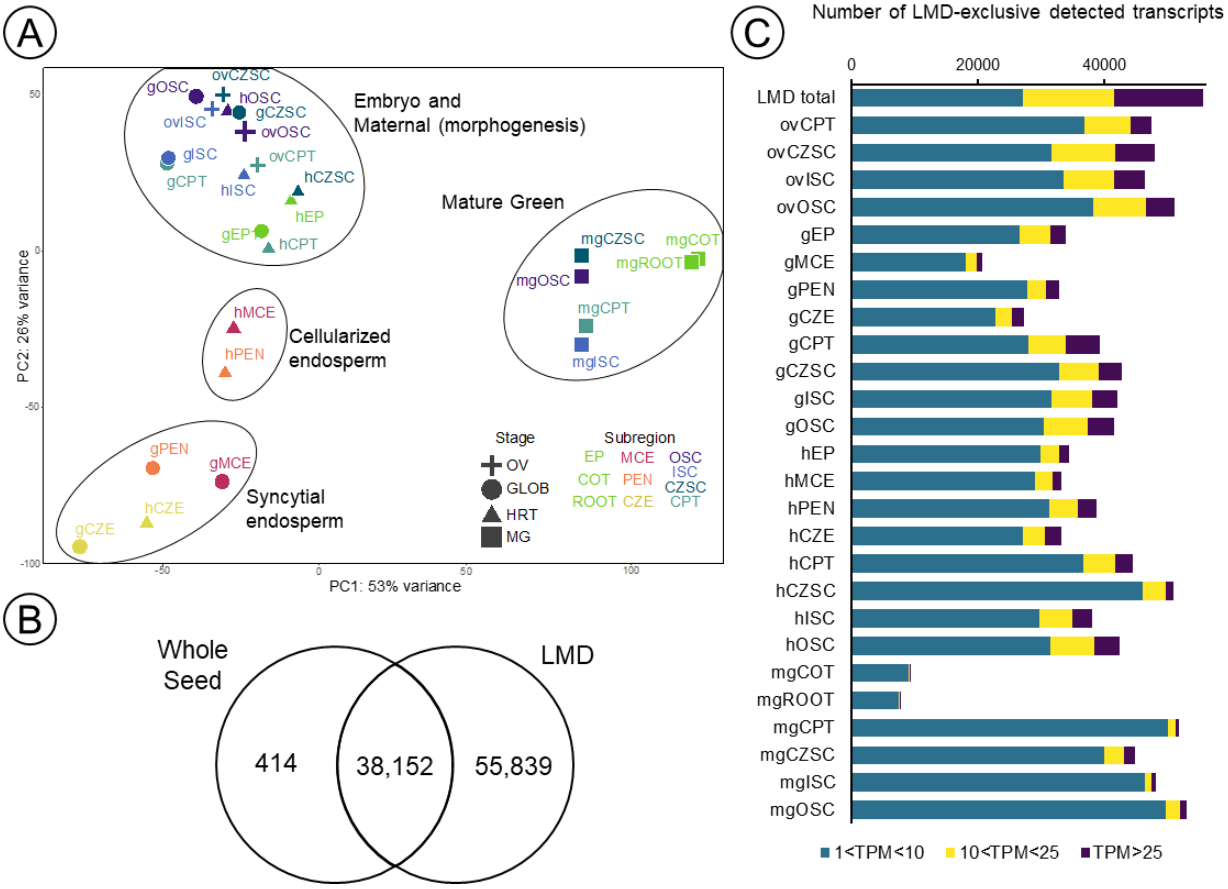


Figure S1. (A) Principle Component Analysis (PCA) comparing the transcriptome profiles of the *Brassica napus* seed subregions. Each stage is marked with a unique identifier (OV = plus sign, GLOB = circle, HRT = triangle, MG = square). (B) Transcripts detected (TPM>1) in whole seed and laser-microdissected seed transcriptomes over the OV, GLOB, HRT, and MG stages. Increasing spatial resolution yields 44% greater transcript detection in the developing *B. napus* seed. (C) Transcripts detected only by LMD-RNAseq in the developing *B. napus* seed. Transcripts are sorted into low (1<TPM≤10, teal), medium (10<TPM≤25, yellow), and high (TPM>25,) accumulation levels.

4.7 Contribution of Authors

D.J.Z, D.K, J.M, and M.G.B performed experiments. D.J.Z and D.K analyzed the data. D.J.Z and

M.F.B wrote and reviewed the manuscript.

