Functional effects of inter-genome interactions on seed development in *Arabidopsis thaliana*

Volume 1 of 1

Gilles Aurélien DUPOUY

For the degree of Doctor of Philosophy

School of Natural Sciences

Plant and AgriBioscience Centre, Plant and Biotechnology Lab

Under the supervision of Professor Charles Spillane

Supported by the Irish Research Council

Submitted on the 31rst August 2021

Table of Contents

List of publications and conferences

Posters/presentations

- 23rd of February 2018: Poster presentation “Genomic imprinting as an epigenetic modifier of seed development in plants” at the Ryan Institute Research day in NUI Galway, Ireland.
- 31rst of May 2019: Poster presentation “Imprinting genes under Positive Darwinian Selection as Modifiers of F1 Seed Size in Arabidopsis thaliana” at the Ryan Institute Research day in NUI Galway, Ireland.

PhD Course Work

- AgriFood Sustainability & Agri-Resilience Challenges (2016)
- Understanding Ireland’s Agriculture & AgriFood Sector (2016)
- Understanding AgriBusiness & AgriFood Market Trends (2017)
- AgriFood Career, Communication & Impact Pathway Skills (2017)
- Journal Club Programme (2018)
- Research Paper Publication (2018)
- Research Integrity (2019)

Conferences attended

- Ryan Institute Research day 2018, NUIG, Ireland
- Ryan Institute Research day 2019, NUIG, Ireland
- IPSAM 2019, Carlow, Ireland

List of Tables

List of Figures

List of Abbreviations

ABA: Abscissic acid

ANU%: Percentage of Aborted, Normal and Unfertilized seeds

BR: Brassinosteroids

BPH: Best parent heterosis

cDNA: complementary DNA

CHD: Chromodomain-helicase-DNA-binding

CK: Cytokinins

cTP: Chloroplast transit peptide

DAP: Days after pollination

DEG: Differently expressed genes

DNA: Desoxyribonucleic acid

EBN: Endosperm balance number

easiRNA: epigenetically activated small interfering RNA

FIS: Fertilisation independent seed

GA: Gibberelic acid

GFP: Green fluorescent protein

GO: Gene ontology

iMEG: Imprinted maternally expressed gene

iPEG: Imprinted paternally expressed gene

MPH: Mid parent heterosis

MS: Murashige and Skoog

MTZ: Maternal to zygote

NH: Negative heterosis

NLS: Nuclear localisation signal

PcG: Polycomb group

PCR: Polymerase Chain reaction

PDS: Positive Darwinian selection

PGRs: Plant growth regulators

PRC2: Polycomb repressive complex 2

PRP: Plastid ribosomal protein

PRPL5: Plastid ribosomal protein L5

PKL: PICKLE

PKR2: PICKLE RELATED 2

QTL: Quantitative time loci

RdDM: RNA-directed DNA-methylation

RILs: Recombinant inbred lines

RNA: Ribonucleic Acid

RP: Ribosomal protein

siRNA: Small interfering RNA

TAIL-PCR: Thermal Asymmetric Interlaced PCR

TE: Transposable element

TEM: Transmission electron microscopy

Trx: Trithorax

WPH: Worst parent heterosis

WT: Wild-type

Declaration

I certify that this thesis is my own work, and that I have not used this work in the course of another degree, either at National University of Ireland Galway, or elsewhere.

Signed: 

_____

Gilles Dupouy

Acknowledgements

Many thanks to Professor Charles Spillane to have allowed me to undertake this PhD in the Plant and biotechnology lab from NUI Galway. Many thanks also to Sara Farrona, Peter McKeown and Alexander de Menezes for being such wonderful, welcoming and helpful GRC committee members. I also want to warmly thank my lab co-workers and other members of PABC for supporting me and helping me progressing through my different projects and keeping my sanity when things did not go according to plan: Kallyne Ambrosio, Alberto Abrantes, Godwin James, Lorene Lefèbvre, Charlene Linderhoff, Ronan Cashell, Mélanie Thomas, Brendan Hallahan, Chiara Cazzaniga, James Friel, Rosa Castillo, Francesca Lopez, Brian Joyce, Linyi Lai, Sandesh Rao, Steven McLoughlin, Henrique Pires, Joelle de Groot, Antoine Fort and Marcus McHale. Special thanks to Galina Brychkova and Peter McKeown for the feedbacks on my different chapters, and especially to Noel Lucca for all the time, help and support through the last years.

I would also like to send a huge thank you to the Galway music sphere: all the members of NUIG orchestra society and Choral Society, NUIG staff singers, TestosterTones, Galway Early Music, ConTempo Quartet, Luminosa String orchestra, Galway Adult string orchestra, Galway Traditional Orchestra and especially Josef Steiner and Ramin Haghjoo for all the good craic performing music which definitely helped me getting through these 5 intensive years.

Finally, I would like to warmly thank my family, the people from Naduir Farm and the other members of the Culcul Clan des quatre for all their love and support through all these years and for believing in me until the very end.

Abstract

Plant seed size determination is of high economic interest for modern food production worldwide as many crops are mainly consumed in the form of seeds, such as rice, wheat or corn. Parental conflict and genomic imprinting have been intensively studied during the last 20 years with many reports of their manifestations to significantly affect seed size determination in plants. However it has so far mostly been studied from the point of view of the nuclear genome leaving out the organelle genomes from the mitochondria and the chloroplast. In this Ph.D. thesis we explore how the interactions between these genomes can post-zygotically influence the development of seeds in angiosperms using the plant molecular model *Arabidopsis thaliana*.

After establishing a state-of-the-art of the current knowledge on the molecular manifestations of maternal-paternal conflict over the energy allocated to the offspring, I studied the effect of these genomic interactions using different approaches. I first studied the impact of parentally imprinted genes under positive Darwinian selection on the development of intra-species hybrid seeds of *A. thaliana* and revealed an essential role of the CHD3 chromatin remodeler *PICKLE RELATED 2* (*PKR2*) on the establishment of post-hybridity F1 seed size heterosis in disregard to its imprinting status. I further studied the evolution of *PKR2* in comparison to its paralog *PKL* and showed that *PKR2* is under a process of neofunctionalization since its apparition in the Brassicaceae lineage after the duplication of *PKL,* explaining their previously reported antagonistic effects on seed size in *A. thaliana.* I then explored the impact of a disruption of cytonuclear relationships (cybridity) and showed that cybridity and hybridity have an additive effect on F1 seed size determination. Finally, I performed a case-study on the impact of cytonuclear relationships on seed development by characterising the gene *PLASTID RIBOSOME PROTEIN L5* and have shown it to be necessary for post-globular embryo development in *A. thaliana*. These results highlight the interconnection of the different plant genomes for seed size determination as well as its importance for future plant breeding programs.

Keywords: Seed size; Genome interaction; Organelles; Kinship; Chromatin

Chapter 1 : General introduction

Introduction

Why is seed size biologically/biotechnologically of interest?

Angiosperm seeds, which are the basis of human food security, are the product of the double fertilisation of a female gametophyte (ovule) by a male gametophyte (pollen grain) (Nawaschin 1898; GUIGNARD 1899; Raghavan 2003). Seeds are crucial for human consumption of carbohydrates (e.g. wheat, maize or rice), oil (e.g. rapeseed, sunflower) or proteins (e.g. legumes) and are the basis for seed production systems. They can be used as a preservation state of species of interest for future generations, and are also the best way to exchange germplasm collections of varieties across countries. A need to understand the different actors of seed development, as well as seed content storage processes has thus arisen and is of key interest for breeding programs aiming to improve nutrient characteristics in crop species. As an increased seed size is usually correlated with increased seed content such as fatty acids and proteins and is also correlated with the production of more vigorous plants with high crop yield (Sundaresan 2005; Makinde et al. 2020), the mechanisms controlling final seed size in plants are of critical need of understanding for the improvement of seed nutritional values in crop species.

Seed components participating in seed size determination

The molecular mechanisms of seed development and its different components has been thoroughly studied (Grossniklaus and Schneitz 1998; Berger 1999; Köhler and Makarevich 2006; Curtis and Grossniklaus 2008; Chandler and Werr 2019; Ingram 2020). From these we understand that seed development is mainly influenced by the maternal plant: it depends on how much resources are available, how much of them are used by the maternal plant for survival and how much is actually available for reproductive functions. The quantity of offspring produced by the plant is also of key importance for the determination of available resources for each. The regulation of maternal seed coat development through the action of genes such as *SEEDSTICK* which appears to influence cell cycle via direct interaction with the transcription factor *E2Fa* has also been shown to majorly affect seed size (Paolo et al. 2021). Minor variations from growth rates set by maternally derived parameters are brought up by

the addition of paternally-inherited genetic components, in a different way across two of the three other main component of the seed: the embryo, i.e. the usually diploid progeny resulting from the fertilization of the maternal haploid egg cell by one of the paternal haploid sperm cell, and the endosperm, i.e. the usually triploid sink organ for maternal nutrient accumulation destined for the embryo and resulting from the fertilisation of the diploid maternal central cell by the other paternal haploid sperm cell. The seed coat, last usually diploid main component of the seed is solely of maternal origin and is a tissue assuming a role of protection for the developing embryo (integuments) and of nutrient exchange from the plant in direction of the endosperm (chalaza).

These three main elements each have a role to play in final seed size determination. I will not detail here the role of the maternal seed coat in the regulation of seed size as it has been thoroughly reviewed in *Arabidopsis thaliana* (Orozco-Arroyo et al. 2015) and in other species of the Brassicaceae lineage as well as maize (*Zea mays*), rice (*Oriza sativa*), wheat (*Triticum aestivum*) and other cereals (Li et al. 2019b; Olsen 2020). Apart from these maternal tissues, final seed size is mostly determined by the growth of the endosperm as a syncytium. The size expansion of the endosperm has been reported to be powered by an increase in turgor pressure followed by the cell multiplication and expansion of the maternal integuments surrounding it. The endosperm turgor pressure appears to decrease simultaneously with the initiation of the cellularization of the syncytium, setting the end of its expansion (Beauzamy et al. 2016). The growth of the integuments could be highlighted as a major part of the overall seed expansion, but has been shown to actually develop in response to the expansion of both embryo and endosperm, seemingly as a way to leave room for their development (Boisnard-Lorig et al. 2001).

Compared to the important role of the endosperm on seed size determination, the embryo seems to have a minor role, identified as negligible. The proliferation of the syncycial endosperm indeed appear to be the main actor setting the growing rate of the seed, until the endosperm is consumed by the growing embryo (for exalbuminated seeds). After this consumption the seed has not been reported to grow further than the limitations previously set by the endosperm and the seed coat

under normal growing conditions (Bremner et al. 1963; Sundaresan 2005). However, a more recent Knock-Out mutation of *AGC1-4* which encodes a serine-threonine kinase has been shown to lead to an increase in embryo size during the late stages of seed development correlated with an final seed size increase in comparon to Wild-type (Zhang et al. 2020c). It is unfortunately not clear from the data provided if the increased seed size originates solely from the embryo development, or if the endosperm proliferation is still paving the process.

Potential for parental conflict over resources supporting seed growth

Differential involvement of maternal and paternal genome in the process of seed development has been shown to be an important part of the molecular regulation influencing seed size. Indeed, apart from the maternal seed coat tissue, the embryo nuclear genome is normally equally maternal and paternal of origin, and the endosperm nuclear genome is two parts maternal and one part paternal of origin (Scott et al. 1998). Organelle genomes of angiosperms were previously thought to mostly be inherited from the maternal side with few paternal leaks at a frequency of 3.9×10^{-5} in *A. thaliana (Dilkes and Comai 2004; Azhagiri and Maliga 2007b)*, but a more recent study on Yew (*Taxus brevifolia*) has revealed that this assumption might be misguided. *T. brevifolia* indeed displays a paternal inheritance of plasts, and a biparental inheritance of mitochondria (Anderson and Owens 1999). It would thus be safer to assume a species-dependant parentally-biased inheritance of organelles.

Kinship theory, or parent-offspring conflict theory (Trivers 1974) was elaborated to model the energy exchange between maternal parents and their offspring, resulting on a conflict towards how much of the maternal energy is to be allocated to the offspring. It was later adapted to seed development, modelling further the maternal-offspring relationship as resulting from a conflict of interest between maternally-inherited and paternally-inherited genomes in both embryo and endosperm (Haig and Westoby 1989). According to this theory, the paternally-inherited genome would aim to increase energy uptake by the embryo and endosperm which it has fertilised, whereas the maternally-inherited genome would aim to equally distribute maternal energy resources to all the seeds being produced (figure 1.1). This arms race over the

allocation of nutrient uptake by the offspring would have a direct effect on final seed size, especially in an outbreeding context where maternal and paternal actors have evolved independently and fail to suppress each other's efforts to regulate seed size in their own interest (Willi 2013).

Theoretical framework for assessing evidence for parental conflict

In this introduction, I will be reviewing our current understanding of the role of maternally-inherited and paternally-inherited molecular actors on seed size determination under the light of the kinship theory. I will assess the current knowledge on the effect of an imbalance of maternal and paternal genome dosage on seed development as it is currently known to have a critical impact especially on endosperm proliferation and cellularization. From this, I will also explore the current knowledge on epigenetics, commonly admitted as representing the ensemble of marks associated to genomic DNA and histones which can have an influence on the expression of nearby genes on a stable and transmissible manner (McKeown and Spillane 2020). I will especially investigate the role of genomic imprinting characterised by an allelic differential expression of genes originated from gametogenesis-inherited epigenetic marks (Reik and Walter 2001; Köhler et al. 2012; Batista and Köhler 2020), as well as their association with plant growth regulators (PGRs). Imprinted genes seems to intervene significantly in genome dosage variations effects on seed development, with some having a critical impact on endosperm development. Finally a last source or parental conflict will be explored in the interactions between nuclear and organelle genomes.

As most of the published molecular work has been done on the plant molecular model *Arabidopsis thaliana*, I will go over the most recent data concerning this species. However as *A. thaliana* it is a quasi-exclusive inbreeder, I will also review the growing number of scientific studies published on outbreeder crop species.

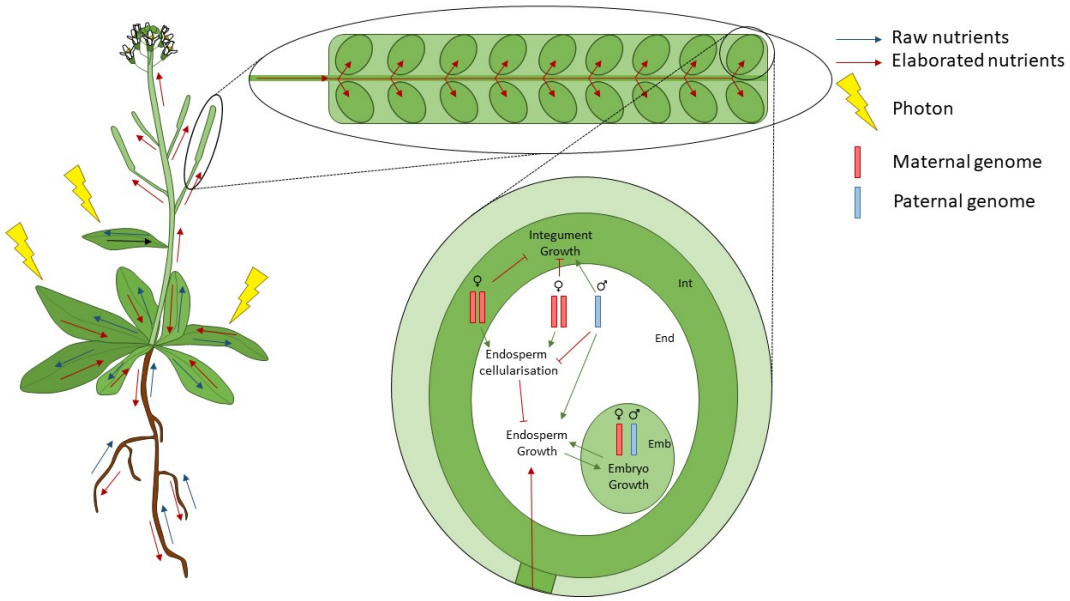


Figure 1.1: Maternal – kin tug-of-war on resources allocation
The energy available for one seed is highly dependent on the energy available in the environment, the capacity of the maternal plant to absorb raw nutrients from the soil and transform them into the components of the elaborated sap using the light available and also on the number of seeds produced per plant. The parental Kinship theory comes from a theoretical conflict between maternal plant and kin on the allocation of energy to the embryo, especially from the mechanism controlling endosperm growth before cellularization. Parental excess experiments in which the endosperm balance number (EBN) differs from the usual 2m:1p highlight conflictual relationships between paternally- and maternally-inherited genomes. In this theory maternally-inherited genomes in endosperm and seed coat maternal genome tend to promote endosperm cellularization over its proliferation whereas paternally-inherited endosperm genome tend to promote the opposite. This results in an underdeveloped endosperm in maternal excess triploid seeds and an over-proliferated endosperm without cellularization in paternal excess triploid seeds.

Int: Integument; End: Endosperm; Emb: Embryo

Genetic and Epigenetic determination of seed size

Parental genome balance has an major effect on seed size

Apart from seed coat intervention, final seed size appears to be primarily determined by endosperm development highly dependent of parentally-inherited factors. Impairment of reciprocal maternal/paternal genome dosage in the endosperm compared to an observed optimum ratio, i.e. 2m:1p or Endosperm Balance Number (EBN), has been reported to have dramatic consequences on seed size in *A. thaliana* (Scott et al. 1998), and more recently in genus *Nicotiana* (He et al. 2020). A decrease in seed size due to a reduced endosperm proliferation was indeed observed in case of an excess of maternally-inherited genome copy number (maternal excess) while an increase of seed size due to an excessive endosperm proliferation was observed in case of an excess of paternally-inherited one (paternal excess). This paternal excess-related endorperm overproliferation was further correlated with a down-regulation of G2 & M mitosis phases-associated genes and an up-regulation of G1 & S phases-associated genes promoting the endoreduplication process without cellularization Tiwari et al. (2010). This was also shown in *fis* and *msi1* mutants which display a similar phenotype in seeds than paternal excess triploids. This suggests the existence of maternal endosperm growth inhibiting factors, and of paternal promoting ones.

An excess of paternally-inherited genome is associated to an endosperm cellularization defect leading to seed failure in most seed progeny (Nishiyama and Inomata 1966; Lin 1984; Dilkes and Comai 2004; Stoute et al. 2012), a phenomenon called triploid “block”. This block was shown by Dilkes et al. (2008) to arise in an accession-dependant manner in *A. thaliana*, with Col-0 being one of the accessions with the heaviest block (80% of seed development failure), whereas triploid block in Ler-0 appears much weaker (20% of seed failure). This genome dosage effect cannot however be fully explained from a whole gene expression doubling from either parental sides. A recent study by Song et al. (2020) has indeed shown that genome duplication can scarcely be correlated with gene expression doubling with putative dosage sensitive genes especially appearing to be down-regulated as opposed to expectations. This may suggest that gene regulation differs within progeny

individuals, which could explain why some paternal-excess triploid individuals survive the block while an accession-dependant number of them abort.

Viability of paternal-excess triploid embryos was recently discussed as *in vitro* polyploidisation of zygotes of *Oryza sativa* has shown embryo arrest in three quarters of cases of paternal excess triploidy (Toda et al. 2018). However, polyspermy fertilisation in *A. thaliana* has more recently allowed the development of triploid paternal excess embryos with a "normal" triploid endosperm to bypass the triploid block (Mao et al. 2020). It thus seems that modifications of the EBN would be the determining factor creating the "block" in paternal excess triploid, a regular 2m:1p ratio in the endosperm being preferred for triploid embryo survival. Triploid embryos development systematically failed in the absence of an endosperm, even when cultured in a media containing required nutrients and PGRs. This correlates the dysregulations of genes participating in nutrient reservoir activity pathways found in triploid hybrids by Fort et al. (2017). With the hypothesis of a "quality-check" control by paternal dosage sensitive genes in the endosperm to explain this block (Walbot and Evans 2003), it seems logical to explore parentally biased epigenenetic actors, such as genomic imprinting, in order to understand EBN effects on seed size and the extreme cases of seed failure associated to endosperm overproliferation.

Evidences of the role of epigenetics and genomic imprinting in seed size determination

The definition of epigenetic marks highlight the fact that they can be transmitted from one cell to daughter cells by mitosis. There is high potentiality for these marks to be inherited by gametes through meiosis, as well as these marks to remain on the genome inherited by both parents after fertilization of both egg-cell and central cell. Genomes inherited by both parents thus can be transmitted with epigenetic marks reflecting an optimal expression of particular genes relative to each parent. First discovered in 1970 in Maize with the parental dependence of the R-mottled phenotype (Kermicle 1970) and then in mammals (McGrath and Solter 1984), this parental epigenetic imprint can create an imbalance of expression between both alleles of a given gene, creating imprinted maternally expressed genes (iMEG) when

the maternal allele is more expressed than the paternal allele, and imprinted paternally expressed genes (iPEG) in the opposite case (Köhler et al. 2012; Batista and Köhler 2020).

Often present near transposable Element (TE)-like sequences (Walker 1998), the only cases of imprinting so far have been identified in flowering plants and placental mammals, suggesting that the presence of accessory tissues involved in transfer of maternal nutrients to the offspring may be a driver of the evolution of genomic imprinting (Killian et al. 2001). In Angiosperms, imprinting has been mostly identified in the endosperm although some genes are also expressed in an imprinted manner in the embryo. Most imprinted genes in the embryo however progressively revert to biallelic expression during embryo maturation (Jahnke and Scholten 2009). The current models of imprinting genomics is mostly relying on two different epigenetic repression marks deposition pathways. On one hand the fertilization independent seed (FIS) - Polycomb group (PcG) Repressing complex 2 (PRC2) was shown to deposit H3K27me3 repressive marks on the maternal allele of most iPEGs in the central cell. On the other hand other actors such as DNA METHYL TRANSFERASE 1 (MET1) and the RNA-directed DNA-methylation (RdDM) pathway associated with epigenetically activated small interfering RNAs (easiRNA) have been shown to deposit CpG islands methylation repressive on the paternal alleles of iMEGs and maternal TE, while a DEMETER (DME)-associated wide genome demethylation is observed in the maternal gametophyte prior to fertilization (Batista and Köhler 2020; Erdmann et al. 2017; Borges et al. 2018; Anderson et al. 2020).

This genomic imprinted phenomenon in the endosperm appears to be an important factor of seed survival after hybridization as strong dysregulations of imprinted genes have been reported a common feature in abortive hybrid seeds (Florez-Rueda et al. 2016; Muyle et al. 2018). Genomic imprinting and particularly iPEGs seems to have a particularly tight connection with EBN (Jullien and Berger 2010) and have been linked with interploidy hybridization barriers, first shown in *A. thaliana* (Wolff et al. 2015; Jiang et al. 2017) and more recently in the genres *Capsella* (Lafon-Placette et al. 2018) and *Mimulus* (Kinser et al. 2018). An overexpression of paternally-biased siRNA was also linked to the establishment of postzygotic barrier and seed failure in paternal-excess triploids in *A. thaliana* (Martinez et al. 2018). The deregulation of some

repressive chromatin remodelers participating in H3K9me2 deposition was correlated with paternal-excess triploid seed endosperm failure (Jiang et al. 2017). Genomic imprinting would thus be a good candidate to explain the impact of parental genome imbalance on seed size and seed survival.

An increasing number of endosperm iMEGs and iPEGs have been predicted so far in a large panel of studies in *A. thaliana* (Gehring et al. 2011; Hsieh et al. 2011; McKeown et al. 2011; Wolff et al. 2011; Pignatta et al. 2014; Schon and Nodine 2017; Köhler et al. 2018), in the genus *Capsella* (Hatorangan et al. 2016; Lafon-Placette et al. 2018), in *Z. mays* (Waters et al. 2011), in *O. sativa* (Luo et al. 2011; Rodrigues et al. 2013; Yuan et al. 2017; Chen et al. 2018a) and in *Glycine max* (An et al. 2017). Among then different components of the PRC2, most being encoded by iMEGs such as *MEA* and *FIS2*, are required to prevent an ectopic proliferation of the endosperm prior to fertilisation and thus has a major control on seed size determination (Chaudhury et al. 1997; Spillane et al. 2000; Sørensen et al. 2001; Köhler et al. 2003a; Köhler et al. 2003b; Makarevich et al. 2006; Hermon et al. 2007). Members of PRC2 were also shown to be downregulated in the integuments after central cell fertilisation in order to deactivate the block on the seed coat proliferation (Figueiredo et al. 2016). Overall, a Knockout of the different FIS-class iMEGs tend to result in an autonomous overproliferation of the endosperm prior to any fertilization and without any cellularization, a phenotype resembling the previously reported paternal excess triploid block. A global de-repression of the maternal alleles of iPEGs in the endosperm of these *fis* mutants has also been identified (Figueiredo and Köhler 2018). A Knockout of iPEGs such as *PKR2* and *ADM* was interestingly shown to restore the impaired endosperm cellularization in both *fis* mutants and paternal excess triploids (Huang et al. 2017). A paternal Knockout of the RNA Polymerase IV (Pol IV) known to control the polymerisation of 21-, 22- and 24-nt siRNA was also shown to restore endosperm cellularization and formation of viable seeds in paternal-excess triploid seeds, showing the importance of the RdDM pathway in the control of the endosperm proliferation (Martinez et al. 2018). Interestingly, most of reported imprinted genes in *O. sativa* have been shown to participate in the regulation of nutrient metabolism and endosperm development, hence having a likely effect on seed size (Yuan et al. 2017).

All these arguments tend to point towards a significant intervention of genomic imprinting in the endosperm, especially iPEGs, in the creation of an apparent tug-of-war between maternally- and paternally-inherited genomes towards the determination of seed size.

PGR signalling is parentally biased during seed development

A key element linking parental biases and endosperm proliferation seems to be found in PGRs signalling. Dramatic effects of genome imbalance as described in the previous sections have revealed the importance of endosperm cellularization timing throughout seed development on the determination of seed size. The maximum size appears to be obtained upon the moment the cellularization process starts resulting in endosperm proliferation arrest. The timepoint of endosperm cellularization closely correlates shifts in PGRs balances during seed growth, as summarized by Locascio et al. (2014) for *A. thaliana* (figure 1.2) and *Z. mays*.

A first PGR balance is formed by a progressive increase in Auxin and decrease in Cytokinins (CK) and brassinosteroids (BR) concentrations throughout the whole seed development. More recent reports refine this model by revealing a decrease in Auxin concentration in the endosperm at a timepoint correlating with the initiation of cellularization suggesting a more restricted role of auxin promoting endosperm proliferation Batista et al. (2019a). This highlights an antagonistic roles of auxin and CK towards the development of the seed. Auxin was reported to promote endoreduplication in the syncytial endosperm in *Z. mays* (Bernardi et al. 2012) and CK was linked to endosperm cellularization (Li et al. 2013) and was reported to participate in the limitation of seed expansion during the morphogenesis stage, as seen with *ahk2; ahk3; ahk4* triple mutant (Riefler et al. 2006).

The second hormonal balance is characterized by an increase of Gibberelic acid (GA) and decrease of Abscissic acid (ABA) concentrations during the endosperm proliferation phase (morphogenesis), followed by an inversion of the balance upon endosperm cellularization (maturation phase) with finally another increase of GA and plateauing of ABA concentrations during the desiccation phase. ABA was also linked to endosperm cellularization by Cheng et al. (2014), which correlates the increase in ABA accumulation observed upon this timepoint.

The high concentrations on auxin throughout the whole development process suggest an important role of this PGR for seed development indicating that it might be a major actor in seed size determination. Auxin is especially generated in the endosperm and is then further transported into the seed coat in an AGL62-dependant manner (Lur and Setter 1993; Figueiredo et al. 2015). Seed coat proliferation has also been associated with an increased GA signal, likely coming from an auxin-mediated repression of PRC2 in the seed coat. BR production and signalling from the seed coat coupled with auxin was also shown to participate in endosperm proliferation in *O. sativa* (Zhang et al. 2020b), while ABA has been described as having an important role in nutrient accumulation during morphogenesis, likely influencing seed size (Locascio et al. 2014).