4.8 References

Alptekin B, Akpinar BA, Budak H (2017) A comprehensive prescription for plant miRNA identification. Front Plant Sci **7**: 2058

Amich F, García-Barriuso M, Bernardos S (2007) Polyploidy and speciation in the orchid flora of the Iberian Peninsula. Botanica Helvetica **117**: 143–157

An J, Lai J, Sajjanhar A, Lehman ML, Nelson CC (2014) MiRPlant: An integrated tool for identification of plant miRNA from RNA sequencing data. BMC Bioinformatics **15**: 275

Ando A, Kirkbride RC, Jones DC, Grimwood J, Chen ZJ (2021) LCM and RNA-seq analyses revealed roles of cell cycle and translational regulation and homoeolog expression bias in cotton fiber cell initiation. BMC Genomics **22**: 1–16

Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA **19**: 740–751

Axtell MJ, Meyers BC (2018) Revisiting criteria for plant microRNA annotation in the Era of big data. Plant Cell **30**: 272–284

Axtell MJ, Westholm JO, Lai EC (2011) Vive la différence: Biogenesis and evolution of microRNAs in plants and animals. Genome Biol **12**: 1–13

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V (2019) Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun. doi: 10.1038/S41467-019-13730-0

Batista RA, Figueiredo DD, Santos-González J, Köhler C (2019a) Auxin regulates endosperm cellularization in Arabidopsis. Genes Dev **33**: 466–476

Batista RA, Figueiredo DD, Santos-González J, Köhler C (2019b) Auxin regulates endosperm cellularization in Arabidopsis. Genes Dev **33**: 466–476

Bedi S, Sengupta S, Ray A, Nag Chaudhuri R (2016) ABI3 mediates dehydration stress recovery response in Arabidopsis thaliana by regulating expression of downstream genes. Plant Science **250**: 125–140

Beeckman T, De Rycke R, Viane R, Inzé D (2000) Histological study of seed coat development in Arabidopsis thaliana. J Plant Res **113**: 139–148

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, et al (2013a) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci U S A **110**: E435–E444

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, et al (2013b) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci U S A **110**: E435–E444

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015) The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis **53**: 474–485

Bevilacqua C, Makhzami S, Helbling JC, Defrenaix P, Martin P (2010) Maintaining RNA integrity in a homogeneous population of mammary epithelial cells isolated by Laser Capture Microdissection. BMC Cell Biol **11**: 1–13

Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ (2017) The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol **18**: 65

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP (2021) Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. New Phytologist **230**: 354–371

Bleckmann A, Alter S, Dresselhaus T (2014) The beginning of a seed: Regulatory mechanisms of double fertilization. Front Plant Sci. doi: 10.3389/FPLS.2014.00452/ABSTRACT

Bologna NG, Voinnet O (2014) The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis . Annu Rev Plant Biol **65**: 473–503

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nature Reviews Molecular Cell Biology 2015 16:12 **16**: 727–741

Boscá S, Knauer S, Laux T (2011a) Embryonic development in arabidopsis thaliana: From the zygote division to the shoot meristem. Front Plant Sci. doi: 10.3389/fpls.2011.00093

Boscá S, Knauer S, Laux T (2011b) Embryonic development in arabidopsis thaliana: From the zygote division to the shoot meristem. Front Plant Sci **2**: 16239

Bouyer D, Kramdi A, Kassam M, Heese M, Schnittger A, Roudier F, Colot V (2017) DNA methylation dynamics during early plant life. Genome Biol. doi: 10.1186/s13059-017-1313-0

Brown RC, Lemmon BE, Nguyen H (2004a) Comparative anatomy of the chalazal endosperm cyst in seeds of the Brassicaceae. Botanical Journal of the Linnean Society **144**: 375–394

Brown RC, Lemmon BE, Nguyen H (2004b) Comparative anatomy of the chalazal endosperm cyst in seeds of the Brassicaceae. Botanical Journal of the Linnean Society **144**: 375–394

Brown RC, Lemmon BE, Nguyen H, Olsen OA (1999) Development of endosperm in Arabidopsis thaliana. Sex Plant Reprod **12**: 32–42

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014a) Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. Science (1979) **345**: 950–953

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014b) Early allopolyploid evolution in the post-neolithic Brassica napus oilseed genome. Science (1979) **345**: 950–953

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al (2014c) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345**: 950–953

Che NY, Yang Y, Li YD, Wang LL, Huang P, Gao Y, An CC (2009) Efficient LEC2 activation of OLEOSIN expression requires two neighboring RY elements on its promoter. Sci China C Life Sci **52**: 854–863

Chen C, Zeng Z, Liu Z, Xia R (2018) Small RNAs, emerging regulators critical for the development of horticultural traits. Hortic Res. doi: 10.1038/s41438-018-0072-8

Chen D, Yuan C, Zhang J, Zhang Z, Bai L, Meng Y, Chen L-L, Chen M (2012) PlantNATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acid Res **40**: 1187–1193

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X (2018) Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants 2018 4:5 **4**: 258–268