Auxin biosynthesis and signalling has been identified as the pain participant to the overproliferation of the endosperm causing the seed abortion phenotype observed in paternal excess triploid seeds. Auxin is indeed sufficient to trigger endoreduplication in the syncytial endosperm as well as endosperm proliferation (Batista et al. 2019a). Overexpression of genes intervening in auxin biosynthesis and signalling was also reported to cause a paternal excess triploid seeds-like phenotype. The repression of such genes has also shown to rescue paternal excess-associated phenotypes. This suggests that auxin signalling in the seed might be paternally biased, which was confirmed in some ways by the identification of some auxin-related genes as iPEGs in *A.thaliana*. Among them *YUC10* has been confirmed as an iPEG in the endosperm by Batista et al. (2019a) after being identified in previous studies (Hsieh et al. 2011; Pignatta et al. 2014), alongside with *TAR1* (Wolff et al. 2011) and *TAA1* (Pignatta et al. 2014). Homologs of these genes were also reported to be paternally imprinted in the endosperm of *A. lyrata*, *C. rubella*, *O.sativa* and *Z.mays*, and several species of the genre *Brassica* (Figueiredo and Köhler 2018).

The iPEG *PHE1* (Köhler et al. 2005; Makarevich et al. 2008) has been identified as being likely expressed upstream of the activation of these auxin-related genes. It was indeed shown to act as a master regulator of multiple iPEGs including ones involved in auxin biosynthesis and signalling (*YUC10*, *TAR1* and *PGP10*) but also other genes involved into metabolic processes such as carbohydrate transport and triglyceride biosynthesis (Batista et al. 2019b). The paternal expression of *PHE1* upon fertilisation

followed by an activation cascade of several *AGAMOUS-LIKE* (AGL) genes and iPEGs resulting in an increase of auxin biosynthesis in the endosperm seems thus to be a likely pathway promoting endosperm proliferation prior to cellularization (Figure 1.2). The possibility of a synergistic role of *PHE1* and *AGL62* on endosperm cellularization has also been hypothesized (Wang et al. 2020).

A decrease of the expression of iPEGs controlling auxin signalling was observed in diploid WT seeds upon endosperm cellularization in the micropylar endosperm where the cellularization is initiated (Batista et al. 2019a). This down-regulation was not observed in paternal triploid excess seeds, correlating with the absence of endosperm cellularization in these seeds. The presence of two active paternal alleles of *PHE1* in paternal excess triploid seeds seems likely to be enough in order to trigger an overproduction of auxin causing endosperm overproliferation represented by the up-regulation of auxin-related genes such as YUC10, with *PHE1* also demonstrated as up-regulated (Tiwari et al. 2010).

The same phenotype was shown in the *mea* mutant and in hybrids seeds obtained from crosses between *A. thaliana* and *A. arenosa*, in which PRC2 fails to maintain the repression of the maternal allele of *PHE1* which is then biallelic expressed (Kiyosue et al. 1999; Köhler et al. 2005; Josefsson et al. 2006). An overexpression of paternally-biased genes involved in Auxin synthesis and signalling thus appears to be a good explanation for the endosperm overproliferation without cellularization in paternal-excess triploid seeds.

However in the case of maternal-excess triploid seeds the genetic state of *PHE1* in the endosperm post fertilisation is of one paternal active allele, and four maternal repressed alleles (Scott et al. 1998). If *PHE1* is the key initiator of the auxin signalling pathway in the endosperm, we would expect its proliferation to be as WT. In this particular case however a nearly inexistent proliferation of the endosperm is observed with a cellularization process being initiated earlier than in WT. The reason behind this can probably be found with the deregulation of auxin-related genes: *PHE1* and *YUC10* have indeed been found down-regulated in maternal excess, which would hint for a potential maternally-biased regulator repressing them in this context (Tiwari et al. 2010). The gene *MAX1* which participates in the regulation of the abundance of PIN transporters was also found up-regulated in maternal excess

triploids. *PIN4* was found down-regulated as a result, linking a decreased transport of auxin with the reduction of endosperm proliferation. This suggests the existence of a dosage sensitive maternal effector, which overexpression in maternal-excess triploid seeds would result in an early repression of *PHE1* and auxin signalling *AGLs* and iPEGs. This would limit auxin signalling very early in the seed development, facilitating an early cellularization of the endosperm.

Auxin might thus not be the only actor controlling endosperm proliferation. I have written earlier that ABA signalling had been linked to endosperm cellularization (Cheng et al. 2014), so it could be hypothesized that an Auxin/ABA balance to be another vector of control for the switch from endosperm proliferation to cellularization. *ABI5* which was reported to promote ABA signalling by inducing the repression of the HAIKU pathway in the endosperm (Kang et al. 2013) and has been shown in a more recent report to be stabilised by the timely exportation of the mobile element TERMINAL FLOWER 1 (TFL1) from the chalazal endosperm (Zhang et al. 2020a). The knockout of *TFL1* results in a delayed endosperm cellularization and an increase of seed size compared to WT further suggesting that the stabilisation of ABI5 is a factor promoting endosperm cellularization (Zhang et al. 2020a). This pathway could be interpreted as a maternal control over seed size from the close proximity of the chalazal endosperm to maternal vascular tissues by the mean of controlling the timely trigger of endosperm cellularization.

Another model emerges from these results, which mainly highlights a parentally biased balance of auxin against ABA and CK determining the frame of endosperm proliferation until the overthrow of the balance initiates the endosperm cellularization process. The activity of *CKX2* which was previously linked to the repression of CK signalling by CK oxidation appears to correlated this (Li et al. 2013). *CKX2* was indeed found up-regulated in paternal excess triploids and both *fis* and *msi1* mutants while repressed in maternal excess triploids, alongside GA and BR metabolism-associated genes Tiwari et al. (2010). This shows that CK signalling tend to be repressed in paternal excess triploids and both *fis* and *msi1* mutants, correlating with the loss of endosperm cellularization in these contexts.

However, as this model explains well the control over endosperm cellularization one also has to be careful with generalisation, as some species do not develop any endosperm syncytium, which in this case starts proliferating directly by cell division like in tomato for instance. In such species no paternal imprint of auxin-related genes have been reported, forcing us to evaluate the possibility of a different mechanism underlying endosperm development.

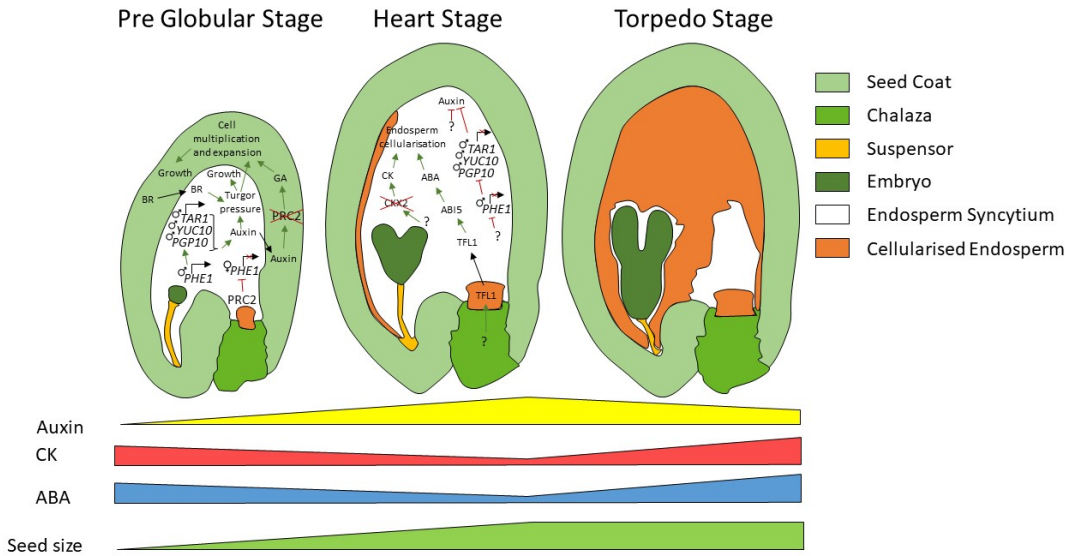


Figure 1.2: Seed growth driven by Endosperm development under parentally-biased PGR influence
The maternal allele of *PHE1* is repressed by PRC2 in the central cell of the female gametophyte. Expressed paternal allele is introduced in the endosperm upon central cell fertilization. The paternal allele directly activates auxin signalling related genes such as *YUC10*, *TAR10* and *PGP10*. Auxin level rises in the endosperm, promoting an increase of turgor pressure by a yet unknown mechanism which promotes the proliferation of the endosperm syncytium indirectly resulting in seed coat. Auxin is exported from the endosperm in the seed coat cells, promoting cell proliferation and expansion of the seed coat. Around heart stage, an unknown inhibitor represses the expression of *PHE1* paternal allele, thus repressing *YUC10*, *PGP10* and *TAR1* expression, resulting in a decrease in Auxin levels. In parallel, the TFL1 factor is imported from the chalazal endosperm to the endosperm where is stabilises ABI5 which promotes ABA signalling. A yet unknown mechanisms results in the repression of *CKX2* promoting CK signalling. The overthrow of the balance between auxin on one side and ABA and CK on the other side results in the arrest of syncytial endosperm proliferation and in the initiation of cellularization which stops the seed from increasing further in size.

Disruption of cytonuclear relationships shows potential for seed size variation

This introduction has so far highlighted the role of parent-of-origin effects in seed development by the importance of the maternal context and of inherited epigenetic features depending on the paternal or maternal inheritance of nuclear genes. It has been previously argued that parent-of-origin effects in plant reproduction in general must arise from some combination of epigenetic inheritance and long-lasting maternal effects, but also cytoplasmic inheritance (Duszynska et al. 2013; Duszynska et al. 2019). There is indeed potential to obtain parent-of-origin effects on seed development mediated by the inheritance of organelles since the mitochondrion and, in plants and algae, plastids possess their own genomes. These genomes may enter in conflict with the nuclear genome and affect fitness since they are not equally inherited by both parents, as explained above (Havird et al. 2019). In cases of biparental inheritance of plastids and mitochondria, as no recombination of genomes occur between organelles, this would create a competition between maternally and paternally inherited organelles. The lipid composition of organelle membrane has been identified as a major factor of this competition as it determines the organelle stability and division rate (Sobanski et al. 2019). Most mitochondrial (Sew et al. 2016) and plastid functions (Joyard et al. 2010; Tejos et al. 2010) are known to be crucial in *A. thaliana* seed development and their impairment usually lead to embryo maturation defect. Mitochondrial functions have been especially linked to the maternal inheritance of cytoplasmic male sterility in which pollen development is impaired (Schnable and Wise 1998).

Because both organelles are mostly inherited by the maternal side in angiosperm (with exceptions depending on the species), an evolutionary coupling of maternally inherited nuclear and organelle genes has been proposed (Figure 1.3). The model behind this hypothesis is elaborated on the crucial maintenance of the physical cohesion of protein complexes which subunits are encoded in different nuclear/cytoplasmic compartments (Wolf and Hager 2006; Wolf 2009). This can be observed for instance for the mitochondrial and plastid ribosomes which components are encoded by genes which have migrated from organelle genomes to nuclear genome through evolution (Allen 2018). Recent work done on *A. thaliana*

leaves has revealed that such genes tend to be coordinated in expression even after whole genome duplication, while the overall mitochondria transcripts ratios to nuclear abundance increased and plastidial transcript ratios decreased (Coate et al. 2020). This cytonuclear coupling has been experimentally demonstrated in *A. thaliana* with the plastid-encoded protein ClpP1, a subunit part of the plastid Caseinolytic Protease (Clp) complex. ClpP1 was reported to display a fast amino-acid substitution rate compared to the rest of the proteome in the same lineage (Williams et al. 2019) correlating with a previously reported amino-acid substitution rate acceleration in the nuclear-encoded Clp subunits (Weng et al. 2016). This cytonuclear linkage (i.e. the non-random association of organelle and nuclear alleles) model seem to favour self-reproduction for the inheritance of the maintenance of cytonuclear genome coupling, as maternal nuclear genes and cytonuclear genes would be inherited in block without being disturbed by the presence of a paternal outbreeder. This was theoretically highlighted by Wade and Goodnight (2007) and further bioinformatically confirmed by Moison et al. (2010) in *A. thaliana*.

A cytonuclear linkage disequilibrium by modifying the inheritance of coupled nuclear and organelle genes would thus be expected to affect plant fitness and possibly seed size. This has been recently correlated to hybridity effects observed in *Daucus carota* (Ramsey et al. 2019) and more recently demonstrated to have on its own an impact on seed size determination, at an intensity close to nuclear heterozygocity (Flood et al. 2020). An example of maternal cytonuclear linkage importance in *A. thaliana* seeds has been observed with *NUWA*, an iMEG encoding a P-subfamily pentatricopeptde (PPR) protein required for mitochondrial function in the early embryo (He et al. 2017). Late aborting *nuwa* mutant seeds displayed a reduced endosperm proliferation coming from a delayed endoreduplication in the syncytium in comparison to WT. Since NUWA products were identified in the early endosperm, this would suggest that *NUWA* has a role in endosperm proliferation and may thus have a role to play in seed size determination.

Even though there are only have few examples correlating this model, there seem to be some potential for cytonuclear relationships to create parental effects on seed size. The techniques allowing the study of cytonuclear linkage disequilibrium impacts

on plant fitness are however relatively new, but we can hope for an increased interest in the investigation of the role of cytonuclear linkage in the near future.

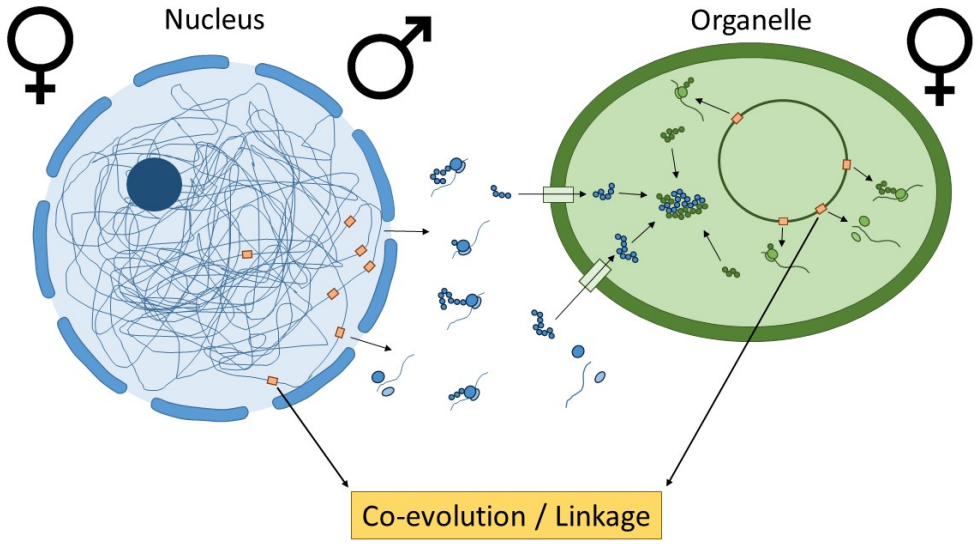


Figure 1.3: Inter-genome interaction and role of protein complexes cohesion for fitness.
Organelle protein complexes are composed of proteins encoded by genes which have for a major part been transferred over to the nuclear genome through evolution in a species-dependant way. The parental inheritance (usually maternal) of organelle encoded genes is highly dependent of the species, creating an evolutionary linkage between organelle and nuclear loci which is necessary to maintain the structural cohesion within the said complexes.

Intervention of gamete-inherited factors

One last aspect of parentally-biased mechanisms impacting seed development and size which needs to be discussed concerns the gametophytically inherited factors. These factors are usually mRNA produced and untranslated in the sperm cell to be later translated in the egg-cell upon fertilisation and initiate signalling cascades. The PPR protein-encoding *NUWA* appears to be a good example of this, as its maternal-specific product (RNA and proteins) are detected in mature gamete before pollinisation (He et al. 2017). The relative short mRNA half-life for this class of gene however suggests a short maternal-to-zygote (MTZ) transition for this gene. The role of NUWA as a maternally inherited factor might thus to be relatively small before zygotic production supplements it.

Another well-studied example can be found with the mitogen-activated protein kinase gene *YODA* (Figure 1.4), which promotes pavement cell identity in the embryo

by phosphorylating WRKY2 and later activating the axis patterning gene *WOX8* (Musielak and Bayer 2014; Ueda et al. 2017). The YODA cascade was reported to be initiated by the action of the paternally-inherited mRNA factor SHORT SUSPENSOR PROTEIN (SSP) (Bayer et al. 2009). Mutants of *YODA* mainly show an underdeveloped suspensor and an embryo which fails to elongate, whereas its overexpression displays suspensor overproliferation (Lukowitz et al. 2004). SSP was reported to accumulate in both zygote and micropylar endosperm, which would suggest its action and the following YODA cascade to be required in both fertilization products. Although it was discussed earlier in this chapter that under normal growth conditions the embryo does not play a major role in comparison to the endosperm and seed coat tissues in increasing final seed size, these mutants however do display an overall reduced seed size compared to wild-type with a reduced embryo and endosperm development.

SSP appears however to be partially redundant with another factor since the observed phenotype in *ssp* knockout mutants is less extreme than in *yda* mutants. This other factor might be found with the three maternally-inherited and central cell accumulated EMBRYO SURROUNDING FACTOR 1 (ESF1) peptides (Costa et al. 2014) and the gametophyte and embryo maternally-expressed HOMEODOMAIN GLABROUS 11/12 (Ueda et al. 2017). These factors have also been shown to be required for the YODA signalling pathway, and were reported to act synergistically with the paternally-inherited factor SSP, with their knockout mutants also resulting in smaller seeds compared to WT. Two maternally-biased repressors of *YODA* i.e. the *YODA* promoter binding ANGUSTIFOLIA 3 (AN3, Meng et al. (2017)) and the YODA binding ETHYLENE INSENSITIVE 3 (EIN3, Meng et al. (2018)) have been reported to repress seed growth, as their respective mutants result in an increased seed size compared to WT. It is interesting to highlight that YODA signalling is on one side promoted by one paternally- and two maternally-inherited/expressed effectors, and is on the other side repressed by two other maternally-expressed effectors. This can be interpreted as interfering maternal signals towards the YODA signalling cascade, although the promoting signal appears more upstream than the repressing one as most of its effectors are maternally-inherited while the repressing signal is zygotically derived.

Embryo growth through the inheritance of parentally-inherited factors thus appears to have a significant role to play in the determination of seed size, likely from some yet unknown mechanism of exchange with the endosperm which impacting its growth. These reported gametophytically-inherited signals interestingly seems to interfere with embryo maternal imprinting. Since genomic imprinting was reported only in early stages of embryo development (Jahnke and Scholten 2009), inherited factors might be of help to understand the role of the disappearing genomic imprinting in the embryo during seed development.

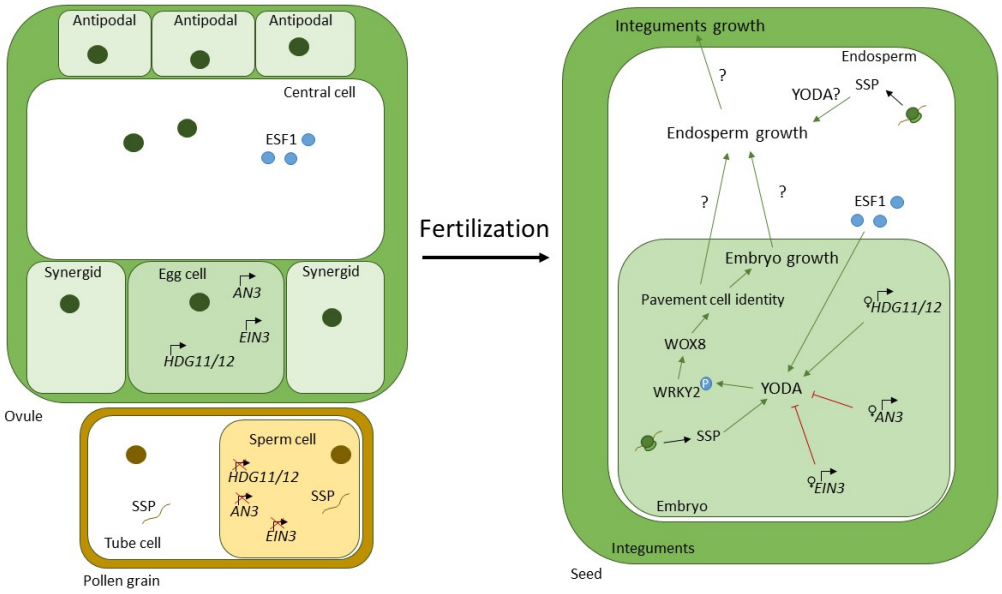


Figure 1.4: Parentally inherited factors and YODA signalling role in seed size determination.
The YODA cascade is necessary for normal WOX 8 controlled pavement cell identity during embryo development starting from YODA phosphorylating WRKY2. YODA and is under the control of different inherited factors and embryo parentally imprinted genes: it is activated thanks to the paternally inherited mRNA *SSP*, the three central cell accumulated peptides ESF1, and the iMEGs *HDG11* and *HDG12*, but is also repressed by the iMEGs *AN3* and inhibited by the iMEG *EIN3*. This shows conflicting maternal signals over the activation of the YODA phosphorylation cascade.

Arguments against the kinship theory

Aside of the influence of the maternal tissues, I have shown above that there is a significant importance of inter-genomic conflict and parentally-biased control of PGRs in the determination of seed size, especially through the role of paternally-biased genes. However, as good as this model seems to hold, there are also a

significant amount of counter arguments to be highlighted and which will have to be explained in order to fully understand the control of seed size determination.

Limitations of genomic imprinting as source of maternal-paternal conflict

The previous section summarised the different cases in which genomic imprinting could be seen as a molecular ground for the manifestation of the kinship theory during seed development, but this can unfortunately not be generalized to the global occurrences of parental imprinting. Only few cases of imprinted genes have indeed been reported to have a significant effect on seed size as most of them, especially iPEGs, have not shown any link with seed size determination (Wolff et al. 2015). For instance, *PKR2* does not appear to display any impact on seed size *per se* even if its knockout rescues the endosperm overproliferation of paternal-excess triploid seeds (Carter et al. 2016). This could suggest a higher impact of parental imprinting in outbreeding species where hybridization is the norm in comparison to what can be observed in inbreeders such as *Arabidopsis thaliana*. Indeed in outbreeding species the genomic imprinting of loci can drastically differ between the different mates as they usually have evolved in response to different environments. This can lead to a more severe competition between these loci in comparison to inbreeders (Willi 2013; Brandvain and Haig 2018). Thus the absence of effect of imprinted genes on seed size depicted in *Arabidopsis thaliana* may be imputable to the inbreeding characteristic of the species.

In other cases, some imprinted genes appear to behave in an opposite manner to what the kinship theory would predict. For instance the iMEGs *MEG1* in *Z. mays* (Costa et al. 2012) and *FH5* in *A. thaliana* (Gerald et al. 2009) are shown to act towards promoting nutrient allocation to the offspring as their knockout mutants display a reduction of the chalazal endosperm compared to WT. The *fh5* phenotype is rather counter-intuitive since according to the Kinship theory maternal imprinting impairment is rather characterised by an increase of endosperm proliferation without cellularization.

The homeodomain transcription factor FWA is another counter-intuitive iMEG which was identified as maternally imprinted in the endosperm of *A. thaliana* by Kinoshita et al. (2004). Its impairment was not reported to have any effect on seed

development and it appears up-regulated in *fis* mutants and paternal excess triploids seeds and down-regulated in maternal excess triploid (Tiwari et al. 2010). This gene expression behaviour is more closely related to iPEGs than to iMEGs. However, *FWA* was also shown down-regulated in *msi1* mutants even though these mutants mimic the endosperm overproliferation phenotype seen in paternal excess triploid and *fis* mutant seeds. *msi1* mutants are however different to other *fis* mutants since they display less peripheral endosperm nuclei than WT (Guitton et al. 2004), suggesting that *FWA* expression would linked to a yet undiscovered mechanism dysregulated in a different fashion between *fis* and *msi1* mutants.

The genomic imprinting phenomenon was also shown to be more variable than expected: recent studies have revealed some imprinted genes to display an accession-dependant pattern of allelic expression as it can be observed with the IV homeodomain-Leucine zipper gene *HDG3*. *HDG3* was shown to be expressed as an iPEG in Col-0 X Cvi crosses but also to be biallelically repressed in reciprocal Cvi X Col-0 crosses (Pignatta et al. 2018). The regain of paternal expression of *HDG3* in Cvi was linked with an early endosperm cellularization reducing seed size, which differs from the phenotype observed in Col-0.

Few explanations have been advanced so far to explain the behaviour of these counter-intuitive imprinted genes. One hypothesis suggests that the establishment of imprinting can be a way to avoid allele dominance effects in heterozygotes, and may also have appeared on some genes to match the expression of their interacting partners, even if the newly imprinted gene itself is not essential (Wolf 2013). This further suggest that genomic imprinting would rather be a consequence than a cause of the conflict over resource allocation to the offspring when two new breeding partners interact.

PGR signalling is not overly parentally biased

As discussed above, auxin-related genes such as *YUC10* or *YUC11* appear to be paternally imprinted in the endosperm of multiple species, suggesting a paternal control over auxin signalling (Figueiredo and Köhler 2018). However, a generalization over the whole of embryophytes seems impossible as homologs of these genes haven't been reported as imprinted in some species like the tomato (Florez-Rueda et

al. 2016) in which the endosperm develops as cellularised from the first stages, instead of starting its development as a syncytium.

Furthermore, a potential paternal control over auxin cannot be generalized either in species where auxin-related genes have been identified as paternally-biased. Indeed the auxin transporter *PIN1* which was not detected as imprinted in *A. thaliana* (Gehring et al. 2011; Hsieh et al. 2011; McKeown et al. 2011; Wolff et al. 2011) has surprisingly been identified as an iMEG in *Z. mays* (Xin et al. 2013). This rather suggest a combination of maternal and paternal factors over the control of auxin signalling under normal conditions, even though the paternal side of this control seems to become toxic for seed development in cases of paternal-excess genome imbalance when the maternal side is not. It can be added that even though some auxin-related genes are parentally-biased in *A. thaliana*, no gene related to any other PGR signalling has so far been reported as imprinted. Therefore, although there is a good correlation between PGR balances fluctuation and the timely establishment of endosperm proliferation and cellularization processes, a parental control over PGR signalling during the seed development cannot definitively be taken as a rule.

We also have to highlight the fact that so far only few of auxin-related genes have been identified as impacting seed development among their large numbers (Cao et al. 2020). The role of auxin in seed development is as yet not fully understood and can differ between species. For instance, a couple of auxin interactors including the IAA-glucose hydrolase *TGW6* have been surprisingly shown to negatively impact seed development in *O. sativa* (Ishimaru et al. 2013). The Overproliferation of endosperm in PRC2-impaired mutants of this species has been also revealed to be quite puzzling in terms of PGR accumulations as more CK and less auxin had accumulated in mutants compared to WT (Cheng et al. 2021). These are totally opposite results to what would be expected when compared to *A. thaliana*.

PGR signalling and impact on seed development thus needs to be more thoroughly investigated in order to better understand their versatile roles in seed development in Embryophytes.

Outstanding research questions

The Kinship theory on the resource allocation to the offspring as applied on seed development appears to have some ground on the molecular scale, with some parentally-biased genes having a significant role over development. Parental genome imbalance compared to optimal EBN does indeed results in seed development phenotypes which correspond to what is expected from the contribution of each parents in the light of the kinship theory. Paternal genome excess outcomes on seed development especially suggests the existence of paternally biased genes significantly involved in the control of seed size, which over-representation in the genome causes an endosperm overproliferation. Many genes responding to these characteristics have been discovered and studied so far, especially the iPEG master regulator *PHE1* which is at the source of a cascade of gene activations controlling auxin signalling during seed development. The opposite effect on seed size displayed from maternal excess genome imbalance can also be linked to an excess of maternally-biased genes outside of maternal tissues contribution, however yet not understood in detail.

We could be tempted to impute to parentally biased genes the role of being the molecular representative of a maternal-paternal “arms-race” over seed size, but other results have unfortunately shown us that this is far from being true. Some key parentally imprinted genes indeed have a significant role in seed size determination, but the vast majority of them however appear to have only minor to no role whatsoever in an isogenic context. A significant amount of genes were also reported as critical for seed size development without being parentally biased on either side. Parentally biased phenotypes on seed size development can thus only be imputable to a small selection of genes and parentally-inherited factors, both of them not being representative of the intricate mechanisms of seed development and seed size determination which cannot be reduced to the kinship theory.

It would be interesting to investigate more thoroughly the impact of PGRs-related genes on the development of the different seed tissues during seed development which appears to variate significantly between species. Our understanding of mechanisms governing these signalling pathways thus need to be slightly altered and

optimised to the species in question in order to understand the specific actors influencing seed size in this species. These mechanisms need especially to be thoroughly understood as slight manipulations of their fine tuning seem to end up in endosperm failure-derived seed non-viability.

Chapter 2 : The paternally imprinted gene under Positive Darwinian Selection *PICKLE RELATED 2* contributes to Post-hybrid F1 seed size heterosis in *Arabidopsis thaliana*

Introduction

Nearly 40 years ago the kinship theory was enunciation by Trivers (1974) hypothesizing a conflict between mother and kin on the amount of resources allocated to the kin through its development. This theory was further extrapolated to a maternal-paternal conflict by Haig (2000) arguing that the maternal side usually acts in a way as to equally distribute its resources to all its offspring, whereas the paternal side acts rather in a way as to increase the energy intake by the one offspring resulting from the ovule that it has fertilized. In angiosperms, since the embryo consumes the resources stored within the seed before the germination process, the amount of resources allocated to the offspring is generally correlated to the size of the seed itself. This maternal-paternal "arms-race" has been hypothesized to be one of the drivers of heterosis, defined as a capacity for a hybrid offspring to develop traits which value exceeds the value of the same trait for its parents. The mechanism of heterosis from a genetic point of view is however not yet fully understood (Chen 2013; McKeown et al. 2013; Schnable and Springer 2013). Heterosis can be decomposed in four different intensity levels depending on the offspring output for one given trait: 1) Best parent heterosis (BPH) when the trait value of the hybrid offspring exceeds the one of the best of the two parents; 2) Mid-parent heterosis (MPH) when it exceeds an average value of the two parents; 3) Worst parent heterosis (WPH) when it exceeds the value of the worst of the two parents; 4) Negative heterosis (NH) for the cases of offspring displaying a trait value being worse than the worse of the two parents.

One of the most prominent mechanisms hypothesized to be at the source of paternal/maternal "arms-race" towards seed size determination in the last 20 years has been genomic imprinting (Scott et al. 1998; Haig 2000). However, since there is so far no significant effect of genomic imprinting on seed size in an isogenic background (Wolff et al. 2015), this earlier hypothesis is put to controversy. Interestingly, genomic imprinting seems to have a much stronger impact in a hybrid background as it was reported to be systematically disrupted in abortive hybrid seeds of angiosperms species such as *Solanum licopersicum*, *Silene latifolia* and in the genus *Oriza* (Florez-Rueda et al. 2016; Muyle et al. 2018; Tonosaki et al. 2018). The genomic

imprinting phenomenon in angiosperms happens mostly in the endosperm and to some lesser extent in the embryo from the deposition of parentally-biased epigenetic marks (Reik and Walter 2001; Köhler et al. 2012; Batista and Köhler 2020). The consequence of this is that imprinted genes are differentially expressed depending whether they are maternally-inherited or paternally inherited. As such, a maternally expressed imprinted gene (iMEG) is defined as a gene which expression is biased towards the maternally-inherited allele and a paternal expressed imprinted gene (iPEG) as a gene which expression is biased towards the paternally-inherited allele (Köhler et al. 2012). On one hand the imprinting of iPEGs is usually associated with the deposition of PcG-mediated H3K27me3 with heterochromatic H3K9me2 and CHG methylations as silencing marks on the maternal allele. On the other hand iMEGs are associated with repressive DNA methylation marks on the paternal allele while the maternal allele is maintained expressed as a consequence of a global DEMETER-dependant demethylation in the gametophyte (Köhler et al. 2018; Batista and Köhler 2020). Genes intervening in the establishment of imprinting can also be imprinted themselves, as for instance *MEDEA* (*MEA*), i.e. an iMEG involved in the polycomb repressive complex 2 (Vielle-Calzada et al. 1999), and *PKR2*, i.e. an iPEG encoding a Chromodomain-helicase-DNA-binding 3 (CHD3) chromatin remodeler (Wolff et al. 2011). These genes are known to impact seed development and seed size with mutants of MEA and other PcG genes resulting in an endosperm overproliferation phenotype increasing seed size (Kinoshita et al. 1999; Ohad et al. 1999; Spillane et al. 2000; Köhler et al. 2003a), whereas a mutation of PKR2 was shown to rescue the said overproliferation phenotype and reduces seed size (Huang et al. 2017). This would further suggest an evolution-acquired maternal and paternal control of their own parental imprinting mechanisms and appears to correlate the kinship theory of maternal and paternal conflict over the energy allocation to offsprings (Trivers 1974; Haig 2000). Some of these imprinted genes were recently shown to inexplicably mutate faster than others in the course of evolution (Tuteja et al. 2019) and henceforth were identified as being under positive Darwinian selection (PDS). PDS is defined as the ratio of the number of non-synonymous (amino acid changing) mutations (D_n) over the number of synonymous (silent) mutations (D_s) being strictly superior to one ($D_n/D_s > 1$). An opposite ratio of $D_n/D_s < 1$ is considered highlighting

purifying selection which maintains the protein in a highly conserved state in the course of evolution (Wolff et al. 2011). The fact that genes under PDS tend to evolve faster than average have raised the hypothesis that such genes could be implicated in functions highly valued for trophic competition between individuals, possibly also between mates over energy allocation to the offspring as stated by the kinship theory (Haig 2000). The iMEG *MEA* for instance is known to be under PDS and to dramatically affect the development of the seed in *Arabidopsis thaliana* (Spillane et al. 2007).

In this chapter, I investigate on F1 seed size heterosis in the molecular plant model *A. thaliana* the effect of a subset of iMEGs and iPEGs identified in past reports (Gehring et al. 2011; Hsieh et al. 2011; McKeown et al. 2011; Wolff et al. 2011; Pignatta et al. 2014; Schon and Nodine 2017) and further identified as under PDS (Tuteja et al. 2019).

Materials and Methods

Plant material and growth conditions

Arabidopsis thaliana seeds were surface sterilized with Chlorine gas (3:1 bleach:Hydrochloric acid in a bell jar for one hour). Seeds were germinated on 0.5x Murashige and Skoog (MS) medium (Murashige & Skoog, 1962) containing 1% sucrose and 0.8% agar, and grown in a Percival Tissue Culture cabinet under a 16:8 hr light: dark (21 °C/18 °C) regime (Boyes et al. 2001) until they were transferred to soil (five parts Westland compost [Dungannon, N. Ireland]: 1 part perlite: 1 part vermiculite). Plants were grown in chambers under fluorescent lamps at 200 µmol m^{-2} s^{-1} with the same photoperiod. Plants from a same experiment were all kept at the centre of the same shelf to avoid an effect on seed size of the light discrepancy between shelves as well as between the centre and the edges of the shelves. All Knock-out mutants were in Col-0 background. All crosses were manually done in triplicates at the same developmental stage (5-7 flowers from the main inflorescence from three different plants of the same lineage were emasculated after 4 siliques had appeared on the plant and all used as ovule donor to be manually pollinated by a same pollen donor plant). All *pds* mutant lines and their genotyping primers are described in supplementary table S2.1. Both knockout lines of *PKR2* were previously

published by Aichinger et al. (2009). Mutant line *pkr2-1* (SALK_109423) was genotyped with forward primer 5'-CGTAAAAGCTTCATTTGCGTC-3' and reverse primer 5'-TTGCTCACGACGTATGAGATG-3' and *pkr2-2* (SALK_115303) was genotyped with forward primer 5'-TTCTGATTTTTCAACCGATGG-3' and reverse primer 5'-GGGGAGGAGTATCTGGTGAAG-3'. Presence of the SALK T-DNA was verified using the reverse Lbb1.3 primer 5'-ATTTTGCCGATTTCGGAAC-3'.

Plant DNA extraction

Plant genomic DNA was extracted using 20 mg of rosette leaf which was grinded with glass beads and incubated in DNA extraction buffer [200 mM of Tris-HCl pH 7.5, 250 mM of NaCl, 25 mM of EDTA, 0.5% of SDS] for 10 min at 60°C. Samples were next mixed with equal volumes of ice cold isopropanol (1:1) and DNA was precipitated at -20°C for 10 min followed by a centrifugation. Precipitated DNA was washed once with 70% ethanol, left over to dry, re-suspended in water and incubated at 60°C for 10 minutes.

RNA extraction

RNA was extracted from snapfrozen seeds from at least 30 siliques per replicate collected at 4 DAP according to the protocol described by Allorent et al. (2010), with some modifications. In particular, seeds were homogenised in 2ml Eppendorf tubes using a plastic reusable pestle and mixed with 1 ml RNA extraction buffer as described by Oñate-Sánchez and Vicente-Carbajosa (2008). The protocol continued as described (Allorent et al. 2010). DNase treatment was performed on 1µg of crude RNA using the DNase I amplification grade kit (Invitrogen, UK), followed by clean-up and concentration for RNAseq using RNA clean and concentrator kit from Zymo Research (Cambridge Bioscience, UK). RNA integrity was assessed by agarose gel electrophoresis, and at least 400 ng of RNA per sample were used for mRNA sequencing (Novogene, Cambridge, UK).

F1 Seed size Analysis

F1 seeds from manual crosses were harvested at maturity and left to dry for a week. Dry seeds were imaged with a Scanner at 9 DPI resolution and seed size was measured with ImageJ. Obtained values were trimmed between 0.05 mm^2 and 0.25 mm^2, in order to remove any signal coming from silique debris which might have ended up on the scanning surface.

ANOVA statistical test between samples were performed using R Studio software, followed by a Tukey Post-hoc test using a corrected $\alpha = 0.0001$ in order to take natural variation between cross replicates into account (intra-sample replicates often displaying a seed size significant variation with a p-value comprised between 0.05 and 0.0005).

Preparation for DIC imaging

Siliques at a specific time point after pollination were extracted and fixed overnight in an ethanol and acetic acid solution (9:1) at 4°C. Resulting fixed siliques were then progressively rehydrated in solutions of decreasing ethanol concentration (10 min incubation per step) and rinsed in water. Siliques were cut open and seeds were mounted on a microscopic slide in Hoyers solution (15 ml distilled water, 3.75g gum Arabic, 2.5 ml glycerine and 50g chloral hydrate) and left overnight at 4°C for clearing before being imaged by Differential interfering contrast (DIC) under an Olympus BX51 epifluorescence microscope (Dublin, Ireland). Images were captured with a Leica DFC7000 T camera (Leica microsystems, Ashbourne, Ireland).

RNA sequencing and bioinformatics analysis

DNase treatment was performed on 1µg of crude RNA using the DNase I amplification grade kit (Invitrogen, UK), followed by clean-up and concentration for RNAseq using RNA clean and concentrator kit from Zymo Research (Cambridge Bioscience, UK). RNA integrity was assessed by agarose gel electrophoresis, and at least 400 ng of RNA

per sample were sent for mRNA sequencing to Novogene (Cambridge, UK). RNAseq was performed on two biological replicates per sample, 20 Million reads depth and pair end read. Bioinformatic analysis was performed sing the Galaxy servers (Afgan et al. 2018). Low quality sequences with a PHRED Quality number < 30 were removed using Fastp (Ver. 0.20.1, Chen et al. (2018b)). Resulting high-quality RNA sequences were mapped to reference genome TAIR10.1 using TopHat (Ver. 2.1.1, Kim et al. (2013)). Cufflinks (Ver. 2.2.1.3., Trapnell et al. (2010)) was used to determine gene relative expression using the Fragment per Kilobase per Million (FPKM) method and then to determine Differentially Expressed genes (DEGs) using geometrical method and negative binomial method for dispersion estimation (Poisson law). DEGs were considered as truly differentially expressed with a q-value < 0.05 and a Log2 fold change >2. Venn diagrams and heatmaps were generated using R-studio (Ver. 3.5.1) using the VennDiagram and ViridisLite packages. Gene ontology enrichment was evaluated using the Panther Algorithm (Mi et al. 2013) and dysregulated pathway enrichment with ShinyGAM (version 0.99.5-8-gec900f3, Sergushichev et al. (2016)).
Validation of RNAseq results were done by RT-qPCR. Generation of cDNAs was done on previously treated crude RNA using the Superscript III First Strand Synthesis kit (Invitrogen, UK), according to manufacturer instructions. The RT-qPCR was performed using PowerUp SYBR green Master Mix (Applied BioSystem, UK) and corresponding primers (supplementary table S3.2). Expression of target genes was normalized to *EF-1α* housekeeping gene (Wang et al. 2014) and relative expression was calculated in reference to a control sample using the Livak method. Primers are summarized in supplementary table S2.

Results

Some imprinted genes under PDS show an effect on seed size heterosis

From a total of 6 iMEGs and 23 iPEGs obtained from previously published genomic imprinting work on *A. thaliana* (Gehring et al. 2011; Hsieh et al. 2011; McKeown et al. 2011; Wolff et al. 2011; Pignatta et al. 2014; Schon and Nodine 2017) and recently identified as being under PDS (Tuteja et al. 2019), I report in this chapter the investigation of a potential effect of 2 iMEGs and 10 iPEGs on F1 seed size heterosis

between *A. thaliana* Col-0 and Ler-0 accessions. I crossed 2 iMEG knockout lines (*pds21* and *pds24*) and 10 iPEGs knockout lines (*pds9*, *pds11*, *pds14*, *pds15*, *pds17*, *pds19*, *pds20*, *pds22*, *pds29*, *pds38*) in Col-0 background to WT Ler-0 line as either ovule or pollen donor. All information about *pds* lines and genotyping primers are summarized in the supplementary table S1.

Three iPEG knockout lines, i.e. *pds15*, *pds17* and *pds38,* (figure 2.1A), and one iMEG Knockout line, i.e. *pds24*, (figure 2.1C) have displayed a significant variation of F1 seed size when used as a ovule donor crossed by Ler-0 pollen donor in comparison to WT Col-0 X Ler-0 crosses. In opposition, 4 iPEG Knockout lines, i.e. *pds11*, *pds17*, *pds20* and *pds22* (figure 2.1B) and no iMEG knockout line (figure 2.1D) displayed a significant F1 seed size variation when used as a pollen donor in comparison to WT Ler-0 X Col-0 crosses. One of the iPEG Knock-Out lines, namely *pds17* with a T-DNA insertion in *PICKLE RELATED 2* (*PKR2*) exon number 5, systematically displayed an effect on F1 seed size heterosis in comparison to WT crosses when used as both ovule and pollen donor. Therefore, I decided to further investigate the effect of this gene on F1 seed size heterosis. The seed size of all samples (each made from pooling together the results from three ovule donor plant replicates all crossed by the same pollen donor) displayed a normal distribution. One exception however appears with the cross involving *pds24* mutant line as an ovule donor which tend to display a bimodal distribution with a peak around 0.15 mm^2 and another around 0.175 mm^2. This bimodal distribution however appears to be imputable to variations across plant replicates, with one of the three replicates displaying a large amount of seeds in sizes ranging between 0.16 and 0.20 mm^2, whereas the median value is still centred around 0.15 mm^2 like the other two replicates. This is reflected by the median value of all three replicates pooled together still close to the main distribution peak around 0.15 mm^2. The F1 seed size results from all replicates remain normally distributed. This discrepancy between replicates of the same cross was taken into account by correcting the p-value threshold for the statistical tests to a value of $\alpha = 0.0001$.

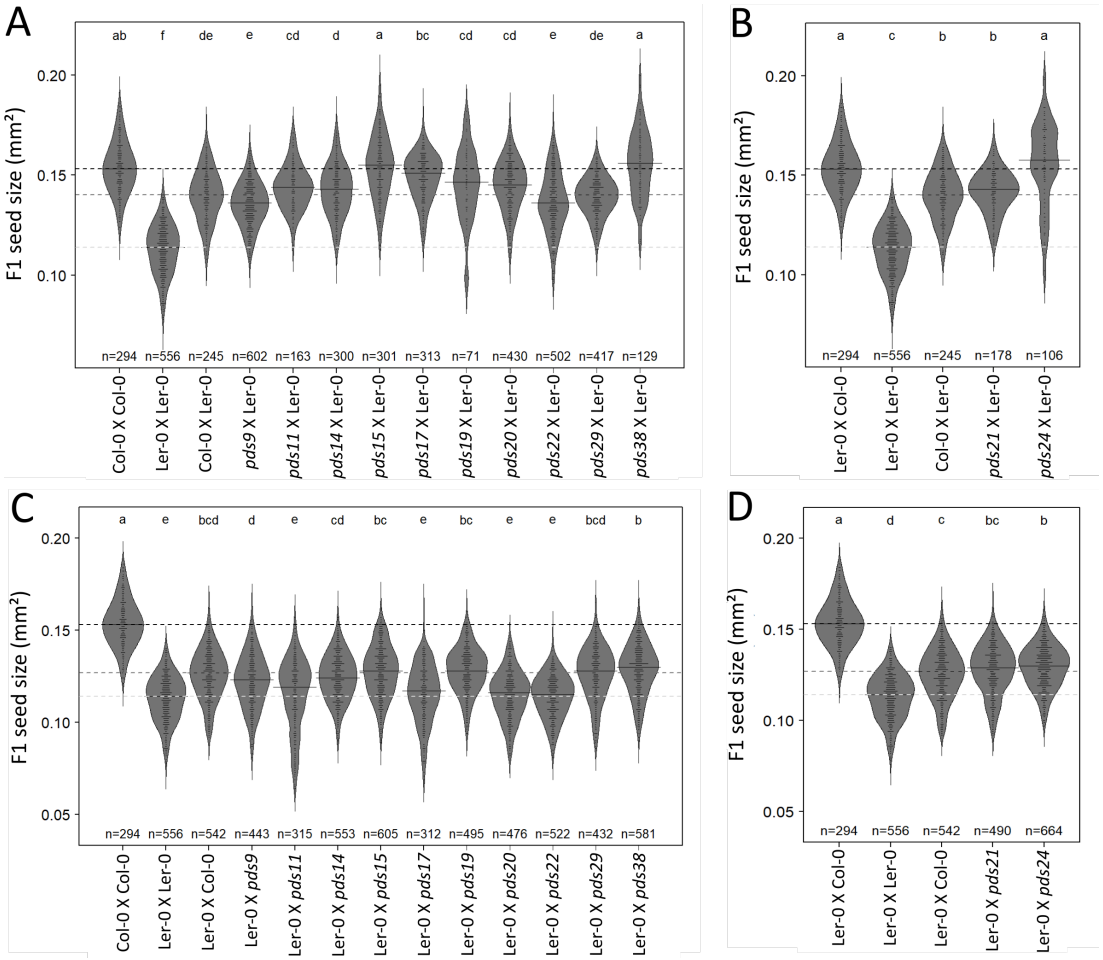


Figure 2.1: Knock-out of some iPEGs and iMEGs under PDS modifies F1 seed size heterosis between Col-0 and Ler-0.

Crossing experiment between Knockout mutant lines of iPEGs and iMEGs in Col-0 background and WT line Ler-0. Bean plot density distribution of F1 seed size from Heterotic crosses between Col-0 and Ler-0 using iPEG (**A** and **C**) and iMEG (**B** and **D**) Knock-Out Col-0 lines as either maternal (**A** and **B**) or paternal parent (**C** and **D**).

Black dashed line: Median F1 seed size for Col-0 X Col-0. Light grey dashed line: Median F1 seed size for Ler-0 X Ler-0. Dark grey dashed line: Median F1 seed size for Col-0 X Ler-0 (**A** and **B**) and Ler-0 X Col-0 (**C** and **D**). Each bean plot represents the pooled results from three ovule donor plant replicates crossed by the same pollen donor. The black bar in each bean plot represent the median value for each pool of three plant replicates.

Both alleles of PKR2 modify seed size in Col-0/Ler-0 and Col-0/Zu-0 hybrid backgrounds

Two different already published knockout T-DNA lines of *PKR2* in Col-0 background, hereafter referred as *pkr2-1* and *pkr2-2* in accordance to past publications (See Material and method section), were used to investigate the effect of *PKR2* on heterosis. Heterotic crosses were performed between two sets of *A. thaliana* accessions: Col-0/Ler-0 (figure 2.2A and 2.2B and Supplementary figure S2.1D and S2.1E) and Col-0/Zu-0 (figure 2.2C and 2.2D and Supplementary figure S2.1F and S2.1G). As above for all samples the seed size results from three ovule donor plant replicates all crossed by the same pollen donor were pooled together to be displayed as beanplots (figure 2.2). Each beanplot displayed a normal distribution at the exception of Ler-0 X Ler-0, Ler-0 X *pkr2-1* and Ler-0 X *pkr2-2* crosses in figures 2.2A and 2.2 B as well as Col-0 X Col-0 and Zu-0 X *pkr2-2* crosses in figure 2.2C and 2.2D, all presenting a double peak distribution. As above this double peak distribution displayed in beanplot pools seems however to be imputable to a discrepancy between the different sample replicates which remain normally distributed. This discrepancy between replicates of the same cross was taken into account by correcting the p-value threshold for the statistical tests to a value of $\alpha = 0.0001$.

The mutation of *PKR2* from the maternal side in all *pkr2-1* X Ler-0, *pkr2-2* X Ler-0, *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 crosses led to a significant reduction of F1 seed size compared to WT crosses Col-0 X Ler-0 and Col-0 X Zu-0 (figures 2.1A and 2.1C). Neither of these mutant crosses displayed a significant F1 seed size difference to WT manually self-pollinated Col-0.

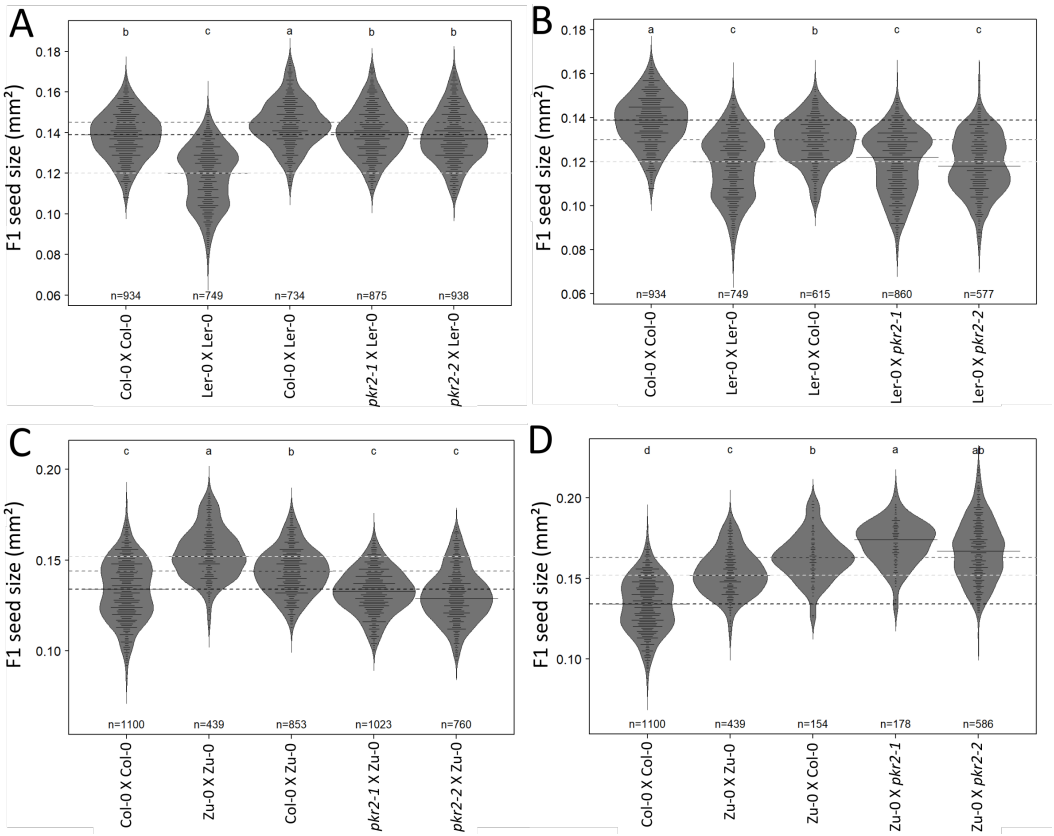


Figure 2.2: Mutation of *PKR2* suppresses paternal contributions in F1 seed size from heterosis crosses.

Heterosis crossing experiment between two Knockout lines pkr2-1 and pkr2-2 and WT Ler-0 and Zu-0. Bean plot of F1 seed size from crosses between *pkr2* mutant lines and WT Ler-0 (**A and B**) and between *pkr2* mutant lines and WT Zu-0 **(C and D)**. F1 Seed size variation resulting from the mutation of either the maternal allele (**A and C**) or the paternal allele (**B and D**) of PKR2.

Mutation of the maternal allele of *PKR2* in Col-0 X Ler-0 (**A**) and paternal allele of *PKR2* in Ler-0 X Col-0 crosses (**B**) shows a decrease of F1 seed size heterosis. Knock-out of maternal allele of *PKR2* in Col-0 X Zu-0 crosses (**C**) also shows a decrease of F1 seed size heterosis but Knock-out of paternal allele of *PKR2* in Zu-0 X Col-0 crosses (**D**) shows no significant F1 seed size variation in both crosses.

Black dashed line: Median F1 seed size for Col-0 X Col-0. Light grey dashed line: Median F1 seed size for Ler-0 X Ler-0 (**A** & **B**) and for Zu-0 X Zu-0 (**C** & **D**). Dark grey dashed line: Median F1 seed size for Col-0 X Ler-0 (**A**), Ler-0 X Col-0 (**B**), Col-0 X Zu-0 (**C**) and Zu-0 X Col-0 (**D**). Each bean plot represents the pooled results from three ovule donor plant replicates crossed by the same pollen donor. The black bar in each bean plot represent the median value for each pool of three plant replicates.

The reciprocal hybrid backgrounds however show different results. On one side the mutation of *PKR2* from the paternal side in both Ler-0 X *pkr2-1* and Ler-0 X *pkr2-2* crosses also display a significant reduction of F1 seed size in comparison to WT Ler-0 X Zu-0 crosses and did not display a significant F1 seed size difference to WT manually self-pollinated Ler-0 maternal parent. On the other side Zu-0 X *pkr2-1* crosses display a statistical F1 seed size increase in comparison to WT Zu-0 X Col-0 crosses while Zu-0 X *pkr2-2* cannot be considered as statistically different.