Clément-Ziza M, Munnich A, Lyonnet S, Jaubert F, Besmond C (2008) Stabilization of RNA during laser capture microdissection by performing experiments under argon atmosphere or using ethanol as a solvent in staining solutions. RNA **14**: 2698

Coen O, Fiume E, Xu W, De Vos D, Lu J, Pechoux C, Lepiniec L, Magnani E (2017) Developmental patterning of the sub-epidermal integument cell layer in Arabidopsis seeds. Development (Cambridge) **144**: 1490–1497

Conde D, Triozzi PM, Balmant KM, Doty AL, Miranda M, Boullosa A, Schmidt HW, Pereira WJ, Dervinis C, Kirst M (2021) A robust method of nuclei isolation for single-cell RNA sequencing of solid tissues from the plant genus Populus. PLoS One **16**: e0251149

Contreras B, Vives C, Castells R, Casacuberta JM (2015) The impact of transposable elements in the evolution of plant genomes: From selfish elements to key players. Evolutionary Biology: Biodiversification from Genotype to Phenotype. Springer International Publishing, pp 93–105

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res **39**: 155–159

Danzer J, Meilott E, Bui AQ, Le BH, Martin P, Hashimoto M, Perez-Lesher J, Chen M, Pelletier JM, Somers DA, et al (2015) Down-Regulating the Expression of 53 Soybean

Transcription Factor Genes Uncovers a Role for SPEECHLESS in Initiating Stomatal Cell Lineages during Embryo Development. Plant Physiol **168**: 1025

Delmas F, Sankaranarayanan S, Deb S, Widdup E, Bournonville C, Bollier N, Northey JGB, McCourt P, Samuel MA (2013) ABI3 controls embryo degreening through Mendel's i locus. Proc Natl Acad Sci U S A **110**: E3888

DeMarini DM, Mutlu E, Warren SH, King C, Gilmour MI, Linak WP (2019) Mutagenicity Emission Factors of Canola Oil and Waste Vegetable Oil Biodiesel: Comparison to Soy Biodiesel. Mutat Res **846**: 403057

Doll NM, Royek S, Fujita S, Okuda S, Chamot S, Stintzi A, Widiez T, Hothorn M, Schaller A, Geldner N, et al (2020) A two-way molecular dialogue between embryo and endosperm is required for seed development. Science **367**: 431–435

Dorrity MW, Alexandre CM, Hamm MO, Vigil AL, Fields S, Queitsch C, Cuperus JT (2021) The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nature Communications 2021 12:1 **12**: 1–12

Errampalli D, Patton D, Castle L, Mickelson L, Hansen K, Schnall J, Feldmann K, Meinke D (1991) Embryonic Lethals and T-DNA lnsertional Mutagenesis in Ara bidopsis.

Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C (2016) Auxin production in the endosperm drives seed coat development in Arabidopsis. Elife. doi: 10.7554/ELIFE.20542

Figueiredo DD, Batista RA, Roszak PJ, Köhler C (2015) Auxin production couples endosperm development to fertilization. Nature Plants 2015 1:12 **1**: 1–6

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011a) Cabbage family affairs: The evolutionary history of Brassicaceae. Trends Plant Sci **16**: 108–116

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011b) Cabbage family affairs: The evolutionary history of Brassicaceae. Trends Plant Sci **16**: 108–116

Freeling M, Xu J, Woodhouse M, Lisch D (2015) A solution to the c-value paradox and the function of junk DNA: The genome balance hypothesis. Mol Plant **8**: 899–910

Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N (2008) Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol **26**: 407–415

Gao P, Quilichini TD, Yang H, Li Q, Nilsen KT, Qin L, Babic V, Liu L, Cram D, Pasha A, et al (2022) Evolutionary divergence in embryo and seed coat development of U's Triangle Brassica species illustrated by a spatiotemporal transcriptome atlas. New Phytologist **233**: 30–51

Gardiner LJ, Quinton-Tulloch M, Olohan L, Price J, Hall N, Hall A (2015) A genome-wide survey of DNA methylation in hexaploid wheat. Genome Biol **16**: 273

Gómez LD, Baud S, Gilday A, Li Y, Graham IA (2006) Delayed embryo development in the ARABIDOPSIS TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. Plant Journal **46**: 69–84

Griffiths-Jones S (2006) miRBase: the microRNA sequence database. Methods Mol Biol **342**: 129–138

Guo Z, Kuang Z, Wang Y, Zhao Y, Tao Y, Cheng C, Yang J, Lu X, Hao C, Wang T, et al (2020) PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Res **48**: D1114–D1121

Harada JJ (2001) Role of arabidopsis LEAFY COTYLEDON genes in seed development. J Plant Physiol **158**: 405–409

Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM, Vaughan RM, Li X, Chen W, Xue Y, Zhong Z, et al (2018) A DNA methylation reader complex that enhances gene transcription. Science (1979) **362**: 1182–1186

Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J, et al (2017) Transcriptomic analysis of seed coats in yellow-seeded Brassica napus reveals novel genes that influence proanthocyanidin biosynthesis. Front Plant Sci **8**: 286392

Huang AHC (2018) Plant Lipid Droplets and Their Associated Proteins: Potential for Rapid Advances. Plant Physiol **176**: 1894

Huang C-H, Zhang C, Liu M, Hu Y, Gao T, Qi J, Ma H (2016) Multiple Polyploidization Events across Asteraceae with Two Nested Events in the Early History Revealed by Nuclear Phylogenomics. Mol Biol Evol **33**: 2820–2835

Ingram GC, Penfield S (2017) Dying to live: cell elimination as a developmental strategy in angiosperm seeds. J Exp Bot **68**: 785–796

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463–467

Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperusa JT (2019) Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. Plant Cell **31**: 993–1011

Jo L, Pelletier JM, Harada JJ (2019) Central role of the LEAFY COTYLEDON1 transcription factor in seed development. J Integr Plant Biol **61**: 564–580

de Jong GW, Adams KL (2023) Subgenome-dominant expression and alternative splicing in response to Sclerotinia infection in polyploid Brassica napus and progenitors. Plant J **114**: 142–158

Kawakatsu T, Nery JR, Castanon R, Ecker JR (2017) Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol **18**: 1–12

Khan D, Millar JL, Girard IJ, Belmonte MF (2014) Transcriptional circuitry underlying seed coat development in Arabidopsis. Plant Science **219–220**: 51–60

Khan D, Ziegler DJ, Kalichuk JL, Hoi V, Huynh N, Hajihassani A, Parkin IAP, Robinson SJ, Belmonte MF (2022) Gene expression profiling reveals transcription factor networks and subgenome bias during Brassica napus seed development. The Plant Journal **109**: 477–489

Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res **47**: D155–D162

Krasileva K V. (2019) The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. Curr Opin Plant Biol **48**: 18–25

Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. Development **130**: 6065–6073

Kuang Z, Wang Y, Li L, Yang X (2019) miRDeep-P2: accurate and fast analysis of the microRNA transcriptome in plants. Bioinformatics **35**: 2521–2522

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS (2018) Impact of whole-genome duplication events on diversification rates in angiosperms. Am J Bot **105**: 348–363

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75**: 843–854

Li M, Wang R, Wu X, Wang J (2020) Homoeolog expression bias and expression level dominance (ELD) in four tissues of natural allotetraploid Brassica napus. BMC Genomics **21**: 1–15

Li Q, Qiao X, Yin H, Zhou Y, Dong H, Qi K, Li L, Zhang S (2019) Unbiased subgenome evolution following a recent whole-genome duplication in pear (Pyrus bretschneideri Rehd.). Hortic Res **6**: 1–12

Liao P, Woodfield HK, Harwood JL, Chye ML, Scofield S (2019) Comparative Transcriptomics Analysis of Brassica napus L. during Seed Maturation Reveals Dynamic Changes in Gene Expression between Embryos and Seed Coats and Distinct Expression Profiles of Acyl-CoA-Binding Proteins for Lipid Accumulation. Plant Cell Physiol **60**: 2812–2825

de Lima JC, Loss-Morais G, Margis R (2012) MicroRNAs play critical roles during plant development and in response to abiotic stresses. Genet Mol Biol **35**: 1069–1077

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al (2014a) The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nature Communications 2014 5:1 **5**: 1–11

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al (2014b) The brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun **5**: 1–11

Liu Y, Wang J, Ge W, Wang Z, Li Y, Yang N, Sun S, Zhang L, Wang X (2017) Two highly similar poplar paleo-subgenomes suggest an autotetraploid ancestor of salicaceae plants. Front Plant Sci. doi: 10.3389/fpls.2017.00571

Liu Z, Wang R, Wang J (2020) Comprehensive Transcriptomic Analysis for Developing Seeds of a Synthetic Brassica Hexaploid. Plants **9**: 1–17

Lotan T, Ohto MA, Matsudaira Yee K, West MAL, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93**: 1195–1205

Lu J, Magnani E (2018a) Seed tissue and nutrient partitioning, a case for the nucellus. Plant Reprod **31**: 309–317

Lu J, Magnani E (2018b) Seed tissue and nutrient partitioning, a case for the nucellus. Plant Reprod **31**: 309–317

Lu L, Yang H, Xu Y, Zhang L, Wu J, Yi H (2023) Laser capture microdissection-based spatiotemporal transcriptomes uncover regulatory networks during seed abortion in seedless Ponkan (Citrus reticulata). The Plant Journal **115**: 642–661

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. Genome Res **15**: 516–525