We also investigated a potential gametophytic or sporophytic effect of *PKR2* on seed size in an isogenic background. No consistent effect could however be observed for both mutant lines (figure 2.3). A significant increase of seed size was indeed observed for manually self-pollinated *pkr2-2* F1 seed size in comparison to Col-0, but no seed size variation could be observed for manually self-pollinated *pkr2-1* (figure 2.3A). Similarly, I could observe a significant F1 seed size increase in *pkr2-2* X Col-0 (figure 2.3B) but not in Col-0 X *pkr2-2* in comparison to WT (figure 2.3C). No F1 seed size variation was observed either in Col-0 X *pkr2-1* (figure 2.3B) or in *pkr2-1* X Col-0 crosses (figure 2.3C) in comparison to WT Col-0. I however observed a difference in size for both *pkr2-1* and *pkr2-2* mutants seeds which appear smaller than WT Col-0 seeds from 4 to 6 DAP for *pkr2-2* and from 5 to 6 DAP for *pkr2-1* (figure 2.3D). No significant difference in size for the embryo was however observed for both mutant lines in comparison to WT Col-0 (figure 2.3E). The variation of seed size between WT and mutants at these stages thus seem to be imputable to endosperm proliferation, however it does not impact final seed size in an isogenic background.

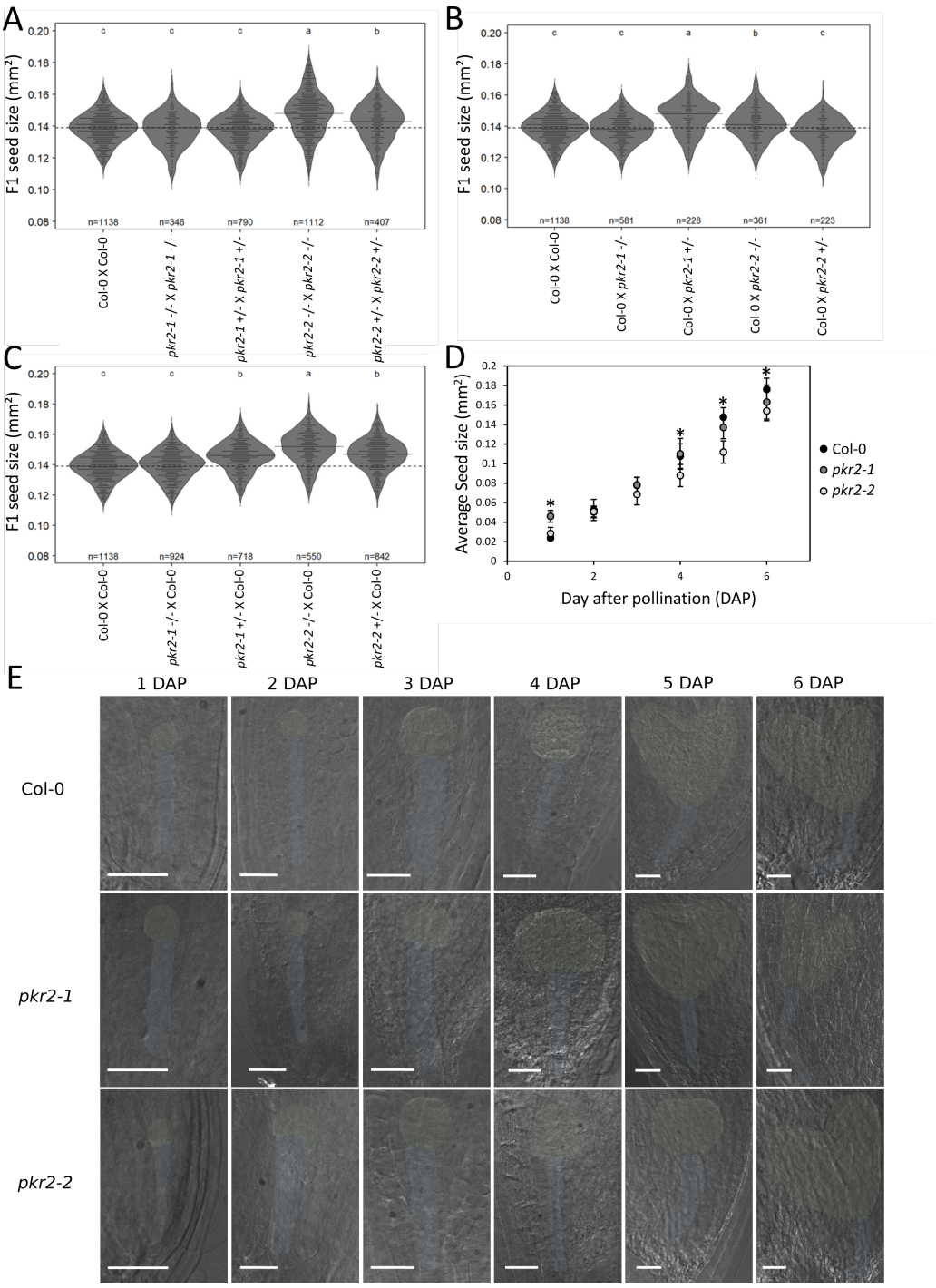


Figure 2.3: The mutation of *PKR2* does not display a consistent significant effect on F1 seed size in Arabidopsis thaliana Col-0 background.

Crossing experiment of knockout lines *pkr2-1* and *pkr2-2* in isogenic (Col-0) background and timecourse experiment of seed development from 1 to 6 days after pollination (DAP). **A.** Homozygous *pkr2* mutants show a F1 seed size decrease compared to WT Col-0. **B.** Paternal allele mutation of *PKR2* shows an F1 seed size increase tendency compared to Col-0 WT. **C.** Maternal allele mutation of *PKR2* shows an F1 seed size decrease compared to Col-0 WT. Each bean plot represents the pooled results from three ovule donor plant replicates crossed by the same pollen donor. The black bar in each bean plot represent the median value for each pool of three plant replicates. **D**. Time course of size of developing seeds (1-6 DAP) from manually selfed mutant lines *pkr2-1* and *pkr2-2* compared to Col-0 WT. Stars indicate timepoints for which both mutant lines are significantly different to WT Col-0 (Student test, p = 0.05). **E**: Time course of embryo growth from 1 to 6 DAP of manually selfed *pkr2-1* and *pkr2-2* compared to WT Col-0. Embryo (green) and suspensor (blue) are highlighted in false colours. White bar represents 25 µm.

Seed size difference between WT and mutants appear around 4 DAP

We performed a time course of seed development from 1 to 6 DAP in Col-0 X Zu-0 background reciprocal crosses and for each timepoint analysed the size of 20 cleared seeds per cross under the microscope. I could observe that seeds from both hybrids involving a mutant line were smaller from 4 to 6 DAP in comparison to WT hybrids Zu-0 X Col-0 and from 5 to 6 DAP in comparison to WT hybrids Col-0 X Zu-0 (figure 2.4A and 4C). F1 seed size of each cross at 5 DAP is represented in beanplots (figure 2.4B and 2.4D). Each beanplot displays a normal distribution with the exception of Zu-0 X *pkr2-2* showing a density which appears bimodal. However since Zu-0 X *pkr2-1* 5 DAP seed size does not display any bimodal distribution this would suggest that the bimodal distribution displayed by Zu-0 X *pkr2-2* 5 DAP seed size would rather be either an artefact resulting from a slightly too small sampling size or a result of the impact of *pkr2-2* background. F1 seed size of Zu-0 X *pkr2-1* and Zu-0 X *pkr2-2* at 5 DAP was similar to WT maternal Zu-0 (figure 2.4D) while F1 seed size of *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 X *pkr2-2* at 5 DAP was smaller than WT maternal Col-0 (figure 2.4B). Differential interference contrast imaging of embryo development further revealed no significant difference between WT crosses and crosses involving a mutant line in terms of embryo development until 6 DAP. I show at 6 DAP WT hybrid seeds to be starting their transition from heart stage to torpedo, while at the same timepoint hybrid seeds involving a mutated allele of *PKR2* are still at the heart stage with apparently smaller cotyledons (figure 2.4E). However this difference appeared to not be consistent and to variate between offsprings obtained from the same silique. Thus with no apparent variation in size of the teguments the variation of seed size at such an early development stage would rather be imputable to the endosperm proliferation. This difference in endosperm proliferation however does not appear to correlate with a difference in final seed size in Col-0 isogenic background the way it does in a hybrid background, suggesting other actors to come in play in addition to the endosperm proliferation for seed size determination.

Following previous data showing *PKR2* as endosperm specific and parentally imprinted in 4 DAP seeds (Wolff et al. 2011; Carter et al. 2016; Huang et al. 2017), as well as the data presented above showing the apparition of a statistical difference in

size between seeds from mutant crosses and WT crosses at that timepoint, I decided to select it for transcriptome analysis in order to understand the impact of a mutation of *PKR2* on the proliferation of the endosperm.

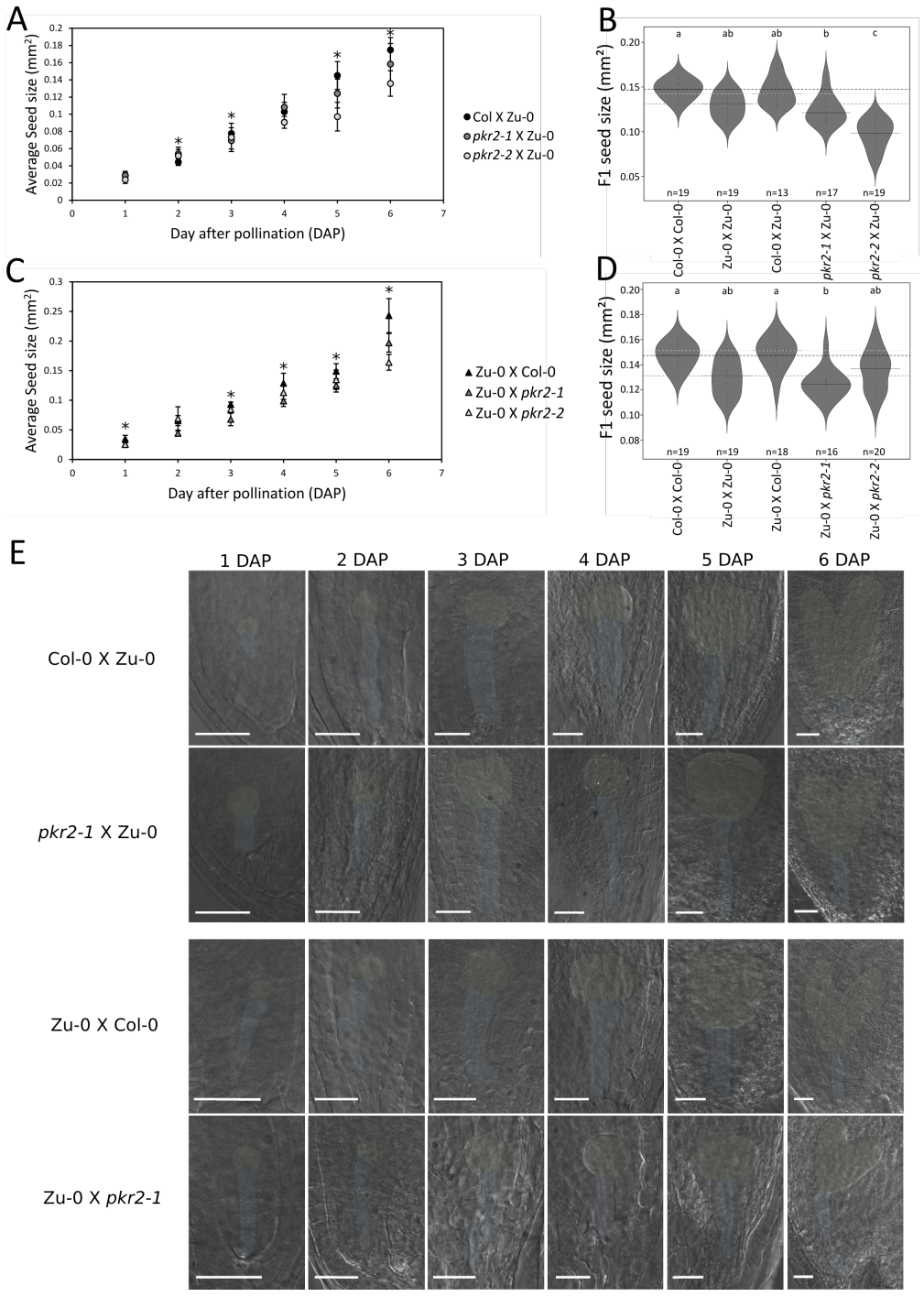


Figure 2.4: Time course of developing seeds in mutated reciprocal Col-0/Zu-0 heterotic crosses. Seed size and embryo development comparison from 1 to 6 DAP between Col-0/Zu-0 heterotic crosses involving Knckout lines pkr2-1 and pkr2-2. **A and C**: Average seed size from 1 to 6 DAP of *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 compared to WT Col-0 X Zu-0 (**A**) and Zu-0 X *pkr2-1* and Zu-0 X *pkr2-2* compared to WT Zu-0 X Col-0 (**C**). Stars highlight timepoints at which the seed size of both crosses involving a KO-*PKR2* line is statistically different from WT control. **B and D**: Density distribution at 5 DAP of seed size of *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 compared to WT Col-0 X Zu-0 (**B**) and of Zu-0 X *pkr2-1* and Zu-0 X *pkr2-2* compared to WT Zu-0 X Col-0 (**D**). WT manually selfed parents Col-0 and Zu-0 are presented for comparison in the first two beans. Statistical ANOVA test and post-hoc Tuckey test displayed as letters above. Different letters represent statistically different samples (p = 0.0001). **E**: Time course of embryo growth from 1 to 6 DAP of *pkr2-1* X Zu-0 and Zu-0 X *pkr2-1* compared to reciprocally Col-0 X Zu-0 and Zu-0 X Col-0. Embryo (green) and suspensor (blue) are highlighted in false colours. White bar represents 25 µm.

PKR2 *is expressed as an iPEG in reciprocal crosses between Col-0 and Ler-0 and between Col-0 and Zu-0*

We generated cDNAs from crude RNA extracted from 4 DAP whole seeds in order to verify the paternal imprinting of *PKR2* in reciprocal crosses between Col-0 and Ler-0 on one hand and Col-0 and Zu-0 on the other hand. The paternal imprinting of *PKR2* in the endosperm was indeed revealed in crosses between Col-0 and Bur-0 at that timepoint by Wolff et al. (2011). However, as previously shown by Wolff et al. (2015), no SNP exist in the *PKR2* CDS sequence between Col-0 and Ler-0 which prevents the direct estimation of relative expression of each maternal and paternal alleles. Three SNPs were predicted in the *PKR2* CDS sequence between Col-0 and Zu-0 by the 1001 genome consortium (magnus.nordborg@gmi.oeaw.ac.at and Consortium 2016) but a verification by sanger sequencing proved these SNPs to not actually be present in the sequence, highlighting the necessity to verify such data from the consortium before using them in technical applications. I could therefore not directly verify the relative expression maternal and paternal alleles of *PKR2* between Col-0 and Zu-0 either. However, I observed a variation of the global expression of *PKR2* in both Col-0/Ler-0 and Col-0/Zu-0 hybrid backgrounds which is visible by RT-PCR (Figure 2.5A) and further verified by RT-qPCR (Figure 2.5B and 2.5C) using *EF-1α* housekeeping gene (reported as one of the most stably transcribed genes in diploid *A. thaliana* by Wang et al. (2014)) to normalize *PKR2* relative expression values. These results correlate the previously established paternal imprinting of *PKR2*. A mutation of the maternal allele indeed results in no significant modification of *PKR2* expression in neither *pkr2-1* X Ler-0 nor *pkr2-2* X Ler-0 in comparison to Col-0 X Ler-0 (verified by RT-qPCR: $p=0.206$, figure 2.5B) nor in *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 in comparison to Col-0 X Zu-0 (verified by RT-qPCR: $p=0.896$, figure 2.5C), whereas the mutation of the paternal allele resulted in a significant 20-fold expression decrease in Ler-0 X *pkr2-1* and Ler-0 X *pkr2-2* in comparison to Ler-0 X Col-0 crosses ($p<0.001$, figure 2.5B) and 3.7-fold expression decrease in Zu-0 X *pkr2-1* and Zu-0 X *pkr2-2* in comparison to Zu-0 X Col-0 ($p= 0.004$, figure 2.5C).

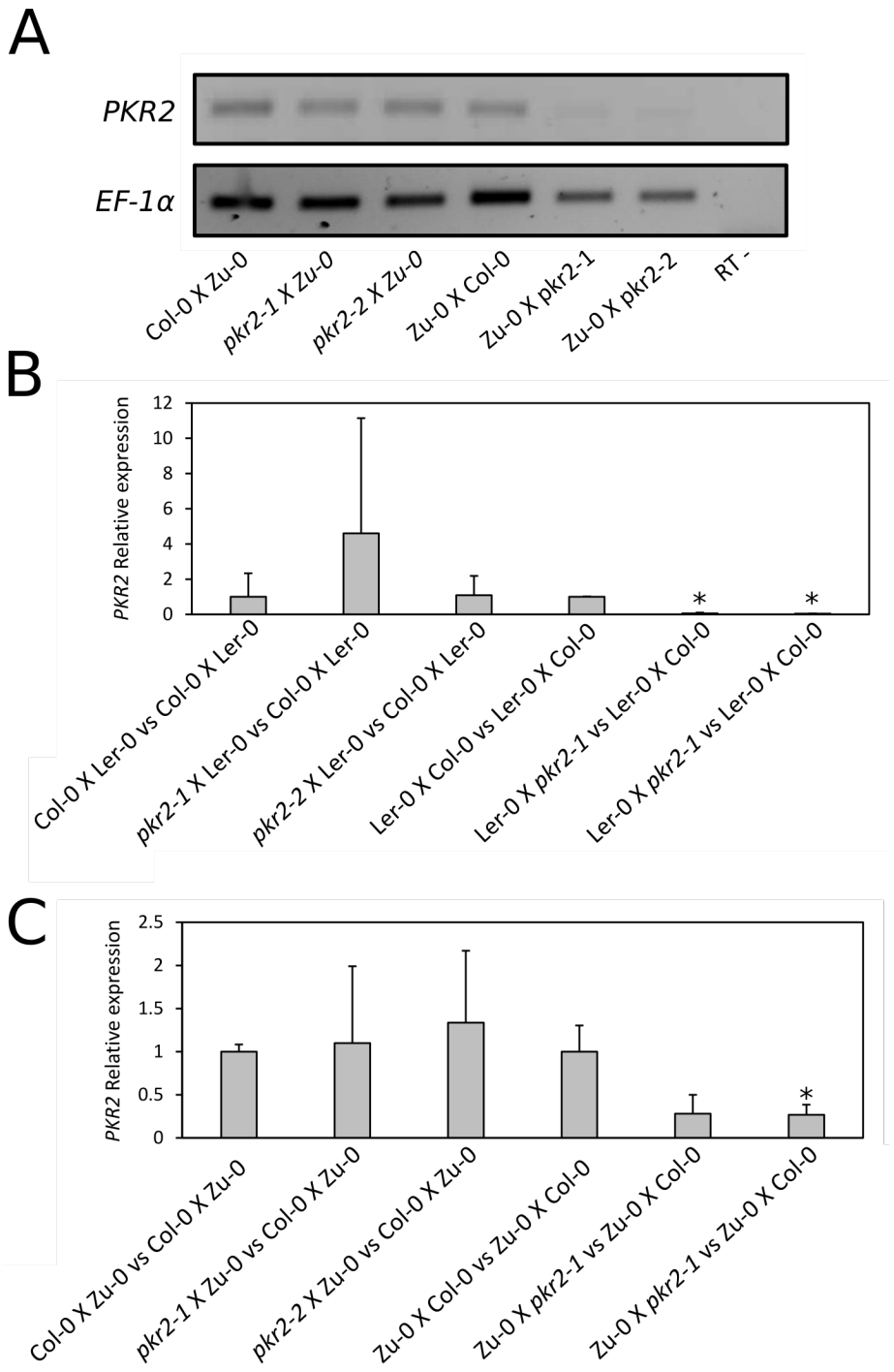


Figure 2.5: *PKR2* expression in mutated Col-0/Ler-0 and in Col-0/Zu-0 crosses.
Evaluation of PKR2 relative expression in 4 DAP seeds from heterotic crosses involving knockout lines pkr2-1 and pkr2-2 using the Livak method. **A**: RT-PCR of *PKR2* expression. *EF1α* is used as a housekeeping control. **B and C**: Fold change of *PKR2* relative expression normalized on *EF-1α* expression in mutant crosses between Col-0 and Ler-0 (**B**) and between Col-0 and Zu-0 (**C**) in comparison to WT crosses. Stars indicate significant fold change according to student test (p = 0.05).

Whole 4 DAP seed transcriptome shows a significant cluster of genes dysregulated by hybridity and reverting to maternal levels in pk2-2 *X Zu-0.*

We performed an RNAseq experiment on 4 DAP whole seed RNA in order to shed some light on the role of a mutation of one allele of *PKR2* on the establishment of post-hybrid seed size heterosis. Low sequences were removed from raw files using fastp and aligned against TAIR10.1 reference genome using Tophat, giving a 96.0% alignment rate for Col-0, 88.6% for Zu-0, 95.2% for Col-0 X Zu-0 hybrids and 88.4% Zu-0 X Col-0 hybrids.

A total of 1284 DEGs were found between WT Col-0 and Zu-0. On one hand RNAseq results reveal a high level of gene dysregulation in both Col-0 X Zu-0 and *pkr2-2* X Zu-0 in comparison to WT Col-0 and between each other (figure 2.6A): 403 DEGs were found between WT Col-0 X Zu-0 and Col-0 (1172 in comparison to Zu-0), 389 between *pkr2-2* X Zu-0 and Col-0 and 550 between *pkr2-2* X Zu-0 and Col-0 X Zu-0. A very low level of gene dysregulation was however found for mitochondrial and chloroplastic genes. On the other hand a much higher level of gene dysregulation was found in Zu-0 X Col-0 and Zu-0 X *pkr2-2* compared to WT Col-0 and Zu-0, with numerous chloroplast dysregulated genes, while a very much lower level of gene dysregulation was found between Zu-0 X *pkr2-2* and Zu-0 X Col-0 (figure 2.6C). I found 1108 DEGs in Zu-0 X Col-0 in comparison to Zu-0 (1596 in comparison to Col-0), 996 DEGs in Zu-0 X *pkr2-2* in comparison to Zu-0 (1303 in comparison to Col-0), and only 95 DEGs between Zu-0 X *pkr2-2* and Zu-0 X Col-0. An important level of gene dysregulation was also found in both mutant crosses compared to *pkr2-2*: 747 DEGs were found in *pkr2-2* X Zu-0, and 1177 were found in Zu-0 X *pkr2-2*.

Gene ontology (GO) analysis in the Col-0 X Zu-0 crossing direction (figure 2.6B) as well as in the Zu-0 X Col-0 crossing direction (figure 2.6D) revealed a significant enrichment of biotic and abiotic stress response GO terms in all DEG subsets. Apart from those GO analysis revealed that the comparison between Col-0 X Zu-0 and Col-0 is enriched genes related to circadian rhythm, transcription, developmental processes and response to hormone while the comparison between *pkr2-2* X Col-0 and Col-0 X Zu-0 showed an enrichment in genes involved in circadian rhythm, cell wall organization, transcription, protein metabolic process and development. The

comparison between *pkr2-2* X Col-0 and Col-0 showed mostly an enrichment in genes involved in cell wall organization. In the opposite direction, the comparison of Zu-0 X Col-0 to Col-0 was enriched in GO terms such as circadian rhythm, cell fate commitment, development, cell wall organization, transcription, nucleic acid metabolic process and response to hormone, while the comparison between Zu-0 X pkr2-2 and Zu-0 X Col-0 only revealed an enrichment in genes involved in circadian rhythm and transcription. The comparison of Zu-0 X *pkr2-2* to Col-0 was similar to the comparison of Zu-0 X Col-0 to Col-0 and was enriched in GO terms such as development, organelle organization and response hormone. An enrichment of cell wall organization genes can also be found in the *pkr2-2* mutant in comparison to WT Col-0. Genes related to Cell wall organization and transcription were also found enriched between Zu-0 and Col-0. These GO terms were also found enriched between *pkr2-2* X Zu-0 and *pkr2-2* X *pkr2-2*, but not between Zu-0 X *pkr2-2* and *pkr2-2* X *pkr2-2*.

In associated to GO terms, I used the ShinyGAM web application to analysed dysregulated pathway enrichments in each DEGs sets. Two major pathways were found enriched in *pkr2-2* X Zu-0 in comparison to Col-0 X Zu-0, namely fatty acid biosynthesis with the upregulation of the alkaline ceramidase encoded by *AT5G58980* (supplementary figure S2.2A), as well as the regulation of cell wall plasticity with the down regulation of pectin lyases such as encoded by *AT1G04680*, *AT1G67750*, *AT3G07850* and *AT3G07010* and the upregulation of pectin methylesterases inhibitors such as encoded by *AT5G51490* (supplementary figure S2.2B). Surprisingly, another pectin methylesterases inhibitor encoded by *AT3G05620* was found downregulated within the same samples. One of the pectin lyases (*AT1G67750*) was also found significantly upregulated as a result of hybridity (supplementary figure S2.2C). The terpenoid biosynthesis pathway was found enriched between *pkr2-2* X Zu-0 and *pkr2-2* with especially the downregulation of the terpenoid cyclase encoded by *AT5G44630* and the terpenoid synthase GGPPS6 encoded by *AT3G14530* (supplementary figure S2.2E). These genes are found upregulated in both *pkr2-2* and Col-0 X Zu-0 in comparison to WT Col-0, and downregulated in *pkr2-2* X Zu-0 in comparisons to both *pkr2-2* and Col-0 X Zu-0.

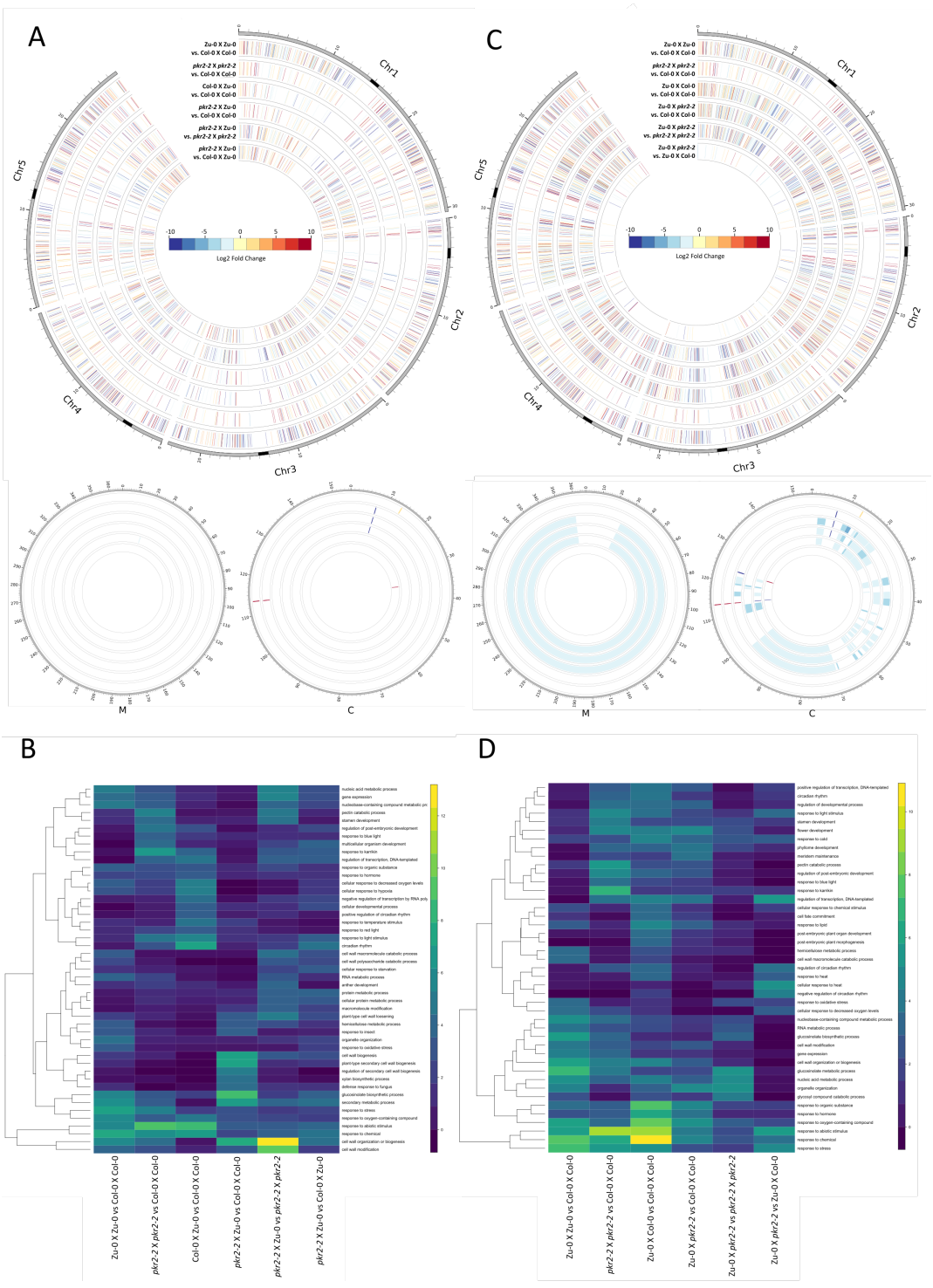


Figure 2.6 : Differentially expressed genes (DEG).
DEGs subsets and GO terms enrichment between transcriptomes of 4 DAP seeds from mutant hybrids, WT hybrids and isogenic Col-0 seeds. **A and C**: Circular genome representation of DEGs observed samples from Col-0 X Zu-0 **(A)** and Zu-0 X Col-0 **(C)** crossing directions. Each bar represent one gene at its position on its chromosome. Colour scale represent the Log2 Expression Fold Change for a DEG between samples. Nuclear genome is displayed on the top, Mitochondrial (M) and Chloroplastic (C) genomes are displayed on the bottom. Nuclear genome is displayed in megabases and mitochondrial and chloroplastic genomes in kilobases. Black bands on nuclear chromosome scales represent centromeres. **B and D:** Heatmap dendrogram of the –Log10 of significant Gene Ontology terms p-value for both Col-0 X Zu-0 **(B)** and Zu-0 X Col-0 **(D)** crossing directions.

The variation of expression of the genes cited above were verified by RT-qPCR in Col-0 X Col-0 and Col-0 X Zu-0 WT crosses as well as in both *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 mutant crosses (Supplementary figure S2.3). Pectin lyase-encoding genes *AT1G04680* and *AT1G67750* were found significantly up-regulated 2.5- and 5.5-fold (p=0.024 and 0.068) respectively in Col-0 X Zu-0 compared to Col-0 as did the terpenoid synthase encoding gene *AT3G14530* (upregulated 2.5-fold, p=0.020). A significant downregulation of the expression of these three genes were further verified in both mutant crosses *pkr2-1* X Zu-0 and *pkr2-2* X Zu-0 in comparison to Col-0 X Zu-0 levels. I thus observed a 2 to 3-fold decrease of *AT1G04680* and *AT3G14530* expression (p=0.017 and 0.024) and a 3 to 8-fold decrease of *AT1G67750* expression (p=0.029). I also confirmed a significant 2 to 3-fold decrease (p=0.040) in the expression of the pectin methylesterase inhibitor-encoding *AT3G05620* within both mutant crosses in comparison to Col-0 X Zu-0, while not showing any significant variation between Col-0 X Zu-0 and Col-0 X Col-0. In comparison, I show a significant 3 to 4.5-fold upregulation (p=0.035) of the pectin methylesterase inhibitor-encoding *AT5G51490* within both mutant crosses in comparison to Col-0 X Zu-0 with no significant variation of expression between Col-0 X Zu-0 and Col-0 X Col-0. However, neither the pectin lyase encoding gene *AT1G07850* nor the terpenoid cyclase encoding gene *AT5G44630* or the alkaline ceramidase encoding gene *AT5G58980* have shown any significant variation of expression between samples by RT-qPCR, differing from our RNAseq results.

Overlap between DEGs was assessed to decipher the presence of genes being dysregulated both by hybridity and by a mutation of *PKR2*. DEGs overlap within the different crossing directions Col-0 X Zu-0 and Zu-0 X Col-0 was analysed separately, and four types of subsets are majorly represented: Zu-0 X Zu-0 vs Col-0 X Col-0 (“WT vs WT” subset), Col-0 X Zu-0 vs Col-0 X Col-0 and Zu-0 X Col-0 vs Col-0 X Col-0 (“Hybrid vs WT” subsets), *pkr2-2* X Zu-0 vs Col-0 X Col-0 and Zu-0 X *pkr2-2* vs Col-0 X Col-0 (“Mutant hybrid vs WT” subsets) and *pkr2-2* X Zu-0 vs Col-0 X Zu-0 and Zu-0 X *pkr2-2* vs Zu-0 X Col-0 (“Mutant hybrid vs hybrid” subsets). In both crossing direction, I can observe a significant number of DEGs being dysregulated in the same direction in hybrid vs WT subsets and WT vs WT subset. I can although observe a more important overlap for the Zu-0 X Col-0 crossing direction (135 downregulated and 106

upregulated genes, figures 2.7D and 2.7E) than for the Col-0 X Zu-0 crossing direction (90 downregulated and 74 upregulated genes, figures 2.7A and 2.7B). A low overlap of genes dysregulated in the same direction was however found between "mutant hybrid vs hybrid" and "hybrid vs WT" subsets. A significant overlap was rather found with genes being oppositely dysregulated in both subsets, especially between DEGs sets of the Col-0 X Zu-0 direction: 82 upregulated DEGs in the *pkr2-2* X Zu-0 vs Col-0 X Zu-0 subset and downregulated in the Col-0 X Zu-0 vs Col-0 X Col-0 subset (figure 2.7A) and 62 downregulated DEGs in the *pkr2-2* X Zu-0 subset and upregulated in the Col-0 X Zu-0 vs Col-0 X Col-0 subset (figure 2.7B). The overlap in the opposite crossing direction was however low with 7 upregulated DEGs in the Zu-0 X *pkr2-2* vs Zu-0 X Col-0 subset and downregulated in the Zu-0 X Col-0 vs Col-0 X Col-0 subset (figure 2.7D), and 12 downregulated DEGs in the Zu-0 X *pkr2-2* vs Zu-0 X Col-0 subset and upregulated in the Zu-0 X Col-0 vs Col-0 X Col-0 subset (figure 2.7E).

Overlapping DEGs displayed above from both crossing direction are represented in two separated heatmap dendrogram (figures 2.7C and 2.7F). In both heatmaps it can observed that the sets of oppositely dysregulated DEGs in "hybrid vs WT" and "mutant hybrid vs hybrid" subsets are not significantly dysregulated between mutant hybrids and WT. More surprisingly, a majority of these DEGs are found dysregulated in the same direction in both hybrids vs Col-0 X Col-0 and *pkr2-2* X *pkr2-2* vs Col-0 X Col-0 DEGs subsets. This overlap of oppositely dysregulated genes is especially enriched in genes involved in transcription and circadian rhythm such as the transcription factors *MYB66* and *MYB90* involved respectively in hypocotyl cell fate determination and anthocyanin production or the MYB family circadian regulation factors *RVE2* and *RVE8*. The pectin lyase encoding gene *AT1G67750* reported above is also found within this overlap, alongside with the alkaline ceramidase gene *AT5G58980* and the pectin methylesterase inhibitor gene *AT5G51490*. Interestingly, not all genes found dysregulated in the same direction in both hybrids vs Col-0 X Col-0 and *pkr2-2* X *pkr2-2* vs Col-0 X Col-0 DEGs subsets are found oppositely dysregulated in "mutant hybrid vs hybrid" DEG subsets. Particularly, one chloroplast gene was found systematically downregulated as a consequence of hybridity, mutation of *PKR2* or both: *psbK* with encodes a subunit of the photosystem II (figure 2.6A, 2.6C, 2.7C and 2.7F).

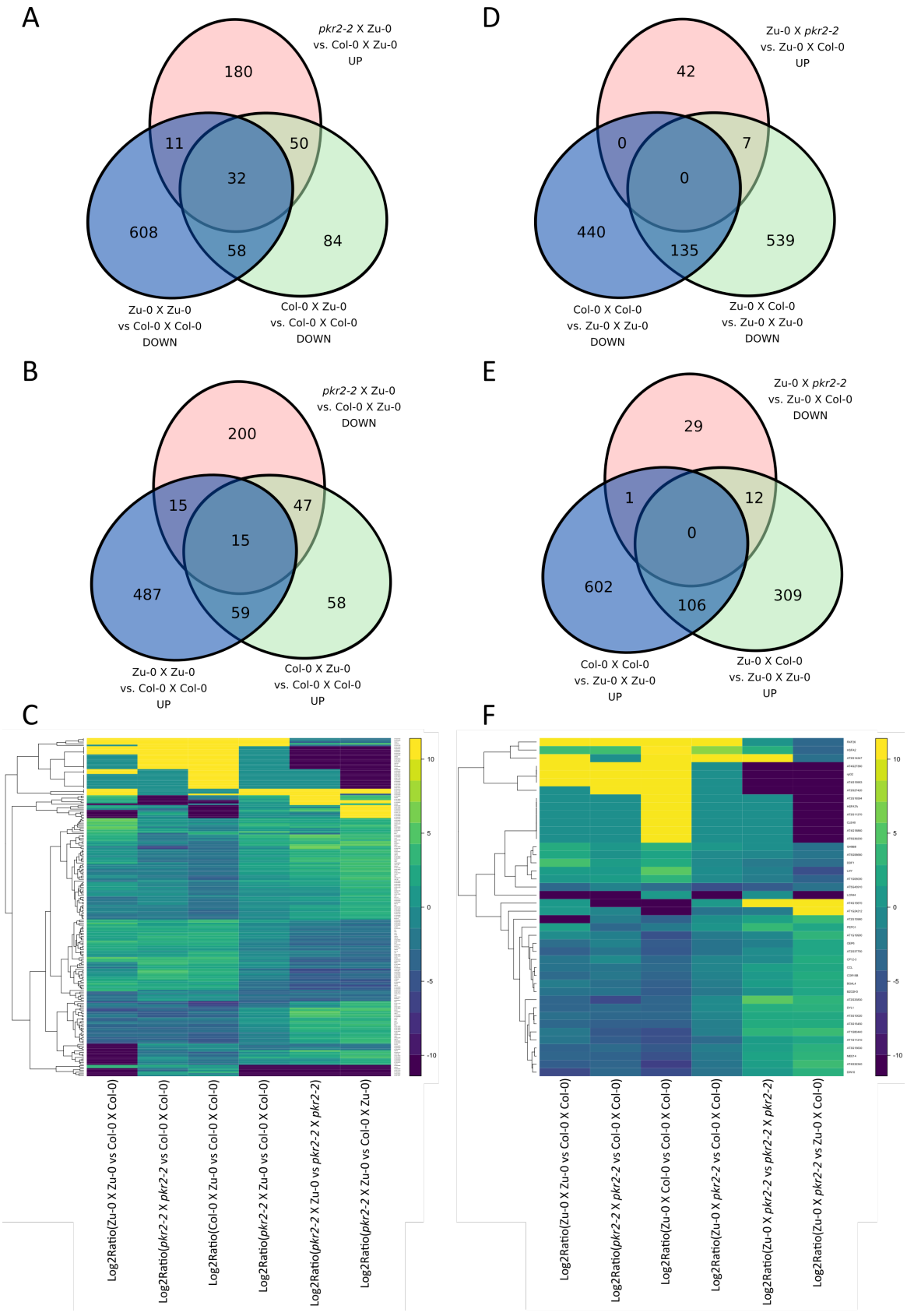


Figure 2.7 : Overlaps of differentially expressed genes (DEG).
Overlap of DEGs subsets between transcriptomes of 4 DAP seeds from mutant hybrids, WT hybrids and isogenic Col-0 seeds. **A, B, D and E**: Venn diagram of Down- **(A and D)** and Up- **(B and E)** regulated DEGs between samples for Col-0 X Zu-0 **(A and B)** Zu-0 X Col-0 crossing directions **(C and D)**. **C and F:** Heatmap dendrogram of overlapping DEGs in Col-0 X Zu-0 **(A)** and Zu-0 X Col-0 **(B)** crossing directions.

Discussion

Only a small subset of the iPEGs and iMEGs used in this chapter appear to display a significant effects on F1 seed size heterosis in a Col-0/Ler-0 heterotic hybrid background. These results are in opposition from previous publications reporting no significant effect of iPEGs on F1 seed size (Wolff et al. 2015). However, these last reports were made from a knockout of iPEGs in an isogenic background, in which I have not found any significant effect either, especially for the iPEG *PKR2* (figure 2.4). I find some variation of seed size in a Col-0 isogenic background but however only with one of the two knockout lines of *PKR2* used in this chapter (*pkr2-2*). This could be attributable to the *pkr2-2* background which might possess one or several unmapped T-DNA insertions in its genome despite the indication that *PKR2* would be the only disrupted gene according to NASC.

Our data show that some genes under parental imprinting can have an impact of F1 seed size in a hybrid context involving heterosis. As genes under parental imprinting are good candidates to justify the kinship theory, it is logical to think that this maternal and paternal “arm’s race” over energy allocation to the embryo would be more severe with mates from different populations, as it is expected to disrupt the relationships between co-adapted linked genomic loci (Haig 2014). This would further explain the apparition of a significant effect of parentally-biased genes on seed development in a hybrid context. Hybridity was indeed shown to create a disruption of co-evolved gene-regulation patterns in tomato (Florence-Rueda et al. 2018). One could advance that these co-evolved gene regulation patterns would especially be disrupted in a self-mating species such as *A. thaliana*.

Our data show a role of *PKR2* in the establishment of F1 seed size post-hybrid heterosis effect. Both *pkr2-1* and *pkr2-2* knockout mutants crossed with either Ler-0 and Zu-0 as a ovule donor, and crossed with Ler-0 as a pollen donor indeed lead to an F1 seed size significantly different to WT crosses but not significantly different to the F1 seed size of the maternal accession. This leads to either an increase of F1 seed size when WT hybrid is displaying MPH (figure 2.1A) or a decrease of F1 seed size when the WT hybrid is displaying BPH (figure 2.2A).

Crosses performed between *A. thaliana* WT accessions Col-0 and Bur-0 has revealed a paternally biased expression of *PKR2* in 4 DAP seeds, and has further demonstrated it as an iPEG with its maternally-inherited allele being maintained repressed by the PRC2 complex (Wolff et al. 2011; Schatlowski et al. 2014). One would thus expect the mutation of the maternal allele of *PKR2* to have no particular effect on seed development, as apparently repressed. However, our data show a reproducible effect of the mutation of the maternal allele of *PKR2* in modifying F1 seed size in a Col-0 X Ler-0 hybrid context (figure 2.2A), further confirmed using an additional heterotic Col-0 X Zu-0 Background (figure 2.2B). In opposition, one would expect the inheritance a mutated paternal allele of *PKR2* to display an effect on F1 seed size since this allele is supposedly maintained activated. Although I indeed observed such an effect in Ler-0 X Col-0 crosses similar to the effect observed in the opposite crossing direction (figure 2.2B), no significant effect is observed in Zu-0 X Col-0 with both mutated lines *pkr2-1* and *pkr2-2*. I do observe a significant seed size increase for Zu-0 X *pkr2-1*, but not for the other mutant line. In the light of these results, one could rightfully hypothesize that *PKR2* expression is not paternally-biased in these hybrid background.

Our transcriptome and RT-qPCR results however seem to correlate a maternal repression and paternal activation of PKR2 in both Col-0/Ler-0 and Col-0/Zu-0 hybrid backgrounds. Further arguments towards the conservation of *PKR2* parentally-biased expression between Col-0 and Ler-0 was also brought up by Huang et al. (2017) using a PKR2::GFP fusion protein. Maternal inheritance of the construct (Col-0 X Ler-0 hybrid background) showed no fluorescence at all whereas paternal inheritance (Ler-0 X Col-0 hybrid background) showed fluorescence restricted in the syncytial endosperm. These results further confirm a PRC2-mediated repression of the maternal allele of *PKR2* in Col-0 X Ler-0 hybrid background as the *fis1* mutation allows fluorescence to be restored in the endosperm.

Analysing the growth of crosses involving mutant lines *pkr2-1* and *pkr2-2* compared to WT crosses further shows that the disruption of *PKR2* seems to hinder the heterosis F1 seed size effect of hybrid crosses. Both Col-0 X Zu-0 and Zu-0 X Col-0 hybrid contexts indeed display a smaller seed size at the same time point of development when either maternal or paternal allele of *PKR2* is mutated. This effect

appears to initiate around 4 DAP and is clearly visible at 5 DAP onwards (figure 2.4A and 4C). Comparison of crosses F1 seed size at 5 DAP reveal that this decrease of size leads mutated seeds to adopt a similar seed size as WT isogenic parents while WT hybrid seeds displays BPH for Zu-0 X Col-0 background to display a smaller seed size than both parents while WT hybrid displays MPH for Col-0 X Zu-0 background (figure 2.4B). *PKR2* is known to have a role in endosperm proliferation as its knockout was shown to partly rescue the endosperm overproliferation phenotype from paternal excess triploid seeds (Huang et al. 2017). Since there is no significant F1 final seed size variation in isogenic mutant lines compared to WT Col-0 this suggest that *PKR2* has probably a minor role on seed size determination in an isogenic background but has a more important effect in case of a strong variation of transcriptome such as occurring in hybrids (Florez-Rueda et al. 2016). *PKR2* however appears to have a significant impact on endosperm proliferation as *pkr2* seeds appear smaller than Col-0 seeds from 4 to 6 DAP in a way not apparently related to the development of the embryo (figure 2.4D and 2.4E). However this effect on endosperm proliferation does not appear to affect final seed size in an isogenic background whereas it results in significant variations of final seed size in a hybrid background. This suggests that hybridity may lead to additional modifications of seed development, possibly a modification of the triggering time of endosperm cellularization, which added to a modification of endosperm proliferation would affect final seed size. The modification of final seed size in *Capsella* has indeed be reported to originate from a modification of endosperm cellularization time (Rebernig et al. 2015).

The results shown by the RNAseq experiment correlates the F1 size variation data observed in Col-0 X Zu-0 reciprocal crosses. I indeed observed an important dysregulation of the transcriptome between *pkr2-2* X Zu-0 and Col-0 X Zu-0 which correlates the observed seed size difference between these samples, while the much smaller effect of the mutation of *PKR2* on transcriptome between Zu-0 X *pkr2-2* and Zu-0 X Col-0 correlates the absence of significant seed size variation between them. That *pkr2-2* X Zu-0 displays an effect on F1 seed size in comparison to Col-0 X Zu-0 despite the fact that I have shown that the mutation of the maternal allele of PKR2 doesn't change *PKR2* global expression in 4 DAP seeds suggests a maternally derived effect. Since *PKR2* encodes a chromatin remodeler it suggests that this maternal

effect could take place as the establishment of different epigenetic marks compared to WT in *pkr2-2* induced by the loss of a functional *PKR2*. These epigenetic marks would then be inherited by the *pkr2-2* X Zu-0 seeds and lead to a different gene expression profile in comparison to Col-0 X Zu-0 seeds. It is important to note at this point that the variation of seed size as well as the dysregulation of 4 DAP seed transcriptome does not seem to be imputable to the mutation of *PKR2* alone but to the combination of hybridity and *PKR2* mutation. Indeed, I show the presence of a cluster of genes which are dysregulated by hybridity in Col-0 X Zu-0 and which expression reverts to maternal levels in *pkr2-2* X Zu-0 but surprisingly these genes are also dysregulated in a similar way in both *pkr2-2* vs Col-0 and Col-0 X Zu-0 vs Col-0 DEG subsets (figure 2.7C). This would suggest that the disruption of *PKR2* and the hybridization mechanism tend to have similar effects on transcriptome regulation, but have an antagonistic effect when combined. It could be explained by a change in the PKR2-dependent epigenetic background in *pkr2-2* which would disrupt the effects of hybridization via post-transcription mechanisms such as small interfering RNAs (siRNAs). Indeed despite not affecting seed size significantly, a knockout mutation of *PKR2* in Col-0 appears to affect an important amount of stress-response genes which action is heavily mediated via siRNAs (Li et al. 2017). This could lead to the cancelling of Col-0 X Zu-0 hybridization effects on gene expression regulations, which would be reset to maternal values after the inheritance of a non-disrupted copy of *PKR2* from paternal Zu-0. Thus maternal inheritance of a non-disrupted *PKR2* seems especially to be a key element in post-hybridity F1 seed size heterosis establishment.

From enriched pathway data it would seem that the in *pr2-2* X Zu-0 might be imputable to a combination of cell wall tightening mechanism as well as a potential reduction of endosperm turgor pressure induced by alkaline ceramidase and a reduction of growth from a decrease of Terpenoid synthase activity. Indeed, pectin lyases (Marín-Rodríguez et al. 2002) and pectin methylesterases (Levesque-Tremblay et al. 2015) activity have been reported act towards a loosening of plant cell walls. Thus it seems logical to observe a decrease in pectin lyases gene expression and an increase in pectin methylesterase inhibitors gene expression in *pkr2-2* X Zu-0 in comparison to Col-0 X Zu-0, as this would be correlated with a tightening of cell walls

preventing seed expansion and thus decreasing final seed size. It could also be linked to a context of enhanced endosperm cellularization process which would lead to an early endosperm cellularization in comparison to WT, thus preventing further seed expansion. This was confirmed by RT-qPCR analysis but I could however not confirm any significant variation of the expression of the alkakine ceramidase-encoding gene *AT5G58980* between sample replicates wish could be due to an important variation between replicates. Terpenoids synthase and cyclase were also reported to participate in the GA biosynthesis pathway (Yamaguchi 2006). GA biosynthetic pathway was previously associated with seed growth as levels of GA peak in the seed during endosperm proliferation (Locascio et al. 2014). In correlation to this, GA metabolism associated genes are found upregulated in paternal excess triploid seeds while downregulated in maternal excess triploid seeds (Batista et al. 2019a). It would thus be logical to associate the observed downregulation of terpenoid biosynthesis genes with a reduction of seed size in *pkr2-2* X Zu-0. These results were also confirmed by RT-qPCR for *AT3G14530* encoding a terpenoid synthase, suggesting overall an effect of *PKR2* on post-hybridity F1 seed size heterosis establishment by promoting a loosening of cell walls and enhancing the synthesis of terpenoids. Lastly, a certain number of circadian cycle related genes appeared dysregulated between samples. This could possibly be due to fact that all samples and replicates were not all harvested in liquid nitrogen at the same time of the day, as the sampling of Arabidopsis thaliana seeds is very time consuming with circa 20 min per replicate. This could have induced some discrepancy in the expression of these genes with some replicates harvested in liquid nitrogen in the morning, and others in the afternoon. Further comparison with previously produced seed/endosperm transcriptome data from a similar timepoint seem in indicate a low impact of hybridization or PKR2 mutation on the mechanisms of endosperm proliferation, with no or poor mention from our data of a modification in the expression of cell cycle or auxin biosynthesis genes, which were previously found to be the most prevalent ones during the different stages of endosperm proliferation (Day et al. 2008). However the mentions to cell wall modifications and stress response appear to correspond to previously reported transcriptome variations reported in intra and inter-species hybrid seeds, which correlate the hypothesis of this chapter of *PKR2* being necessary

in post-hybridity heterosis effect establishement (Burkart-Waco et al. 2013; Alonso-Peral et al. 2017; Fort et al. 2017).

In comparison with previously published results it would appear that *PKR2* has only a small impact on F1 seed size in isogenic conditions, but has a significant impact in a context of significant transcriptome dysregulation such as in hybrids, *pkl* mutants (Carter et al. 2016) or mutants of the FIS complex (Huang et al. 2017). In these cases the mutation of *PKR2* was observed to rescue the seed size/seed development phenotype observed, suggesting that *PKR2* influences seed size determination as a sort of catalyser in response to strong dysregulations of the seed transcriptome.

Conclusion

I have shown in this chapter that some genes under PDS and parentally imprinted in the endosperm of *A. thaliana* have an impact on seed size in a heterotic context. The chromatin regulator *PKR2* especially has an effect on F1 hybrid seed size from either its maternally- or paternally-inherited allele, despite being expressed as an iPEG. The knockout of *PKR2* appears to reduce the speed of endosperm proliferation in either an isogenic or a hybrid context, but this does not seem to have a major impact on F1 final seed size determination in a isogenic context. Dysregulations of genes involved in plant cell wall plasticity probably affecting endosperm cellularization in a timewise fashion as well as genes involved in the terpenoid synthesis pathway appear to be good candidates to explain the variation of F1 seed size in heterotic crosses lacking either a maternal or a paternal allele of *PKR2*, suggesting an effect of PKR2 in the timewise establishment of *A. thaliana* endosperm cellularization.

The precise role of *PKR2* in seed development is however still difficult to understand, but our data correlated with previous studies suggest that the impact of *PKR2* on seed size is highly dependent on transcriptome dysregulations, and may prove to be stronger the more significant these dysregulations are. Being restricted to the Brassicaceae linage, *PKR2* appears be a good study candidate to improve seed size and seed content in agronomic species such as *Brassica rapa*.

Chapter 3 : *PICKLE RELATED 2* is a fast evolving neofunctionalizing gene originating from the duplication of *PICKLE* in Brassicaceae.

Chapter 3: PICKLE RELATED 2 is a fast evolving neofunctionalizing gene originating from the duplication of PICKLE in Brassicaceae.

Introduction

Gene accessibility for transcription by the cell machinery is linked to movements of chromatin condensation and de-condensation and is regulated in a tissue-specific way by epigenetic mechanisms, which are still not well understood today (Batista and Köhler 2020). To our current knowledge, these mechanisms are controlled by two major protein components, named PolyComb (PcG) proteins and Trithorax (Trx)-like Chromodomain-helicase-DNA-binding (CHD) chromatin remodeler family proteins. In plants, the components of the PcG complexes have been described in homology with the PcG proteins identified on *Drosophila melanogaster* (Grossniklaus et al. 1998; Kiyosue et al. 1999; Köhler et al. 2003a; Köhler et al. 2003b; Luo et al. 1999; Ohad et al. 1999). PcG proteins participate in the PcG Repressive Complexes of either type 1 or 2 (PRC1 and PRC2) and it is now hypothesized that both complexes can be recruited independently on gene targets. However, they are also reported to create binding sites for each other in order to strengthen a repressive mark (Mozgova and Hennig 2015).

Trx-like CHD proteins, on the other hand, are ATP-dependant chromatin remodelers which are primarily working as monomers (Ho et al. 2013; Ogas et al. 1999; Zhang et al. 2008; Carter et al. 2016; Carter et al. 2018). They globally possess one or two chromodomains in their sequence depending on the protein (Flanagan et al. 2005) as well as a pair of Plant Homeodomain (PHD) for preferential recognition of H3K4me3 (Wysocka et al. 2006) and tend to act antagonistically from the PcG complexes (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002; Murawska et al. 2008; Aichinger et al. 2011; Jing et al. 2019). For instance, PcG complex genes *SWINGER* (*SWN*) and *EMBRYONIC FLOWER 2* (*EMF2*) were reported as direct targets which expression is repressed by the CHD3 protein PICKLE (PKL) in plant root tissues (Aichinger et al. 2009). The CHD3 family protein (a class of chromatin remodelers members of the subfamily II of CDH remodelers) is composed of 4 members in *Arabidopsis thaliana*: PICKLE (PKL) and PICKLE RELATED 1, 2 and 3 (PKR1, PKR2, PRK3). Among these four proteins, PKR2 is the closest to PKL evolutionary speaking, followed by PKR1 and PKR3 (Ho et al. 2013).