Maggioni L, von Bothmer R, Poulsen G, Lipman E (2018) Domestication, diversity and use of Brassica oleracea L., based on ancient Greek and Latin texts. Genet Resour Crop Evol **65**: 137–159

Mansfield S, Briarty L (1990) Endosperm cellularization in Arabidopsis thaliana L. Arab Inf Serv **27**: 65–72

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264**: 421–424

Mathews S, Kramer EM (2012) The evolution of reproductive structures in seed plants: a re-examination based on insights from developmental genetics. New Phytologist **194**: 910–923

McFaline-Figueroa JL, Trapnell C, Cuperus JT (2020) The promise of single-cell genomics in plants. Curr Opin Plant Biol **54**: 114–121

Meade LE, Plackett ARG, Hilton J (2021) Reconstructing development of the earliest seed integuments raises a new hypothesis for the evolution of ancestral seed-bearing structures. New Phytologist **229**: 1782–1794

Meinke DW (1992) A homoeotic mutant of Arabidopsis thaliana with leafy cotyledons. Science (1979) **258**: 1647–1650

Millar AA, Waterhouse PM (2005) Plant and animal microRNAs: Similarities and differences. Funct Integr Genomics **5**: 129–135

Millar JL, Khan D, Becker MG, Chan A, Dufresne A, Sumner M, Belmonte MF (2015a) Chalazal seed coat development in Brassica napus. Plant Science **241**: 45–54

Millar JL, Khan D, Becker MG, Chan A, Dufresne A, Sumner M, Belmonte MF (2015b) Chalazal seed coat development in Brassica napus. Plant Science **241**: 45–54

Miquel M, Trigui G, D'Andréa S, Kelemen Z, Baud S, Berger A, Deruyffelaere C, Trubuil A, Lepiniec L, Dubreucq B (2014) Specialization of Oleosins in Oil Body Dynamics during Seed Development in Arabidopsis Seeds. Plant Physiol **164**: 1866

Mu J, Tan H, Zheng Q, Fu Y, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, et al (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol **148**: 1042–1054

Nag A, Jack T (2010) Sculpting the flower; the role of microRNAs in flower development. Curr Top Dev Biol. Academic Press Inc., pp 349–378

Nambara E, Hayama R, Tsuchiya Y, Nishimura M, Kawaide H, Kamiya Y, Naito S (2000) The role of ABI3 and FUS3 loci in Arabidopsis thaliana on phase transition from late embryo development to germination. Dev Biol **220**: 412–423

Nambara E, Keith K, McCourt P, Naito S (1995) A regulatory role for the ABI3 gene in the establishment of embryo maturation in Arabidopsis thaliana. Development **121**: 629–636

Nelson SK, Ariizumi T, Steber CM (2017) Biology in the dry seed: Transcriptome changes associated with dry seed dormancy and dormancy loss in the arabidopsis GA-insensitive sleepy1-2 mutant. Front Plant Sci. doi: 10.3389/fpls.2017.02158

Nguyen H, Brown RC, Lemmon BE (2000a) The specialized chalazal endosperm in Arabidopsis thaliana and Lipidium virginicum (Brassicaceae). Protoplasma **212**: 99–110

Nguyen H, Brown RC, Lemmon BE (2000b) The specialized chalazal endosperm in Arabidopsis thaliana and Lipidium virginicum (Brassicaceae). Protoplasma **212**: 99–110

Nguyen TP, Cueff G, Hegedus DD, Rajjou L, Bentsink L (2015) A role for seed storage proteins in Arabidopsis seed longevity. J Exp Bot **66**: 6399–6413

Niñoles R, Ruiz-Pastor CM, Arjona-Mudarra P, Casañ J, Renard J, Bueso E, Mateos R, Serrano R, Gadea J (2022) Transcription Factor DOF4.1 Regulates Seed Longevity in Arabidopsis via Seed Permeability and Modulation of Seed Storage Protein Accumulation. Front Plant Sci **13**: 915184

Nodine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev **24**: 2678–2692

Ono S, Liu H, Tsuda K, Fukai E, Tanaka K, Sasaki T, Nonomura KI (2018) EAT1 transcription factor, a non-cell-autonomous regulator of pollen production, activates meiotic small RNA biogenesis in rice anther tapetum. PLoS Genet **14**: e1007238

Otto SP, Whitton J (2000a) Polyploid incidence and evolution. Annu Rev Genet **34**: 401–437

Otto SP, Whitton J (2000b) Polyploid Incidence and Evolution. Annu Rev Genet **34**: 401–437

Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol **15**: 1–18

Pei L, Zhang L, Li J, Shen C, Qiu P, Tu L, Zhang X, Wang M (2019) Tracing the origin and evolution history of methylation-related genes in plants. BMC Plant Biol **19**: 307