Importantly, that PKR2 is plant-specific since it is not found within CHD3 family proteins in animals suggest that PKL and PKR2 have emerged from a gene duplication event within plants kingdom. However, these genes have been reported to play an antagonistic role on seed size determination in *Arabidopsis thaliana*. On one hand the knockout of *PKL* leads to a significant seed size increase associated with a decreased fertility, while on the other hand the knockout of *PKR2* can rescue the *pkl* seed size and decreased fertility phenotypes despite having no significant effect on seed size *per se* (Carter et al. 2016), suggesting divergent functions for PKL and PKR2 despite being from the same family. The phenotypical effects observed in *pkl* mutants may also be in part imputable to an overexpression of *PKR2* resulting from *PKL* disruption as reported in *A. thaliana* roots by Aichinger et al. (2011). PKR2 was shown to be under positive Darwinian selection (PDS) by Tuteja et al. (2019) with particular sites on the protein having a significantly higher rate of non-synonymous mutations over synonymous ones, but so far PKL hasn't been reported to be particularly fast evolving.

Gene duplication with one of the duplicates being fast evolving while the other one remains under purifying selection is usually an argument towards gene neofunctionalization of the fast evolving duplicate (Gibson and Goldberg 2009; Sandve et al. 2018). In this chapter, I am investigating the evolution of PKR2 and PKL within Embryophytes and studying the hypothesis of a neofunctionalization of *PKR2* in comparison to *PKL* in the Brassicaceae lineage.

Methods

Plant material and growth conditions

Arabidopsis thaliana seeds were surface sterilized with Chlorine gas (3:1 bleach:Hydrochloric acid in a bell jar for one hour). Seeds were germinated on 0.5x Murashige and Skoog (MS) medium (Murashige & Skoog, 1962) containing 1% sucrose and 0.8% agar, and grown in a Percival Tissue Culture cabinet under a 16:8 hr light: dark (21 °C/18 °C) regime (Boyes et al. 2001) until they were transferred to soil (five parts Westland compost [Dungannon, N. Ireland]: 1 part perlite: 1 part

vermiculite). Plants were grown in chambers under fluorescent lamps at 200 µmol m^{-2} s^{-1} with the same photoperiod. Plants from a same experiment were all kept at the centre of the same shelf to avoid an effect on seed size of the light discrepancy between shelves as well as between the centre and the edges of the shelves.

All Knock-out mutants were in Col-0 background: *pkr2-1,* i.e. SALK_109423, *pkr2-2*, i.e. SALK_115303 (Aichinger et al. 2009), *pkl-1* (Ogas et al. 1997) and *pkl-10* i.e. GABI_273E06 (Zhang et al. 2012). Primers used to genotype all four mutant lines are summarised in supplementary table S2.1. All crosses were manually done in triplicates at the same developmental stage (5-7 flowers from the main inflorescence from three different plants of the same lineage were emasculated after 4 siliques had appeared on the plant and all used as ovule donor to be manually pollinated by a same pollen donor plant).

Plant DNA extraction

Plant genomic DNA was extracted using 20 mg of rosette leaf which was grinded with glass beads and incubated in DNA extraction buffer [200 mM of Tris-HCl pH 7.5, 250 mM of NaCl, 25 mM of EDTA, 0.5% of SDS] for 10 min at 60°C. Samples were next mixed with equal volumes of ice cold isopropanol (1:1) and DNA was precipitated at -20°C for 10 min followed by a centrifugation. Precipitated DNA was washed once with 70% ethanol, left over to dry, re-suspended in water and incubated at 60°C for 10 minutes.

RNA extraction and sequencing

RNA was extracted from snapfrozen seeds from at least 30 siliques per replicate collected at 4 DAP according to the protocol described by Allorent et al. (2010), with some modifications. In particular, seeds were homogenised in 2ml Eppendorf tubes using a plastic reusable pestle and mixed with 1 ml RNA extraction buffer as described by Oñate-Sánchez and Vicente-Carbajosa (2008). The protocol continued as described (Allorent et al. 2010). DNase treatment was performed on 1µg of crude RNA using the DNase I amplification grade kit (Invitrogen, UK), followed by clean-up and concentration for RNAseq using RNA clean and concentrator kit from Zymo

Research (Cambridge Bioscience, UK). RNA integrity was assessed by agarose gel electrophoresis, and at least 400 ng of RNA per sample were used for mRNA sequencing (Novogene, Cambridge, UK).

Identification of PKR2 and PKL homolog sequences in embryophytes.

For each gene, an initial search was performed by using the *A. thaliana* sequences as query for a BLASTX search (Mount 2007) in Embryophytes using the Phytozome database (v12.1, Goodstein et al. (2012)). Resulting sequences were compiled and aligned using MUSCLE (version 3.5, Edgar (2004)) on the phylogeny platform. Phylogeny was established using PhyML (version 3.0, Guindon et al. (2010)), and the tree was build using TreeDyn (version 196, Chevenet et al. (2006)). Disorder prediction of PKR2 and PKL in *A. thaliana* was performed using Iupred3 (Erdős et al. 2021).

RNAseq and bioinformatics analysis

RNAseq was performed on two biological replicates per sample, 20 Million reads depth and pair end read. Bioinformatic analysis was performed sing the Galaxy servers (Afgan et al. 2018). Low quality sequences with a PHRED Quality < 30 were removed using Fastp (Ver. 0.20.1, Chen et al. (2018b)). Resulting high-quality RNA sequences were mapped to reference genome TAIR10.1 using TopHat (Ver. 2.1.1, Kim et al. (2013)). Cufflinks (Ver. 2.2.1.3., Trapnell et al. (2010)) was used to calculate gene relative expression using the Fragment per Kilobase per Million (FPKM) method and Differentially Expressed genes (DEGs) were determined using using geometrical method and negative binomial method for dispersion estimation (Poisson law). DEGs were considered as truly differentially expressed with a q-value < 0.05 and a Log2 fold change >2. Venn diagrams and heatmaps were generated using R-studio (Ver. 3.5.1) using the VennDiagram and ViridisLite packages. Gene ontology enrichment was performed using the Panther Algorithm (Mi et al. 2013) and dysregulated pathway enrichment with ShinyGAM (version 0.99.5-8-gec900f3, Sergushichev et al. (2016)).

Validation of RNAseq results were done by RT-qPCR. Generation of cDNAs was done on previously treated crude RNA using the Superscript III First Strand Synthesis kit (Invitrogen, UK), according to manufacturer instructions. The RT-qPCR was performed using PowerUp SYBR green Master Mix (Applied BioSystem, UK) and corresponding primers (supplementary table S3.2). Expression of target genes was normalized to *EF-1α* housekeeping gene (Wang et al. 2014) and relative expression was calculated in reference to a control sample using the Livak method. Primers are summarized in the supplementary table S2.

Results

PKR2 originates from a duplication of PKL at the beginning of the Brassicaceae embranchment.

In order to study the evolution of *PKL* and *PKR2* in land plants, homolog protein sequences of PKL and PKR2 were identified using BLAST within Embyophyte genomes from the Phytozome database. Duplicates were removed and sequences were aligned with PKL sequence from Drosophila melanogaster as outgroup to produce a rooted phylogenetic tree (figure 3.1). Parts of the sequence alignment are shown in supplementary figure S3.1. Three main clads can be identified on this tree: PKL proteins within Brassicaceae, PKL proteins outside of Brassicaceae, and PKR2 proteins within Brassicaceae. This last clade is divided into two groups: PKR2 sequences like in *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Camelina sativa* and *Capsella rubella*, and "PKR2-like" sequences like in *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Eutrema salsugineum* and *Raphanus sativus*. These results further indicate that PKR2 is specific to the Brassicaceae family and seems to originate from a duplication of PKL at the beginning of the Brassicaceae embranchment. Branch size shows that PKR2 is evolving faster than PKL within the Brassicaceae family. Protein sequence alignment shows that in comparison to PKL, PKR2 mostly lacks parts of the PHD zinc finger domain characteristic of the CHD3 family (Supplementary figure S3.1A and S3.2) as well as a small portion of the SNF2-related N-terminal domain related to a predicted loop structure (Supplementary figure S3.1B and S3.2). Specific differences can be spotted on these two locations between PKR2 and PKR2-like clads. While the PKR2

clad lack portions of the PHD domain, the proteins from the “PKR2-like” clad lack this domain entirely. In opposite, the proteins from the “PKR2-like” clad are lacking a smaller portion of the SNF2-related N-terminal domain in comparison to the proteins from the PKR2 clad. Protein disorder prediction with IUPRED3 suggest that this small portion of the SNF2 N-terminal domain corresponds to a high disorder peak region, as this peak is not present in PKR2 (Supplementary figure S3.2).

This prediction of high disorder for the small portion of the SNF2 N-terminal domain present only in PKL correlates with its 3D prediction as a loop structure, suggesting a removal of this structure by positive selection in PKR2 resulting in a local decrease of protein intrinsic disorder. It can also be observed that this loop structure is predicted to be very close to the PKL C-terminal end, despite no clear hydrogen bonds being displayed by the 3D predicted model (Supplementary figure S3.2C). This closeness between the protein C-terminal end and truncated loop structure cannot however be observed in the predicted 3D structure of PKR2 (Supplementary figure S3.2D) and the change in 3D folding of PKR2 C-terminal end in comparison to PKL correlates with a reduction of protein intrinsic disorder around this region (Supplementary figure S3.2A and S3.2B). This further suggests a positive selection for the removal of this loop structure in PKR2 resulting in an overall change of protein 3D folding structure leading to a global decrease of PKR2 intrinsic disorder.

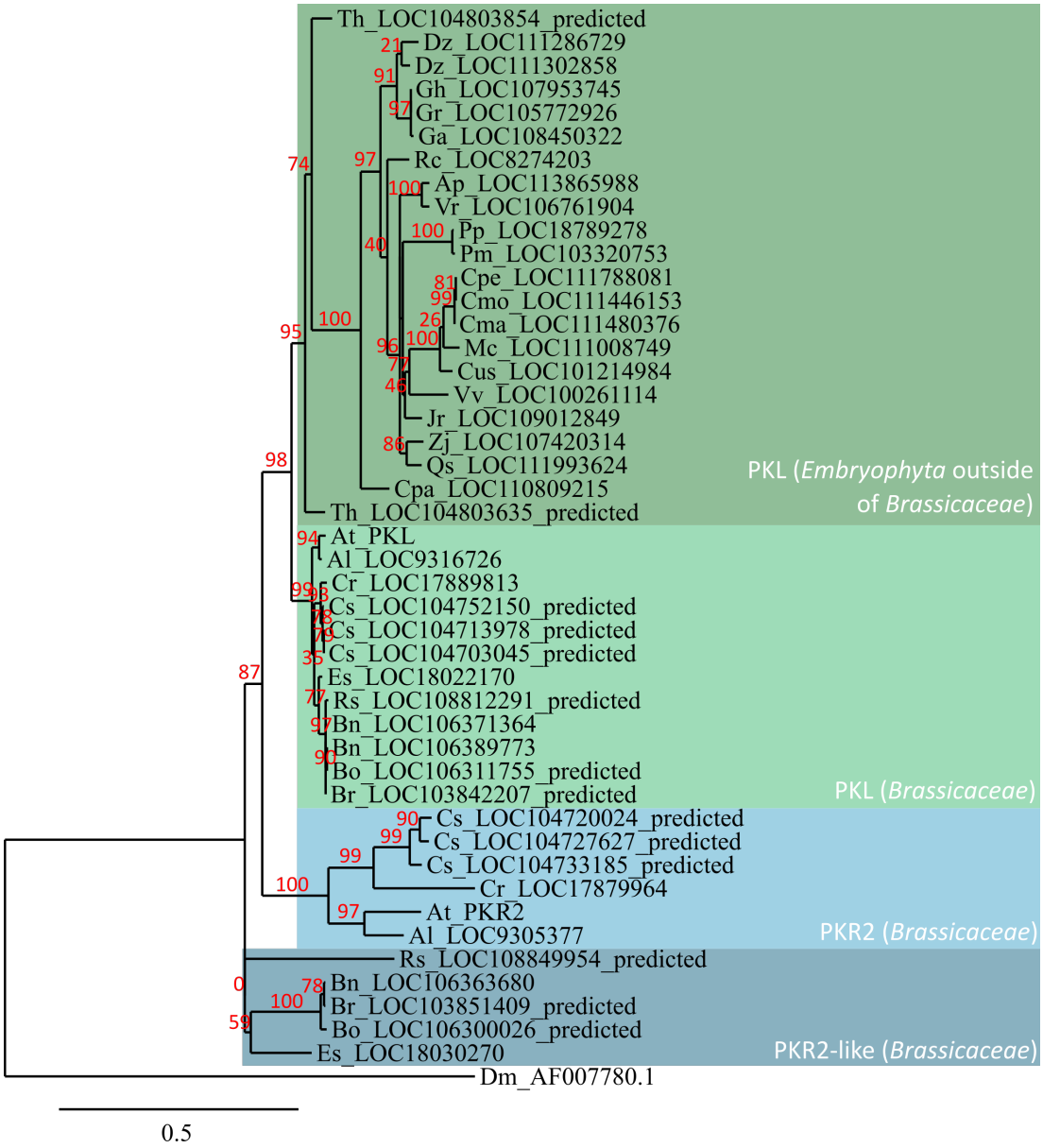


Figure 3.1: **Evolution of PKL and PKR2 in high plants.**
Phylogenetic tree of PKL and PKR2 CHD3 chromatin remodelers in high plants rooted on *Drosophila Melanogaster*. Bootstraps are indicated in red as percentages. Ap: *Abrus precatorius*; At: *Arabidopsis thaliana*; Al: *Arabidopsis lyrata*; Bn: *Brassica napus*; Bo: *Brassica oleracea*; *Br: Brassica rapa*; Cma: *Cucurbita maxima*; Cmo: *Cucurbita moschata*; Cpa: *Carica papaya*; Cpe: *Cucurbita pepo*; Cus: *Cucumis sativus*; Cs: *Camelina sativa*; Cr: *Capsella rubella*; Dm: *Drosophila melanogaster*; Dz: *Durio zibethinus*; Es: *Eutrema salsugineum*; Ga: *Gossypium arboretum*; Gh: *Gossypium hirsutum*; Gr: *Gossypium raimondii*; Jr: *Juglans regia*; Mc: *Momordica charantia*; Pm: *Prunus mume*; Pp: *Prunus persica*; Qs: *Quercus suber*; Rc: *Ricinus communis*; Rs: *Raphanus sativus*; Th: *Tarenaya hassleriana*; Vr: Vigna radiate; Vv: *Vitis vinifera*; Zj: *Ziziphus jujube*.

PKR2 is under Positive Darwinian selection, while PKL is not.

Analysis of PKR2 and PDS amino acid substitution rate ω = Dn/Ds was done using the PAML algorithm. For PKL, 12 sites were selected with ω > 1 but with a probability for positive selection of 0.36 at maximum so they were not considered as truly under PDS. In PKR2, 91 sites were selected with ω > 1, 13 having a probability for positive selection above 0.90. After sequence alignment verification 4 amino acid positions were removed as preceding a deleted/poorly conserved region. The remaining 7 amino acid positions were considered as truly under PDS (Figure 3.2).

One of these amino acid sites is located on a loop at position 39 (N) before the first chromo domain (supplementary figure S3.3A), just before the deleted PHD domain region, according to results obtained using Phyre2 and Alphafold (Kelley et al. 2015; Jumper et al. 2021). This amino acid is not involved in any 3D structure, but is located near an inter-loop hydrogen bond. The two following amino acid sites are located within the second chromo domains at position 169 (R) and 170 (N), both forming a small predicted alpha helix linking two beta sheets (supplementary figure S3.3B). R169 is involved in a double hydrogen bond with D306 and E308 from the nearby alpha helix. N170 does however not seem to be involved in any non-covalent bond. Finally, the remaining four amino acid sites are located within the SNF2-related N-terminal domain at positions 341 (G) and 346 (S) both belonging to a small predicted alpha-helix right after the deleted region from SNF2-related N-ter domain (supplementary figure S3.3C), at 380 (I) belonging to a beta sheet (supplementary figure S3.3D) and 503 predicted to belong to a small alpha helix (P, supplementary figure S3.3E). G341 is involved in hydrogen bonds with M337, V344 and L345 which all participate in the same alpha helix structure. S346 is at the start of the alpha helix and is involved in hydrogen bonds with I342. It is also involved in a hydrogen bond with Q314 from a nearby alpha helix, maintaining cohesion between the two neighbouring helices. I380 is involved in a double hydrogen bond with N245 and V246 and participates in the cohesion of four neighbouring beta sheets. P503 however is not involved in any 3D structure, but is located near an inter-loop triple hydrogen bond between R506 and D818 and G819.

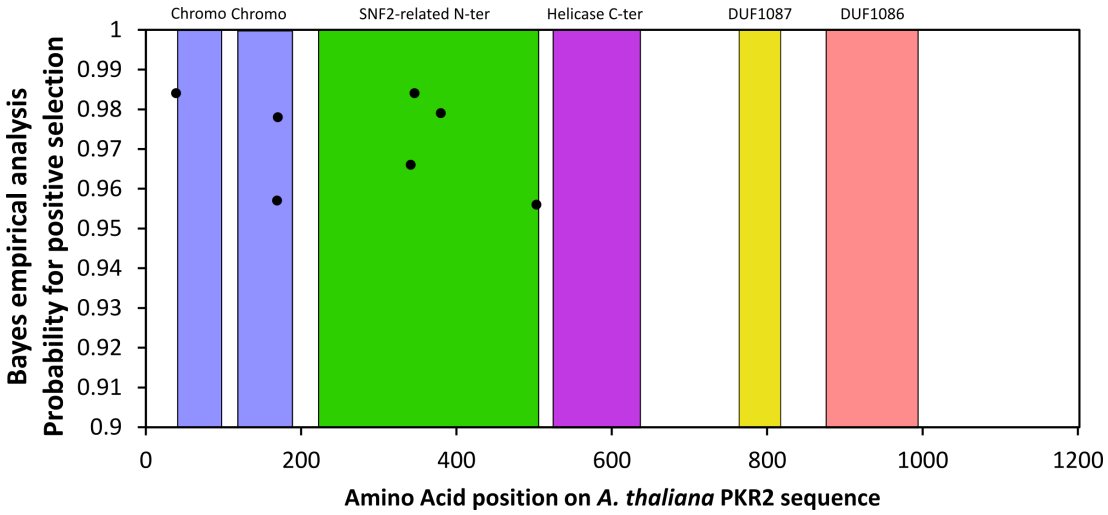


Figure 3.2 : Sites under Positive Darwinian selection (PDS) in PKR2.
Amino acids sites of PKR2 under Positive Darwinian selection. Only amino acid positions considered as truly under PDS are displayed (P>0.90 and not too close from a deleted/poorly conserved region). Each box represent the different domains of the protein as predicted by Pfam.

Despite some redundancy in function, PKL and PKR2 act antagonistically on a specific set of genes.

In order to investigate the different effects of PKL and PKR2 on seed size as well as to further characterise the differences between both proteins, I performed an RNAseq experiment on 4 DAP whole seeds from manually selfed Col-0, *pkr2-2* and *pkl-1* simple mutants and *pkr2-2;pkl-10* double mutant. Sequences were trimmed with fastp and high quality sequences were mapped on TAIR10.1 with Tophat, giving a 96.0% mapping rate for Col-0, 95.5% for *pkr2-2*, 95.9% for *pkl-1* and 95.6 for *pkr2-2*;*pkl-10*. Cuffdiff was used to calculate gene count tables using the FPKM method and sets of DEGs using the geometrical method and negative binomial dispersion estimation method. True DEGs were selected with a q-value < 0.05 and a Log2fold-change of at least 2. A total of 391 DEGs were found between *pkr2-2* and Col-0, 1017 between *pkl-1* and Col-0, 1820 between *pkr2-2*;*pkl-10* and Col-0, 1228 between *pkr2-2*;*pkl-10* and *pkr2-2* and 282 between *pkr2-2*;*pkl-10* and pkl-1 (Figure 3.3A). Interestingly, neither mitochondrial nor chloroplastic transcriptomes appeared significantly impact in both simple mutants *pkr2-2* and *pkl-1* compared to Col-0. However in double mutant *pkr2-2*;*pkl-10* compared to Col-0 or *pkr2-2* a large number of mitochondrial and chloroplastic genes appeared downregulated suggesting a redundant action of *PKL* and *PKR2* on their expression.

Gene ontology (GO) was performed for each DEG set using Panther. All DEGs sets are enriched in GO terms associated to response to biotic and abiotic stresses. In correlation to the more important amount of DEGs found between *pkl-1* and Col-0 in comparison to those found between *pkr2-2* and Col-0 more GO terms appeared in relation to this set as well. While DEGs between *pkr2-2* and Col-0 are enriched in genes related to cell wall organization, development metabolism and transcription, those between *pkl-1* and Col-0 are in addition enriched in genes related to in cell-cell signalling, organelle organization and translation (Figure 3.3B). DEGs found between the double mutant *pkr2-2*;*pkl-10* and Col-0 are enriched in GO terms similar to the ones described above, with the addition of circadian rhythm, photosynthesis, aging and response to hormones. As expected, the comparison between *pkr2-2*;*pkl-10* and *pkr2-2* revealed an enrichment in similar GO terms as between *pkl-1* and Col-0 and comparison between *pkr2-2*;*pkl-10* to *pkl-1* revealed an enrichment in similar GO terms as between *pkr2-2* and Col-0, which shows a similar effect of the knockout of *PKL* and of *PKR2* whether on their own or in association. However the GO terms circadian rhythm, photosynthesis and response to hormone are also found enriched in the DEG sets found between *pkr2-2*;*pkl-10* and *pkr2-2* and between *pkr2-2*;*pkl-10* and *pkl-1* in addition to the ones reported above, which shows that the double mutation has an additional effect in comparison to the sum of each mutation alone. Finally, an significant enrichment in genes involved in DNA repair are especially found between *pkr2-2*;*pkl-10* and *pkr2-2*.

Dysregulated pathways were evaluated for each DEG sets using ShinyGAM (supplementary figure S3.4). Cell wall plasticity was found especially enriched between *pkr2-2* and Col-0 with an upregulation of the pectin lyases-encoding genes *AT1G67750* and *AT3G07010* suggesting an enhanced cell wall loosening. The same pathway was found enriched between *pkl-1* and Col-0 and dysregulated in the same direction with in addition to the two genes mentioned above an upregulation of the pectin lyase-encoding gene *AT3G01270* and a downregulation of *AT2G45220* which codes for a pectin methylesterase inhibitor. Cell wall plasticity was also found dysregulated in *pkr2-2*;*pkl-10* in comparison to Col-0 (upregulation of *AT3G07010* and *AT1G02810*, the latter coding for a pectin methylesterase inhibitor), but a dysregulation of secondary metabolite was also found with the upregulation of

lysine-ketoglutarate reductase-encoding *LKR* and the downregulation of CTP synthase CTPS5-encoding *AT2G34890*. Comparison between *pkr2-2*;*pkl-10* and *pkr2-2* also showed an enrichment in regulation of cell wall plasticity pathway as seen between *pkl-1* and Col-0 with the down-regulation of *AT2G45220* and *AT3G09340* (pectin lyase) and the upregulation of *AT3G10720* (pectin methylesterase inhibitor). Surprisingly I did not find any enrichment in similar pathways between *pkr2-2*;*pkl-10* and *pkl-1* but rather in the fatty acid biosynthesis pathway with the down-regulation of the Alpha/beta hydrolase-encoding gene *AT4G17470* which hydrolyses hexadecanoic acid from proteins.

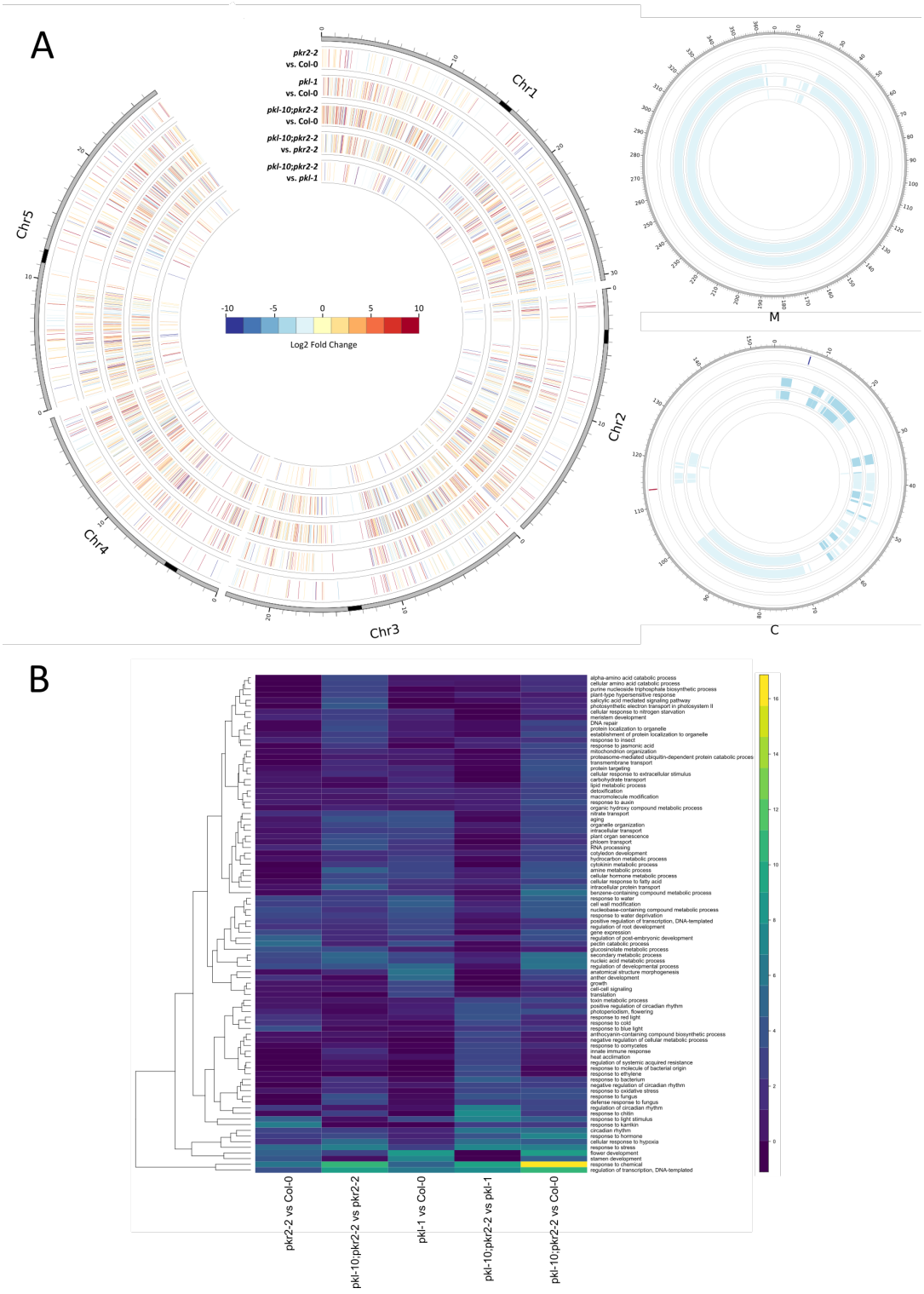


Figure 3.3 : Differentially expressed genes (DEG).

Genomic mapping of DEGs and GO enrichment in comparisons of 4 DAP seeds transcriptomes between pkr2-2, pkl-1, pkr2-2;pkl-10 and WT Col-0. **A**: Circular genome representation of DEGs observed between Col-0, *pkr2-2*, *pkl-1* and *pkr2-2;pkl-10*. Nuclear genome is represented on the large circle on the left, and mitochondrial (M) and chloroplastic (C) genomes on the small circles on the right. Each rectangle represent one gene at its position on its chromosome. Colour scale represent the Log2 Expression Fold Change for a DEG between samples. Nuclear genome is displayed in megabases and mitochondrial and chloroplastic genomes in kilobases. Black bands on nuclear chromosome scales represent centromeres. **B:** Heatmap dendrogram of the –Log10 of significant Gene Ontology terms p-value for DEGs sets between Col-0, *pkr2-2*, *pkl-1* and *pkr2-2;pkl-10*.

Gene targets from dysregulated pathways were validated by RT-qPCR using *pkr2-2*, *pkl-10* and *pkr2-2*;*pkl-10* mutants and WT Col-0 (supplementary figure S3.5). The three pectin lyases encoding genes *AT3G01270*, *AT3G09340* and *AT1G67750* were shown to be increased in expression in mutants in comparison to Col-0*. AT3G01270* displayed a significant 3-fold increase in *pkr2-2* compared to Col-0 (p=0.01), and a 10-fold increase in both *pkl-10* and *pkr2-2*;*pkl-10* compared to Col-0 (p=0.004). *AT3G09340* displayed a 2-fold increase in each mutant although not considered as significant by student t-test (p=0.2). AT1G67750 displayed a 3-fold increase in *pkr2-2* (p=0.132) and a 19 to 22-fold increase in both *pkl-10* and *pkr2-2;pkl-10* mutants (p=0.193). The three pectin methylesterase inhibitors encoding genes were also verified. *AT1G02810* showed a significant 7.5-fold increase in *pkl-10* compared to Col-0 (p=0.001) and a 12-fold increase in *pkr2-2*;*pkl-10* compared to Col-0 (p=0.004) while *AT3G10720* displayed a 11-fold increase in *pkr2-2* vs Col-0 (p=0.06) and a 89-fold increase in both *pkl-10* and *pkr2-2*;*pkl-10* compared to Col-0 (p=0.01). On the contrary but still correlating RNAseq results, AT2G45220 showed a nearly 2-fold decrease in *pkr2-2* (p=0.228), and significant 3-fold decrease in *pkl-10* (p=0.051) as well as a 7-fold decrease in *pkr2-2;pkr-10* (p=0.034). I confirmed no significant variation of expression in *pkr2-2* compared to Col-0 for *LKR* as well as a 25-fold expression increase in *pkl-10* compared to Col-0 (p=0.02). Expression increase of LKR in *pkr2-2*;*pkl-10* compared to Col-0 however only displayed a 17-fold change in comparison (p=0.007). The CTP synthase encoding gene *AT2G34890* was also verified as decreasing significantly in expression 3-fold in *pkl-10* compared to Col-0 and in *pkr2-2*;*pkl-10* compared to Col-0 (p=0.02). A 2-fold decrease was also detected in *pkr2-2* compared to Col-0, however not significant (p=0.10), which correspond to our RNAseq data. Finally, the Alpha/beta hydrolase-encoding gene AT4G17470 was verified as being PKR2-dependant as its expression significantly decreases only in *pkr2-2* and *pkr2-2;pkr-10* in comparison to Col-0 (downregulation 4-fold, p= 0.032) and does not change significantly in *pkl-10*.

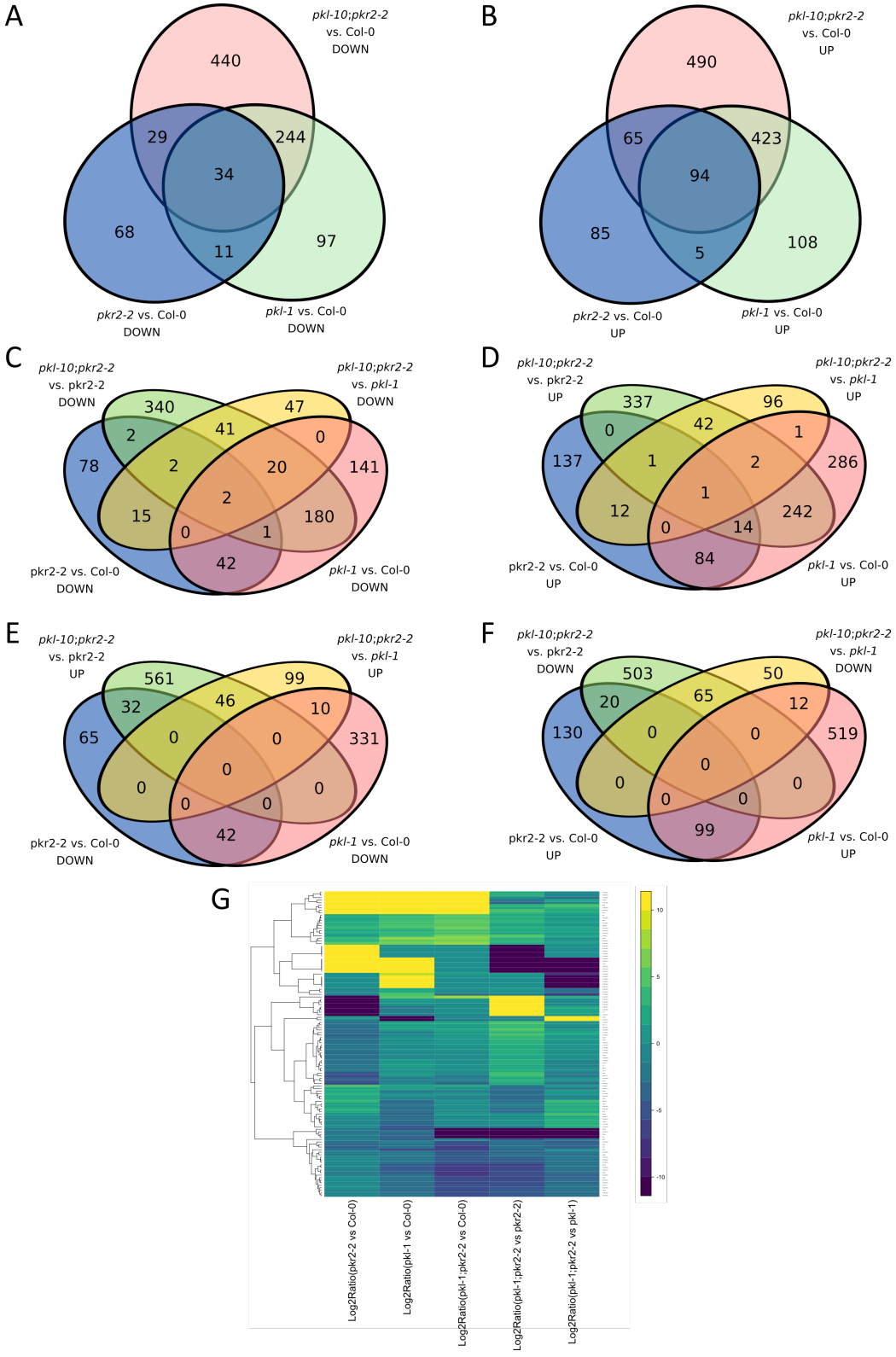


Figure 3.4 : Overlaps of differentially expressed genes (DEG).
Overlaps and clusters of DEGs between the different subsets calculated between 4 DAP seed transcriptomes of *pkr2-2*, *pkl-1*, *pkr2-2;pkl-10* and WT Col-0 4 DAP seeds. **A to F**: Venn diagram of Down- (**A and C**) and Up- (**B and D**) regulated DEGs between *pkr2-2*, *pkl-1* and *pkr2-2;pkl-10* in comparison to WT Col-0 (**A and B**) and between *pkr2-2* and *pkl-1* compared to Col-0 and *pkr2-2;pkl-10* compared to *pkr2-2* and *pkl-1* (**C and D**). **E and F:** Venn diagram of oppositely dysregulated genes between pkr2-2 and pkl-1 compared to Col-0 and *pkr2-2;pkl-10* compared to *pkr2-2* and *pkl-1*. **G**: Heatmap dendrogram of overlapping DEGs between *pkr2-2*, *pkl-1* and *pkr2-2;pkl-10* and WT Col-0.

Overlap between the different DEGs sets was also analysed (figure 3.4). Significant DEGs overlap was found between all three mutants in comparison to Col-0, whether it is down- or up-regulated DEGs (figure 3.4A and 3.4B). A core of 34 down- and 94 upregulated DEGs was especially found between DEGs subsets of three mutants compared to Col-0 and especially enriched in GO terms such as development, response to abiotic stress and cell wall organization (upregulation of pectate lyase-encoding genes *AT59*, *AT1G67750*, *AT3G01270* and *AT3G07010*). Further overlaps were performed between the DEGs sets obtained between *pkr2-2;pkl-10* vs *pkr2-2* and *pk2-2* or *pkr2-2;pkl-10* vs *pkl-1* sets on one side, and between *pkr2-2* vs Col-0 and *pkl-1* vs Col-0 sets on the other side. A significant overlap was found between commonly dysregulated genes coming from these different sets (figure 3.4C and 3.4D). In particular 203 down- and 259 upregulated DEGs were found in common between *pkr2-2;pkl-10* vs *pkr2-2* and *pkl-1* vs Col-0 sets and 19 down- and 14 upregulated DEGs were found in common between *pkr2-2;pkl-10* vs *pkl-1* and *pkr2-2* vs Col-0 sets. Interestingly some overlap was also found between *pkr2-2;pkl-10* vs *pkr2-2* and *pkr2-2* vs Col-0 (7 down- and 16 upregulated DEGs) and between *pkr2-2;pkl-10* vs *pkl-1* and *pkl-1* vs Col-0 (22 down- and 4 upregulated genes). These last two overlaps are composed of similar genes involved as transcription factors, cell wall remodelers and circadian rhythm and photosynthesis associated genes, highlighting a certain redundancy of *pkr2-2* and *pkl-1* mutations. Finally I found a significant overlap of oppositely dysregulated genes between *pkr2-2;pkl-10* vs *pkl-1* and *pkl-1* vs Col-0 on one hand (10 down- and 12 up-regulated DEGs in *pkl-1* vs Col-0 and oppositely dysregulated in the other set) and *pkr2-2;pkl-10* vs *pkr2-2* and *pkr2-2* vs Col-0 in the other hand (32 down- and 20 up-regulated DEGs in pkr2-2 vs Col-0 and oppositely dysregulated in the other set (figure 3.4E and 3.4F). Genes oppositely dysregulated in *pkr2-2;pkl-10* vs *pkr2-2* and *pkr2-2* vs Col-0 sets were especially composed of genes involved in ubiquitination such as *PIT1*, *UBC17*, *PUB24*, *AT1G49210* and *AT1G29800* as well as genes involved as cell wall components such as *AT1G78850*, *AT5G44380* and *AT3G14060.* Genes oppositely dysregulated in *pkr2-2;pkl-10* vs *pkl-1* and *pkl-1* vs Col-0 were rather composed of genes involved in abiotic stress response.

The heatmap dendrogram (figure 3.4G) demonstrates that genes commonly dysregulated between *pkr2-2;pkl-10* vs *pkr2-2* and *pkr2-2* vs Col-0 tend to also be dysregulated in the same direction in *pkr2-2;pkl-10* vs *pkl-1* and *pkl-1* vs Col-0. However, this is true only to some extend since no change in expression below a 2-fold cutoff was considered as significantly dysregulated by cuffdiff. The same can be applied on genes commonly dysregulated between *pkr2-2;pkl-10* vs *pkl-1* and *pkl-1* vs Col-0 in comparison to the two other sets. It can also be observed that oppositely dysregulated genes between either *pkr2-2;pkl-10* vs *pkr2-2* and *pkr2-2* vs Col-0 sets or *pkr2-2;pkl-10* vs *pkl-1* and *pkl-1* vs Col-0 sets don't appear significantly dysregulated in the *pk2-2;pkl-10* vs Col-0 set, suggesting that there are being dysregulated in one direction by either *pkr2-2* or *pkl-1* mutation, but are brought back to WT levels when both mutations are combined. This highlights an antagonistic effect of PKL and PKR2 on the expression of these particular genes.

Discussion

The evolution of PKR2 is quite unique within the CHD3 chromatin remodeler family. The three clades represented by *PKL*, *PKR1* and *PKR3* can indeed be found in both animals and plants, whereas *PKR2* is the only plant specific gene of the family and has been until now classified within the PKL clade for its close resemblance with it despite a lack of PHD domain (Ho et al. 2013). I show here that despite their close resemblance, PKR2 seem to have started to diverge from PKL at the beginning of the Brassicaceae clad formation. The fact that PKR2 is under PDS while PKL is not and that PKL seems to be relatively well conserved in the different *Embryophyta* clades fit the hypothesis of a neofunctionalization of PKR2. Indeed, gene neofunctionalization events are usually associated with a gene duplication followed by one of the duplicate fast diverging while the other remains under purifying selection (Gibson and Goldberg 2009; Sandve et al. 2018). PKL function has indeed proved to be very important in plants for normal cell fate determination through the maintenance of H3K27me3 homeostasis (Ogas et al. 1997; Ogas et al. 1999; Carter et al. 2016; Carter et al. 2018), correlating the apparent necessity for PKL sequence to remain stable. PKR2 function as a chromatin remodeler however has yet remained quite elusive, as

seemingly only having an effect on seed size determination in cases of intense transcriptome dysregulation (Carter et al. 2016; Huang et al. 2017).

The fact that most amino acid positions under PDS in PKR2 sequence are located in both Chromo and in SNF2-related N-terminal domains seems to highlight their importance in PKR2 function. A double chromo domain disposition have been previously shown to lead to a cooperation of both domains to interact with H3 methylated tails (Flanagan et al. 2005). Thus positive selection in chromo domains would suggest an evolution of this cooperative interaction and a selection of new chromatin interactions for PKR2 function in Brassicaceae. This suggests that PKR2 has been positively selected for recognizing a different type of H3 methylated tails in comparison to H3K27me3 which is recognized by PKL (Zhang et al. 2008; Zhang et al. 2012; Carter et al. 2018). Positive selection in the SNF2-related N-terminal domain is however surprising, as this domain possess an ATPase function (Linder et al. 2004; Ho et al. 2013). However I could infer that the fast evolution of these residues are related to the loss of the highly disordered loop structure in the middle of the say domain. Indeed, according to the 3D structure prediction of PKL and PKR2 from the Alphafold Protein Structure database (Jumper et al. 2021), the loss of this loop seems to create an important change of the protein folding, especially the C-terminal part as shown in the supplementary figure S3.2C and S3.2D. In PKL it can be observed that the loop structure seems to be in a close relation with the C-terminal end of the protein (despite the absence of predicted non-covalent bonds). In opposite, the loss of this loop in PKR2 seems to simplify the structure which is no longer in close proximity with the protein C-terminal section. This simplification in PKR2 seems to decrease disorder at that particular location, correlating the suppression of a disorder peak in the IUPRED disorder prediction (Supplementary Figure S3.2B). Furthermore, the apparent release of the C-terminal end of the protein in PKR2 correlates with a smoothening of the predicted disorder signal. Hence, modifications of the PKR2 protein in comparison to PKL seems to go towards a global reduction of disorder within the protein structure.

The sites identified as being under PDS can be found under different configurations within the protein, and apart from N39 and P503 they are all predicted as being

engaged in the enhancement of a 3D structure cohesion by the use of hydrogen bonds. From the previously observed global disorder reduction in PKR2 in comparison to PKL a hypothesis could be emitted that the sites under PDS in PKR2 have been evolving fast as a consequence of a positive selection of a new function for the protein. This might be especially true for the sites positively selected in both chromo domains as this is expected to modify the H3 methylated tails both domains would interact with. The selection of a reduced protein intrinsic disorder would probably be the consequence of the fixation of a newly selected function for the protein. Transition of intrinsic disorder to order within a protein (Disorder-to-order transition) is indeed a hypothetic mechanism which has been raised as an explanation for the subunctionalization or neofunctionalization of a protein after gene duplication events (Ahrens et al. 2016; Ahrens et al. 2017). According to this mechanism some conformations newly obtained by mutations can be stabilized by protein interactions and thus be selected through the course of evolution, followed by a loss of disorder stabilizing the new dominant conformation. Protein intrinsic disorder has been previously described to be a new evolution mechanism which allows the rewiring of protein interactions (Schlessinger et al. 2011; Mosca et al. 2012; Pechmann and Frydman 2014). The reduction of disorder in PKR2 would thus suggest that it is starting to settle in a positively selected new conformation different from the one from PKL in *A. thaliana*.

RNAseq results highlight the fact that the knockout of *PKR2* causes less dysregulation of the transcriptome than *PKL*. This correlates the absence of effect of *PKR2* knockout on seed size in *pkr2* mutants while *pkl* mutants display a significant seed size increase in comparison to WT Col-0 (Carter et al. 2016). Surprisingly, although *pkr2-2;pkl-10* double mutants were reported to rescue the seed size increase phenotype of *pkl-1* mutants, I report here more DEGs in the double mutant than in each simple mutants combined. This suggests a synergistic effect of both *PKL* and *PKR2* on transcriptome and highlights a partial redundancy of both proteins for a same function of gene regulation. This redundancy is further demonstrated with the significant overlap of DEGs between simple and double mutants. Overlapping DEGs however don't represent the majority of DEGs found in each mutants, suggesting a specificity of PKL and PKR2 towards particular cluster of genes. Another surprise can be found on this

RNAseq dataset: whereas previous data reported an upregulation of *PKR2* in A. thaliana roots as a response to a Knockout of *PKL (Aichinger et al. 2011)*, no significant dysregulation of *PKR2* in pkl-1 mutants seeds in comparison to Col-0 can be seen here, suggesting that there might not be any direct regulation of *PKR2* expression by *PKL*, as was previous data suggested. Being a chromatin remodeler PKL was indeed reported to act negatively on the expression of different sets of genes via the deposition and maintenance of H3K27me3 epigenetic marks (Ogas et al. 1999; Zhang et al. 2008; Ho et al. 2013; Carter et al. 2016; Carter et al. 2018), thus a repression of *PKR2* by its ortholog *PKL* has been highly probably so far. A hypothetic reverse interaction of *PKR2* on *PKL* expression also appears to no hold as our dataset and data from the previous chapter do not show a significant modification of *PKL* expression when *PKR2* is knocked out. Any apparent interaction of *PKL* on *PKR2* and vice versa might thus be indirect and highly dependent of the plant tissue, such as the recently reported synergistic effect of PKL and the RNA dependant DNA methylation (RdDM) pathway on gene repression (Yang et al. 2020).

The antagonistic effect of PKL and PKR2 on seed size can also be visible on gene dysregulation between each mutants. A specific antagonistic response between *pkr2-2* or *pkl-1* on one hand and *pkl-2-2;pkl-10* on the other can indeed be observed. Interestingly, this response is composed of DEGs which are specific of either *pkl-1* or *pkr2-2*, but which expression seems to revert to WT Col-0 levels in *pkl-2-2;pkl-10*. The types of proteins mostly represented within this subset being composed of proteins with an ubiquitin-protein transferase activity or components of cell walls would suggest an antagonistic effect of PKL and PKR2 on protein degradation pathways and cell wall organization. These however does not seem to fully explain the variation of seed size observed by Carter et al. (2016), as most of these genes are down-regulated in *pkr2-2* in comparison to Col-0 and re-upregulated in *pkr2-2;pkl-10* in comparison to *pkr2-2*, while they are not significantly dysregulated in *pkl-1* mutant in comparison to Col-0. The portion of the subset reserved to genes dysregulated in *pkl-1* and reverting to WT levels in *pkr2-2;pkl-10* are composed of genes either responding to stress or uncharacterised so far. Thus it is difficult to draw conclusions regarding the effects of genes from this subset on seed size.

However, dysregulated pathways between samples provides more insight on this phenotype. Among 5 pectin lyases found within DEGs, 4 of them have been found up-regulated in mutants in comparison to WT Col-0. In contrast, 3 pectin methylesterase inhibitors were found among DEGs, one being down-regulated in *pkl-1* and *pkr2-2;pkl-10*, while the two others were found up-regulated in *pkr2-2*, *pkl-1* and *pkr2-2;pkl-10*. Pectin lyases and pectin methylesterases activity have previously been linked to a loosening of plant cell walls (Marín-Rodríguez et al. 2002; Levesque-Tremblay et al. 2015). An increase of pectin lyases expression would thus be correlated with the increase of seed size in *pkl-1* as reported by Carter et al. (2016). This expression level of pectin lyases seems maintained in the double mutant, but it can be hypothesised that the global increase of pectin methylesterase inhibitor activity may counterbalance this in promoting cell wall tightening.

An additional mechanism may be added to balance the regulation of cell wall plasticity in *pkr2-2;pkl-10* mutant as a significant decrease of the CTP biosynthesis pathway can be observed. CTP synthases have been reported essential for early endosperm development as regulating nuclei spacing in *Oriza sativa* as their overexpression promotes endosperm development by promoting nuclei replication (Yoon et al. 2021). Among 5 isoforms in A. thaliana, *CTPS2* was reported as being required for complete embryo development and its promoter was characterised as being auxin responsive (Hickl et al. 2021), however no knockout of the other isoform seemed to lead to any particular phenotype so far (Daumann et al. 2018). Among the 5 isoforms, *CTPS5* was the only one being expressed exclusively within the endosperm, but in low levels in comparison to *CTPS2* which seems to be the dominant isoform in this tissue and which expression does not seem significantly different between our samples. The gene encoding for the isoform *CTPS5* (*AT2G34890*) was however found down-regulated in *pkl-10* and *pkr2-2;pkl-10* mutants in comparison to Col-0. A role of *CTPS5* in endosperm development could thus be speculated, even if it is not likely to have an important effect on its own. However its downregulation associated with the dysregulation of the control of cell wall plasticity might help explaining the seed size reduction observed in *pk2-2;pkl-10* in comparison to *pkl-1*. The downregulation of *AT4G17470* encoding an Alpha/beta hydrolase superfamily protein participating as a thioesterase in fatty acid biosynthesis was also observed as

seemingly induced by *pkr2-2* mutation only in both *pkr2-2* and *pkr2-2;pkl-10* mutants. A reduction of thioesterases activity was previously correlated with a decrease of oil content and fatty acid composition in seeds (Moreno-Pérez et al. 2012). This suggest that a reduction of fatty acid accumulation could participate in a decrease of seed size in *pk2-2;pkl-10* in combination with the different dysregulated pathways detailed above. Finally an increase in Lysine-ketoglutarate reductase (*LKR*) expression was also observed among dysregulated pathways. *LKR* was also reported as having a role in the endosperm of *Zea mays* with a peak of expression at 20 DAP (da Silva 1983). It is also involved in the regulation of lysine levels in seeds in *A. thaliana* (Zhu et al. 2001). No effect was however reported on seed size so far. It would appear that *pkl-1* seed size phenotype rescue in *pkr2-2;pkl-10* might not depend explicitly on an antagonistic effect of PKR2 and PKL, but rather by the fluctuation of a balance between different pathways influenced by either *PKR2*, *PKL* or both and seemingly centred around the regulation of cell wall plasticity and control of the endosperm development.

Conclusion

In this chapter I demonstrate that PKR2 originates from a duplication of PKL in the brassicaceae linage and seems to be under a fast evolving neofunctionalization process settling a positively selected newly acquired function for PKR2 leading to it interacting with different H3 methylated tails in comparison to PKL. Some specific amino acid sites seem indeed to be under positive Darwinian selection within important chromatin- and ATP- interacting domains and seemingly guiding PKR2 protein structure towards a reduction of its intrinsic disorder characteristic to fixation of a newly acquired function. I further show that PKL and PKR2 have different effects on *A. thaliana* seed transcriptome with some antagonistic effects on a similar set of genes despite some still apparent functional redundancy. The fluctuating balance between dysregulated pathways influenced by either PKL, PKR2 or both seem to explain previously reported influence of both genes on seed size to some extent.

Functional innovations in PKR2 in comparison to PKL however still remains unexplored, especially in terms of interaction with histone protein and nucleic acids.

Chapter 3: PICKLE RELATED 2 is a fast evolving neofunctionalizing gene originating from the duplication of PICKLE in Brassicaceae.

This would allow a better understanding of the slowly shifting role of the protein on transcriptome regulation as well as its evolutionary advantages in *Brassicacaea*.

Chapter 4 : Cytonuclear and intranuclear hybridity have a cumulative influence on F1 Seed size heterosis in *Arabidopsis thaliana*

Introduction

From the double endosymbiosis event which lead from prokaryotic to eukaryotic lifeforms (Schnepf and Brown 1971; Kutschera and Niklas 2005; Zimorski et al. 2014), mitochondria and chloroplast genomes (both merged, hereafter to be referred as the plasmotype, maternally inherited in *Arabidopsis thaliana*) have established a cross-talk with the nuclear genome (hereafter referred as the nucleotype, both maternally and paternally inherited) of the cell, in order to assure its function. These interactions mostly based on exchanges of metabolites and transcriptional regulation pathways work in two directions: from organelles to nucleus, called retrograde signalling, and from the nucleus to organelles, hereafter anterograde signalling (Bräutigam et al. 2007; Ng et al. 2014). Most protein complexes involved in the chloroplast and mitochondria are made up from subunits encoded in both cytoplasmic and nuclear genomes (Rodermel et al. 1988; Adams and Palmer 2003; Wolf and Brandvain 2014) like most of the mitochondrial and plastid ribosomal proteins. In *Arabidopsis thaliana* most of the genome of the mitochondrion has been duplicated into the chromosome 2 of the nuclear genome (Initiative 2000). Because of these gene transfers, most of the ancestor plasmotype is now located into the nucleotype (Woodson and Chory 2008). In order to function properly, these genes need to be transcribed from the nucleus and the translated protein be sent back to the organelle of origin by the anterograde movement in order to assume its role.

All these interactions, especially between the plasmotype and the maternally inherited nucleotype, inherited together through sexual reproduction, leads to a co-adaptation of these two genotypes and all the molecular complexes they encode within one population (Moison et al. 2010; Sloan et al. 2018; Li et al. 2019a). A cytonuclear disequilibrium or epistasis, i.e. the disruption of the co-evolved relations between plasmotype and nucleotype (Asmussen et al. 1987) has been shown to have an effect on metabolite levels from a significant alteration of pairwise interactions between nuclear loci caused by interactions between cytoplasmic and nuclear genetic loci (Joseph et al. 2013). Such disequilibrium can be interpreted as analogous to linkage disequilibrium happening between nuclear loci (Latta et al. 2001). In Drosophila, it has been suggested to be the cause of decreasing mitochondrial

function happening often in crosses between populations (Sackton et al. 2003), It was also postulated to be at the origin of genetic caste determination between two populations of harvester ants (Linksvayer et al. 2006) and more recently was proposed to be at the origin of the maintenance of genomic imprinting and parent-of-origin gene expression in crosses between populations (Wolf 2009). The existence of such cyto-imprinting interaction effects were further identified in mouse genome using a new statistical model elaborated by He et al. (2014). In some cases it was shown by quantitative trait loci (QTL) analysis of reciprocal pairs of *A. thaliana* recombinant inbred lines that variation in the plasmotype (and its interactions with the nucleotype) can have significant effects on metabolic traits, more powerful than those arising from variations at nuclear loci in some instances (Joseph et al. 2013). Cytonuclear was also shown to be a source of reproductive isolation caused by cytonuclear incompatibility in hybrids of the *Arabidopsis* and *Helianthus* genres (Sambatti et al. 2008; Burton et al. 2013; Muir et al. 2015). It was previously suggested that differences in the plasmotype could trigger heterosis in plants by parent-of-origin effects due to the predominantly maternal inheritance of plasmotype (Duszynska et al. 2013; Ryder et al. 2014). Heterosis here can be defined as an organism trait which exceeds the value of the same trait in its parents (Chen 2013; McKeown et al. 2013; Schnable and Springer 2013). It could have four different levels: Best parent heterosis (BPH) when the trait value of the hybrid offspring exceeds the one of the best of the two parents, Mid-parent heterosis (MPH) when it exceeds the average between the values of two parents, Worst parent heterosis (WPH) when it exceeds the value of the worst of the two parents, and finally negative heterosis (NH) for the cases of offspring displaying a trait value worse than the worse of the two parents. In maize, cytonuclear epistatic QTLs were identified related to both vegetative and reproductive characters, showing the importance of cytonuclear interactions for plant development (Tang et al. 2013). The first cytonuclear epistasis experiments used backcrosses with paternal lines in order to dilute maternally inherited nucleotype into the paternal one (Roux et al. 2016; Boussardon et al. 2019). These so-called cytolines are similar to cytoswap lines, however despite the dilution of the maternally-inherited genome these lines do never harbour 100% of paternal genome. However these studies have provided insights on the effect of cytonuclear

epistasis on plant fitness, later confirmed by Flood et al. (2020) using *A. thaliana* cytoswap lines. This study demonstrated a slightly improved PSII efficiency and biomass dry weight in some cytoswap ecotypes produced with Bur-0, Col-0, Ler-0, C24, Sha, Ws-4 and Ely ecotypes in comparison to their nuclear wild-type parents. It also shows the absence of significant phenotypic difference between WT isogenic lines and cytoswap-generated isogenic ones, confirming the better robustness of the cytoswap procedure compared to cytolines in eliminating maternal genome.