Polydore S, Axtell MJ (2018) Analysis of RDR1/RDR2/RDR6-independent small RNAs in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of short interfering RNA loci. Plant Journal **94**: 1051–1063

Polydore S, Lunardon A, Axtell MJ (2018) Several phased siRNA annotation methods can frequently misidentify 24 nucleotide siRNA-dominated *PHAS* loci. Plant Direct **2**: e00101

Povilus RA, Gehring M (2022) Maternal-filial transfer structures in endosperm: A nexus of nutritional dynamics and seed development. Curr Opin Plant Biol **65**: 102121

Qi X, An H, Hall TE, Di C, Blischak PD, McKibben MTW, Hao Y, Conant GC, Pires JC, Barker MS (2021) Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in Brassica rapa. New Phytologist **230**: 372–386

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biology 2019 20:1 **20**: 1–23

Radchuk V, Borisjuk L (2014) Physical, metabolic and developmental functions of the seed coat. Front Plant Sci **5**: 102496

Ray S, Golden T, Ray A (1996) Maternal effects of the short integument mutation on embryo development in Arabidopsis. Dev Biol **180**: 365–369

Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J, Ma hongma H (2018) Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. Mol Plant **11**: 414–428

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: Polyploidy and crop plants. Am J Bot **101**: 1711–1725

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 Biol **21**: 1–17

Samad AFA, Sajad M, Nazaruddin N, Fauzi IA, Murad AMA, Zainal Z, Ismail I (2017) MicroRNA and transcription factor: Key players in plant regulatory network. Front Plant Sci **8**: 565

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, et al (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635–641

Schauer SE, Jacobsen SE, Meinke DW, Ray A (2002) DICER-LIKE1: Blind men and elephants in Arabidopsis development. Trends Plant Sci **7**: 487–491

Schneitz K, Hülskamp M, Pruitt RE (1995) Wild-type ovule development in Arabidopsis thaliana: a light microscope study of cleared whole-mount tissue. The Plant Journal **7**: 731–749

Schulz P, Jensen WA (1971) Capsella embryogenesis: the chalazal proliferating tissue. J Cell Sci **8**: 201–27

Schwartz BW, Yeung EC, Meinke DW (1994) Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis. Development 120:

Shahan R, Hsu CW, Nolan TM, Cole BJ, Taylor IW, Greenstreet L, Zhang S, Afanassiev A, Vlot AHC, Schiebinger G, et al (2022) A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. Dev Cell **57**: 543-560.e9

Siddiqui ZH, Abbas ZK, Ansari MW, Khan MN (2019) The role of miRNA in somatic embryogenesis. Genomics **111**: 1026–1033

Simon MK, Skinner DJ, Gallagher TL, Gasser CS (2017) Integument development in Arabidopsis depends on interaction of YABBY protein INNER NO OUTER with coactivators and corepressors. Genetics **207**: 1489–1500

Skinner DJ, Brown RH, Kuzoff RK, Gasser CS (2016) Conservation of the role of INNER NO OUTER in development of unitegmic ovules of the Solanaceae despite a divergence in protein function. BMC Plant Biol **16**: 143

SMITH DL (1964) THE EVOLUTION OF THE OVULE. Biological Reviews **39**: 137–159

So KKY, Duncan RW (2021) Breeding Canola (Brassica napus L.) for Protein in Feed and Food. Plants. doi: 10.3390/PLANTS10102220

Springer NM, Lisch D, Li Q (2015) Creating order from chaos: Epigenome dynamics in plants with complex genomes. Plant Cell **28**: 314–325

Srivastava PK, Moturu TR, Pandey P, Baldwin IT, Pandey SP (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. BMC Genomics **15**: 348

Stone SL, Braybrook SA, Paula SL, Kwong LW, Meuser J, Pelletier J, Hsieh TF, Fischer RL, Goldberg RB, Harada JJ (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. Proc Natl Acad Sci U S A **105**: 3151–3156

Stuessy T, Weiss-Schneeweiss H (2019) What drives polyploidization in plants? New Phytologist **223**: 1690–1692

Szyrajew K, Bielewicz D, Dolata J, Wójcik AM, Nowak K, Szczygieł-Sommer A, Szweykowska-Kulinska Z, Jarmolowski A, Gaj MD (2017) MicroRNAs are intensively regulated during induction of somatic embryogenesis in arabidopsis. Front Plant Sci **8**: 18

Tan H, Zhang J, Qi X, Shi X, Zhou J, Wang X, Xiang X (2019) Correlation analysis of the transcriptome and metabolome reveals the regulatory network for lipid synthesis in developing Brassica napus embryos. Plant Mol Biol **99**: 31–44

Tang Y, Yan X, Gu C, Yuan X (2022) Biogenesis, Trafficking, and Function of Small RNAs in Plants. Front Plant Sci **13**: 825477

Traver MS, Bartel B (2023) The ubiquitin-protein ligase MIEL1 localizes to peroxisomes to promote seedling oleosin degradation and lipid droplet mobilization. Proc Natl Acad Sci U S A **120**: e2304870120

Tsai AYL, Kunieda T, Rogalski J, Foster LJ, Ellis BE, Haughn GW (2017) Identification and Characterization of Arabidopsis Seed Coat Mucilage Proteins. Plant Physiol **173**: 1059–1074

Vermon DM, Meinke DW (1994) Embryogenic transformation of the suspensor in twin, a polyembryonic mutant of Arabidopsis. Dev Biol **165**: 566–573

Wan L, Ross ARS, Yang J, Hegedus DD, Kermode AR (2007) Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant Arabidopsis thaliana (thale cress) seeds. Biochemical Journal **404**: 247

Wang F, Perry SE (2013) Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol **161**: 1251–1264

Wang J, Mei J, Ren G (2019a) Plant microRNAs: Biogenesis, homeostasis, and degradation. Front Plant Sci **10**: 360

Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, Sun S, Lei T, Meng F, Wei C, et al (2019b) Polyploidy Index and Its Implications for the Evolution of Polyploids. Front Genet. doi: 10.3389/fgene.2019.00807

Wei L, Du H, Li X, Fan Y, Qian M, Li Y, Wang H, Qu C, Qian W, Xu X, et al (2022) Spatio-temporal transcriptome profiling and subgenome analysis in Brassica napus. The Plant Journal **111**: 1123–1138

West MAL, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. Plant Cell **6**: 1731–1745

Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell **75**: 855–862

Wilkinson LG, Yang X, Burton RA, Würschum T, Tucker MR (2019) Natural Variation in Ovule Morphology Is Influenced by Multiple Tissues and Impacts Downstream Grain Development in Barley (Hordeum vulgare L.). Front Plant Sci **10**: 1374

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y (2018) Homoeolog expression bias and expression level dominance in resynthesized allopolyploid Brassica napus. BMC Genomics **19**: 1–13

Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, et al (2013) The genome of the pear (Pyrus bretschneideri Rehd.). Genome Res **23**: 396–408

Wu Y, Llewellyn DJ, White R, Ruggiero K, Al-Ghazi Y, Dennis ES (2007) Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of the rapidly expanding fibre initial cells on the surface of cotton ovules. Planta **226**: 1475–1490

Xu TT, Ren SC, Song XF, Liu CM (2015) CLE19 expressed in the embryo regulates both cotyledon establishment and endosperm development in Arabidopsis. J Exp Bot **66**: 5217–5227

Xu W, Fiume E, Coen O, Pechoux C, Lepiniec L, Magnani E (2016) Endosperm and Nucellus Develop Antagonistically in Arabidopsis Seeds. Plant Cell **28**: 1343–1360

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

Yang B, Song Z, Li C, Jiang J, Zhou Y, Wang R, Wang Q, Ni C, Liang Q, Chen H, et al (2018) RSM1, an Arabidopsis MYB protein, interacts with HY5/HYH to modulate seed germination and seedling development in response to abscisic acid and salinity. PLoS Genet **14**: e1007839

Yang X, Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics **27**: 2614–2615

Yuan J, Jiao W, Liu Y, Ye W, Wang X, Liu B, Song Q, Chen ZJ (2020) Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. BMC Biol **18**: 1–13

Zhan J, Thakare D, Ma C, Lloyd A, Nixon NM, Arakaki AM, Burnett WJ, Logan KO, Wang D, Wang X, et al (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. Plant Cell **27**: 513–531

Zhang MX, Zhu SS, Xu YC, Guo YL, Yang WC, Li HJ (2020) Transcriptional repression specifies the central cell for double fertilization. Proc Natl Acad Sci U S A **117**: 6231–6236

Zhang Y, Liang W, Shi J, Xu J, Zhang D (2013) MYB56 Encoding a R2R3 MYB Transcription Factor Regulates Seed Size in Arabidopsis thaliana. J Integr Plant Biol **55**: 1166–1178

Zhang Y, Zhang L, Xiao Q, Wu C, Zhang J, Xu Q, Yu Z, Bao S, Wang J, Li Y, et al (2022) Two independent allohexaploidizations and genomic fractionation in Solanales. Front Plant Sci **13**: 1001402

Zhao M, Zhang B, Lisch D, Ma J (2017) Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. Plant Cell **29**: 2974–2994

Zheng P, Zheng C, Otegui MS, Li F (2022) Endomembrane mediated-trafficking of seed storage proteins: from Arabidopsis to cereal crops. J Exp Bot **73**: 1312–1326

Zhou M, Law JA (2015) RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules. Curr Opin Plant Biol **27**: 154–164

Ziegler DJ, Khan D, Kalichuk JL, Becker MG, Belmonte MF (2019) Transcriptome landscape of the early Brassica napus seed. J Integr Plant Biol **61**: 639–650

Ziegler DJ, Khan D, Pulgar-Vidal N, Parkin IAP, Robinson SJ, Belmonte MF (2023) Genomic asymmetry of the Brassica napus seed: epigenetic contributions of DNA methylation and small RNAs to subgenome bias. The Plant Journal **115**: 690–708

(2018) FAOSTAT. http://www.fao.org/faostat/en/#data/QC

(2017) Canola now worth $26.7 billion to Canadian economy - Canola Council of Canada. https://www.canolacouncil.org/news/canola-now-worth-$267-billion-to-canadian-economy/

5 Contribution of Knowledge and Future Works

5.1 Using Subregion-level transcriptomics to increase resolution of developmental networks

The seed is a complex structure, relying on the coordination of many moving parts to protect and nourish the next sporophytic generation. The exceptional complexity within seeds cannot be solely researched with whole seed studies, as the seed undergoes many developmental steps which further differentiate many regions of similar ontogeny, such as the endosperm and seed coat subregions. This work provides a detailed overview of that exceptional complexity with powerful spatial resolution. The benefit of this resolution was conspicuously higher gene detection which resulted in less gene expression being lost to RNA-sequencing noise. Perhaps more importantly, we have gained a much more realistic picture of the transcriptome of smaller subregions of the seed (eg. CPT, endosperm regions, early EP), which otherwise have much of their transcription washed out by the seed coat during morphogenesis or the EP during maturation. Therein, these projects have since demonstrated that the subregions of the *B. napus* seed are transcriptomically and anatomically distinct, with characteristic features defining them. For example, the CPT was consistently divergent from other subregions, with defined anatomical transitions throughout seed development coinciding with a distinct transcriptomic profile. Further, the chalazal subregions are further supported in this work to operate as a signalling hub to orchestrate and coordinate developmental signals, a concept proposed in *Arabidopsis* prior (Belmonte et al., 2013b). Given that *Arabidopsis* (Camelineae) is much more closely related to the basal lineage of the Brassiceae than *Brassica* (Brassiceae), it is possible the chalaza serves this function across the Brassicaceae and warrants

further investigation of these regions in intermediary lineages, especially in current and prospective oilseeds.

Examining the role of the chalaza in coordinating development requires more validation work to be done. In particular, our understanding of how these regions contribute to seed development would benefit greatly from *in situ* hybridization to track the movement of transcripts involved in seed development (FUSCA, LEA, LEC) and/or those known to be involved in seed storage deposition, including OLEO, CRU, SESA. In particular, these data indicate that the EP rapidly declines in transcriptional activity upon reaching maturation, implying the other seed subregions supported and nourished the EP enough to permit it to only accrue storage reserves. *B. napus* is one of the world's most important oilseeds, and accumulates these reservoirs efficiently — as such, this work furthers our understanding of how this oilseed builds upon its reserves to make a healthy, nutritious seed.

5.2 Subgenome biases demonstrate which component genomes create the seed

It is hypothesized crop plants are often polyploid due to enhanced genetic diversity which enables artificial selection and domestication (Qi et al., 2021). *B. napus* is a nascent allopolyploid which became a globally important oilseed likely due in part to this phenomenon. The work within this thesis details the dynamics of subgenome bias as the transcriptomic, DNA methylation, and small-RNA level, and advances our understanding of which constituent parts of the genome are dominant or submissive. We showed non-coding elements of the genome (small RNA loci, TEs) retain more of their epigenetic architecture than the coding and *cis*-regulatory elements across subgenomes, and also described the asymmetry in homologous gene expression within seed subregions. Together these data help paint a clearer picture of how *B. napus* has maintained its genome post-polyploidization, and which homologous genes are

dominant over others —demonstrating that even in processes crucial for seed development, such as seed storage, subgenome dominance is still exerted on specific gene homologs.

Future work in *Brassica* polyploid genetics would benefit from functional genetic validations of submissive and dominant gene pairs, wherein the submissive gene is cloned and assessed for functionality. Furthermore, the accompanied *cis*-regulatory elements of said gene pairs are likely to be affected by these biases and are worth similar investigation. At a larger scale, these homologous gene pairs could be further studied by pulling from the more fractionated genomes (MF1, MF2) to assess if ancient genome triplication and subsequent fractionation has substantially affected gene function or not. At the epigenetic level, ATAC-seq would be an excellent next step in determining if homologous chromosome regions between subgenomes were similarly accessible. Our data presented here are an excellent resource for future validation work and help guide further polyploid research in unravelling the convoluted dynamics of polyploidy in flowering plants.