So far, any potential effect of plasmotype on seed size has been poorly investigated. Unlike for other traits in rice where epistasis effects are majoritary, heterozygosity makes a most significant contribution towards seed size (Yu et al. 1997). In *Campanulastrum americanum*, the weight of seeds obtained from cytolines has shown negative deviations from the predicted additive genetic model (Etterson et al. 2007). The absence of maternal effect on seed size was also demonstrated in Chickpea with two genes acting in a dominant epistatic manner to influence seed size (Upadhyaya et al. 2006). Diverse phenotypical traits linked to particular QTLs in wheat (Soltani et al. 2016) and in *A. thaliana* were investigated in the light of a change of plasmotype (Roux et al. 2016; Boussardon et al. 2019; Flood et al. 2020) but no report has been made concerning an on effect of plasmotype on either seed size or weight.

In order to understand how these cytonuclear epistasis can lead to heterosis in plants, the cytoswap technique would seem the most relevant. It consists in “swapping” plasmotypes between ecotypes without changing nucleotype. This could be achieved using a *cenh3* point mutation from the maternal side in dicotyledonous plants like *A. thaliana* (Ravi and Chan 2010; Kuppu et al. 2015; Kelliher et al. 2019). As the CENH3 histone is necessary for kinetochore activity during cellular division (Lermontova et al. 2011; Samel et al. 2012) this point mutation causes miss-segregation of chromosomes during meiosis causing no or an incomplete transmission of the maternal nuclear genome to the offspring (Ravi and Chan 2010). Hence “cytoswapped” offspring only contain the maternally-derived organelle genome and the paternally-derived nuclear genome. In this chapter, I use the *cenh3* point mutation cytoswap technique to study the effects of cytonuclear epistasis on

the development of seeds between four different accessions of *Arabidopsis thaliana*: Col-0, Ler-0, C24 and Zu-0.

Material and methods

Plant growth condition

Arabidopsis thaliana seeds were surface sterilized with Chlorine gas (3:1 bleach:Hydrochloric acid in a bell jar for one hour). Seeds were germinated on 0.5x Murashige and Skoog (MS) medium (Murashige & Skoog, 1962) containing 1% sucrose and 0.8% agar, and grown in a Percival Tissue Culture cabinet under a 16:8 hr light: dark (21 °C/18 °C) regime (Boyes et al. 2001) until they were transferred to soil (five parts Westland compost [Dungannon, N. Ireland]: 1 part perlite: 1 part vermiculite). Plants were grown in chambers under fluorescent lamps at 200 µmol m^{-2} s^{-1} with the same photoperiod. Plants from a same experiment were all kept at the centre of the same shelf to avoid an effect on seed size of the light discrepancy between shelves as well as between the centre and the edges of the shelves. All crosses were manually done in triplicates at the same developmental stage (5-7 flowers from the main inflorescence from three different plants of the same lineage were emasculated after 4 siliques had appeared on the plant and all used as ovule donor to be manually pollinated by a same pollen donor plant).

Generation of cybrid lines

Haploid inducer lines of *Arabidopsis thaliana (L.) Heynh* were generated from the *cenh3-1* mutant lines (Ravi and Chan 2010; Kuppu et al. 2015) in Col-0 background and provided by Comai lab (UC Davis, California, United States). Offspring from heterozygous *cenh3-1* (*cenh3-1 -/-* are completely sterile) were screened for homozygous seedlings by genotyping, and crossed with the ecotype of interest. Sterile haploid offspring unable to produce elongated siliques were screened among the progeny and validated for the ploidy level by visualizing leaf cell nuclear fluorescence intensity using a flow cytometry ploidy analyser PARTEC system

(Sysmex, Lincolnshire, USA) with previously validated diploid Col-0 leaf cells as a control. The obtained haploid plants were grown in the same conditions as reported and seeds resulting from spontaneous genome doubling (phenomenon reported by Ravi and Chan (2010)) were harvested. Diploid plants were grown from these seeds and nucleotype and plasmotype were subsequently confirmed by genotyping using previously published microsatellites markers (Azhagiri and Maliga 2007a; Păcurar et al. 2012; Ravi et al. 2014). The differentiation between the nucleotypes of Col-0 and Zu-0 was performed using the presence of a SNP in the sequence of *PRPL5* (*AT4G01310*) as predicted by the 1001 genomes project (magnus.nordborg@gmi.oeaw.ac.at and Consortium 2016).

Cybrid lines containing the plasmotype of all accession except Col-0 were obtained by crossing a maternal line of the desired accession with a paternal *cenh3-1* +/- line. F1 generation was screened for *cenh3-1* heterozygous mutation (6 plants), and the F2 was screened for a homozygous *cenh3-1* nucleotype (60 plants). This new haploid inducer line was then crossed as mentioned, to obtain a haploid cybrid line. The diploid offspring was confirmed by genotyping. Each developed line is represented as n[maternally-inherited nucleotype/paternally-inherited nucleotype]cPlasmotype: n[C24/C24]cCol refers to the line obtained by the cytoswap method, that possess the maternal and paternal nucleotypes of C24 and the plasmotype of Col-0.

Seed size Analysis and statistical analysis

F1 seeds from manual crosses were harvested at maturity and left to dry for a week. Dry seeds were imaged with a Scanner at 9 DPI resolution and images were analysed with ImageJ. Obtained values were trimmed between 0.05 mm^2 and 0.25 mm^2, in order to remove any signal coming from silique debris which might have ended up on the scanning surface.

ANOVA statistical test between samples were performed using R Studio software, followed by a Tukey Post-hoc test with $\alpha = 0.0001$ in order to take into account the natural variation between different cross replicates (intra-sample replicates often

displaying a seed size significant variation with a p-value comprised between 0.05 and 0.0005).

Preparation for DIC imaging

Siliques at a specific time point after pollination were extracted and fixed overnight in an ethanol and acetic acid solution (9:1) at 4°C. Resulting fixed siliques were then progressively rehydrated in solutions of decreasing ethanol concentration (10 min incubation per step) and rinsed in water. Siliques were cut open and seeds were mounted on a microscopic slide in Hoyers solution (15 ml distilled water, 3.75g gum Arabic, 2.5 ml glycerine and 50g chloral hydrate) and left overnight at 4°C for clearing before being imaged by Differential interfering contrast (DIC) under an Olympus BX51 epifluorescence microscope (Dublin, Ireland). Images were captured with a Leica DFC7000 T camera (Leica microsystems, Ashbourne, Ireland).

Plant DNA extraction

Plant genomic DNA was extracted using 20 mg of rosette leaf which was grinded with glass beads and incubated in DNA extraction buffer [200 mM of Tris-HCl pH 7.5, 250 mM of NaCl, 25 mM of EDTA, 0.5% of SDS] for 10 min at 60°C. Samples were next mixed with equal volumes of ice cold isopropanol (1:1) and DNA was precipitated at -20°C for 10 min followed by a centrifugation. Precipitated DNA was washed once with 70% ethanol, left over to dry, re-suspended in water and incubated at 60°C for 10 minutes.

RNA extraction

RNA was extracted from snapfrozen seeds from at least 30 siliques per replicate collected at 3 DAP according to the protocol described by Allorent et al. (2010), with some modifications. In particular, seeds were homogenised in 2ml Eppendorf tubes using a plastic reusable pestle and mixed with 1 ml RNA extraction buffer as described

by Oñate-Sánchez and Vicente-Carbajosa (2008). The protocol continued as described (Allorent et al. 2010). DNase treatment was performed on 1µg of crude RNA using the DNase I amplification grade kit (Invitrogen, UK), followed by clean-up and concentration for RNAseq using RNA clean and concentrator kit from Zymo Research (Cambridge Bioscience, UK). RNA integrity was assessed by agarose gel electrophoresis, and at least 400 ng of RNA per sample were used for mRNA sequencing (Novogene, Cambridge, UK).

RNAseq and bioinformatics analysis

RNAseq was performed on two biological replicates per sample obtaining 20 million reads depth and pair end read. Bioinformatic analysis was performed using the Galaxy servers (Afgan et al. 2018). Low quality sequences with a PHRED Quality value < 30 were removed using Fastp (Ver. 0.20.1, Chen et al. (2018b)). Resulting high-quality RNA sequences were mapped to reference genome TAIR10.1 using TopHat (Ver. 2.1.1, Kim et al. (2013)). Gene relative expression was evaluated using the Fragment per kilobase per million (FPKM) method and Differentially Expressed genes (DEGs) were determined using Cufflinks (Ver. 2.2.1.3., Trapnell et al. (2010)), using geometrical method and negative binomial method for dispersion estimation (Poisson law). DEGs were considered as truly differentially expressed with a q-value < 0.05 and a Log2 fold change <2. Gene ontology enrichment was evaluated using the Panther Algorithm (Mi et al. 2013) and dysregulated pathway enrichment with ShinyGAM (version 0.99.5-8-gec900f3, Sergushichev et al. (2016)). Venn diagrams and heatmaps were generated using R-studio (Ver. 3.5.1) with the VennDiagram and ViridisLite packages.

Validation of RNAseq results were done by RT-qPCR. Generation of cDNAs was done on previously treated crude RNA using the Superscript III First Strand Synthesis kit (Invitrogen, UK), according to manufacturer instructions. The RT-qPCR was performed using PowerUp SYBR green Master Mix (Applied BioSystem, UK) and corresponding primers (supplementary table S3.2). Expression of target genes was normalized to *EF-1α* housekeeping gene (Wang et al. 2014) and relative expression

was calculated in reference to a control sample using the Livak method. Primers are summarized in supplementary table S2.

Results

Sixteen cybrid lines were generated in *Arabidopsis thaliana* using wild-type (WT) Ecotypes Col-0, C24, Ler-0 and Zu-0 according to the method described (figure 4.1A and 4.1B). These 16 lines contains four isogenic cybrids, i.e. nuclear genome from either Col-0, C24, Ler-0 and Zu-0 and cytoplasmic genomes from the same accession ([Col/Col]cCol, n[C24/C24]cC24, n[Ler/Ler]cLer and n[Zu/Zu]cZu), that were used here as experiment isogenic technical controls. The remaining 12 lines are composed of the reciprocal cybrids for each combination between the four accessions: Col-0/C24 (n[Col/Col]cC24 and n[C24/C24]cCol), Col-0/Ler-0 (n[Col/Col]cLer and n[Ler/Ler]cCol), Col-0/Zu-0 (n[Col/Col]cZu and n[Zu/Zu]cCol), C24/Ler-0 (n[C24/C24]cLer and n[Ler/Ler]cC24), C24/Zu-0 (n[C24/C24]cZu and n[Zu/Zu]cC24) and Ler-0/Zu-0 (n[Ler/Ler]cZu and n[Zu/Zu]cLer). Each cybrid line was successfully genotyped using microsatellites and SNPs that are able to make the distinction between two nucleotypes and between two plasmotypes (Supplementary table S4.1 and figure 4.2).

These 16 cybrid lines were then grown and manually crossed under the same growth conditions in order to obtain seeds of all the following: Isogenic lines (manually self-crossed isogenic cybrids), Cybrids (manually self-crossed non-isogenic cybrids), Hybrids (reciprocally crossed isogenic cybrids, figure 4.1C), and Hybrid-cybrids (non-isogenic cybrids back-crossed with their plasmotype-of-origin isogenic cybrid, figure 4.1D).

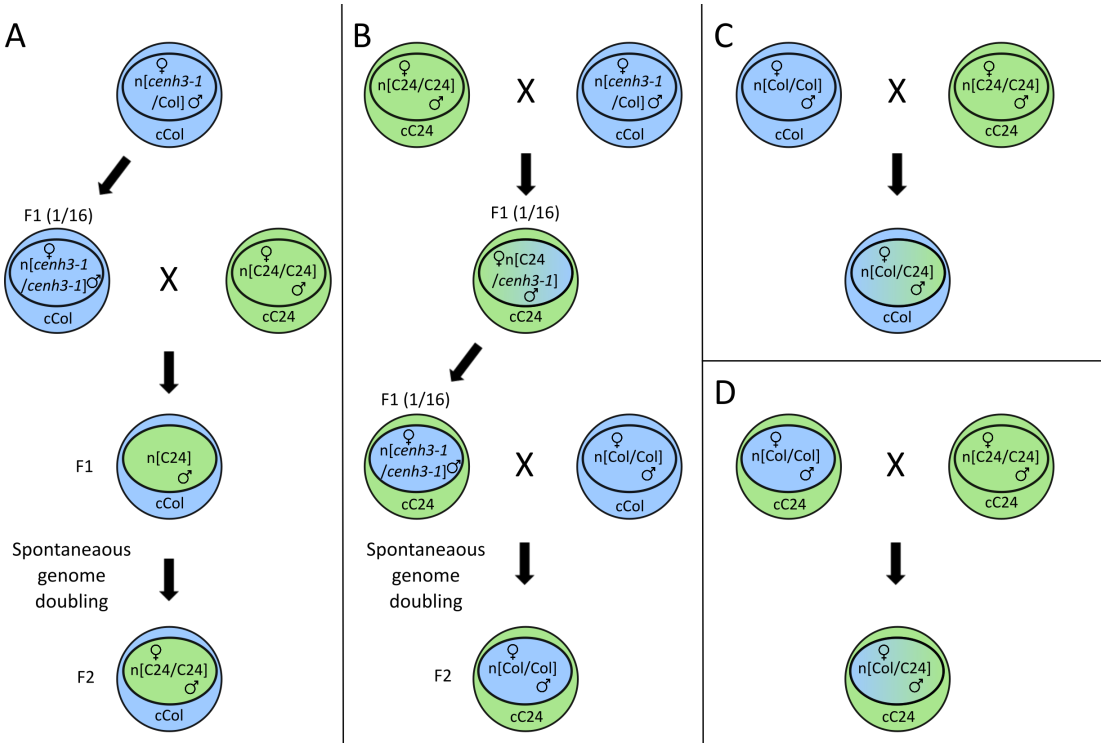


Figure 4.1: Haploid induction and crosses
Crossing procedure for cytoswap lines from a *cenh3* point mutation in Col-0 background **A.** Generation of a stable cytoswap line containing the Col-0 plasmotype from a propagating heterozygous *cenh3* line in Col-0 background. **B.** Generation of a stable cytoswap line containing any other plasmotype from a propagating heterozygous *cenh3* line in Col-0 background. **C.** Generation of a stable hybrid line. **D.** Generation of a stable hybrid-cybrid line from a cybrid line.

Cybridity and hybridity have a cumulative effects on F1 seed size

F1 final seed size was measured on the offspring from three biological replicates consisting of three different ovule donor plants from the same line crossed by the sample pollen donor. The results from these three biological replicates were then pooled together and represented as one beanplots per sample as on figure 3.3. All samples displayed a normal distribution centred on their median value, with the exception of n[C24/Col]cC24, n[Ler/Ler]cCol, n[Ler/Col]cCol, n[Zu/Col]cCol, n[C24/Ler]cC24, n[C24/Ler]cLer, n[C24/Zu]cZu, n[Zu/Zu]cLer and n[Zu/Ler]cLer which appear to display a slight multimodal distribution. However, these multimodal distribution appear to be artefacts created by the variation among plant replicates. The median of these three replicates is centred on the peak of stronger magnitude, which can be considered to be the true peak of the distribution. The first observation

that can be made on F1 seed size between the different crosses performed in this chapter is that each cybrids, hybrids and hybrid-cybrids seeds display a size similar to the accession of the maternal-of-origin nucleotype, with some variations. Furthermore, a positive or negative effect of hybridity and cytonuclear hybridity (cybridity) on F1 seed size arise in opposition between pairs of reciprocal cybrids, hybrids and hybrid-cybrids lines (Figure 4.3 and supplementary figure S4.1). Between Col-0 and C24 (Figure 4.3A and supplementary figure S4.1a and S4.1B), each cybrid seed size does not differ significantly from its maternal-of-origin nucleotype parent (n[Col/Col]cCol for n[Col/Col]cC24 and n[C24/C24]cC24 for n[C24/C24]cCol. Seed size of the hybrid n[Col/C24]cCol does not display a significant difference compared to n[Col/Col]cCol either, but seeds of the hybrid-cybrid n[Col/C24]cC24 are significantly smaller than n[Col/Col]cCol, associated with NH. On the other hand, the reciprocal hybrid n[C24/Col]cC24 displays a significantly increased seed size compared to his maternal-of-origin nucleotype parent n[C24/C24]cC24, associated with BPH, and the hybrid-cybrid n[C24/Col]cCol displays an even greater seed size increase in comparison (figure 4.3A and Supplemental figure S4.1A). Similar effects can be observed within the other genomic combinations, with the Col-0/Zu-0 pair displaying the most dramatic BPH and NH effects on reciprocal cybrids, hybrids and hybrid-cybrids. A similar behaviour can be observed for the C24/Zu-0 combination, with n[C24/C24]cZu displaying WPH and being statistically similar to n[C24/C24]cC24 control, and n[Zu/Zu]cC24 displaying MPH and a seed size closer to n[Zu/Zu]cZu control than to c[C24/C24]cC24. However hybrids show an opposite behaviour with n[C24/Zu]cZu displaying BPH heterosis and n[Zu/C24]cZu displaying WPH (figure 4.3E and Supplemental figure S4.1E). Both hybrid-cybrids of this combination surprisingly display BPH in opposition to all the other combinations where reciprocal hybrid-cybrids tend to behave in opposition depending on their maternal-of-origin nucleotype.

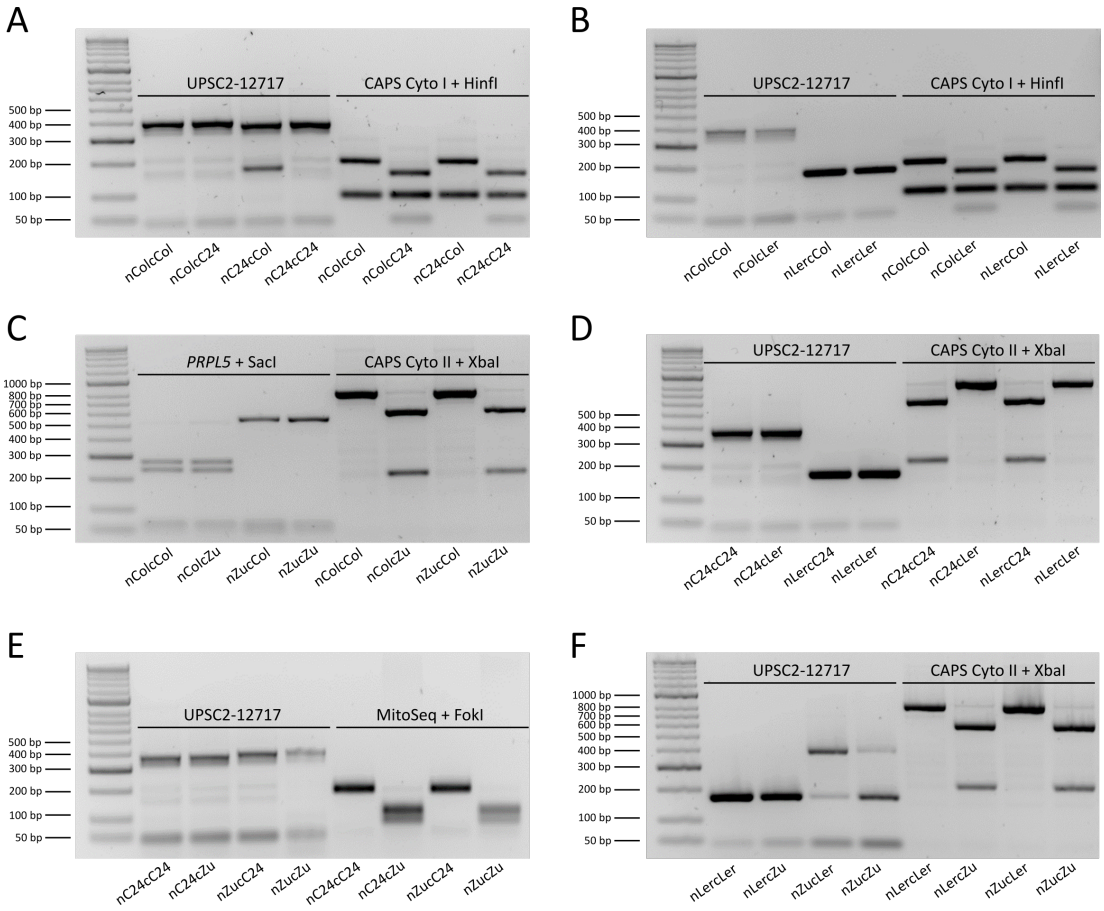


Figure 4.2: Genotyping of Cybrid lines.
Nucleotype and plasmotype genotyping for each cybrid combination between Col-0, C24, Ler-0 and Zu-0. Nucleotype of each line was compared to isogenic lines using the nuclear marker UPSC2-12717 **(A, B, D, E, and F)** and the presence of a SNP in the sequence of *PRPL5* allowing digestion by SacI **(C)**. Cytotype was compared to isogenic lines using microsatellites CAPS Cyto I followed by a digestion with HinfI **(A and B)**, CAPS Cyto II followed by a digestion with XbaI **(C, D and F)**, and MitoSeq followed by a digestion with FokI **(E)**.

Due to the Col-0/Zu-0 combination displaying the most important BPH and NH effects on cybrid and hybrid-cybrid seed size (Figure 4.3C and Supplementary figure S4.1E and S4.1F), it was selected for further seed development analysis aiming to understand the causes of seed size divergence between cybrid and hybrid and hybrid-cybrid lines in comparison to their maternal-of-origin isogenic parent. The eight different lines were separated into two groups centred on their maternally-inherited nucleotype since, according to our observations, it is the nucleotype which most significantly affects F1 seed size determination. I thus separated on one side the maternally-inherited Col-0 (hereafter referred as "Col-0 group") lines n[Col/Col]cCol (isogenic), n[Col/Col]/Zu (cybrid), n[Col/Zu]cCol (hybrid) and n[Col/Zu]cZu (hybrid-cybrid), and on the other side the maternally-inherited Zu-0 lines (hereafter refered as "Zu-0 group") n[Zu/Zu]cZu, n[Zu/Zu]cCol, n[Zu/Col]cZu and n[Zu/Col]cCol. n[Zu/Zu]cZu was kept within the Col-0 group and n[Col/Col]cCol was kept within the Zu-0 group for comparison of each line to their plasmotype and/or paternal-of-origin nucleotype parent (figure 4.4).

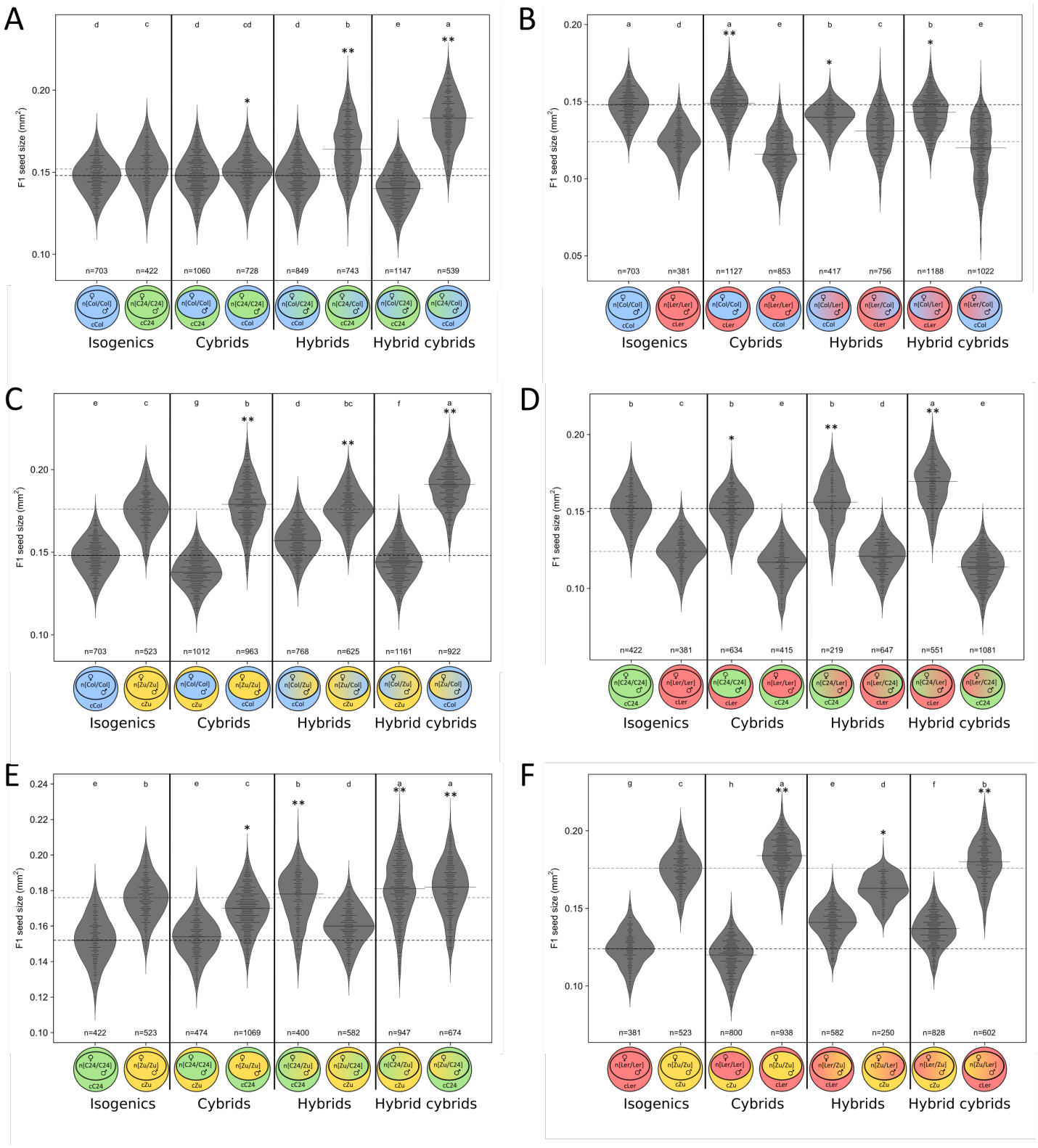


Figure 4.3: F1 seed size.

Density distribution of F1 seed size from manually crossed isogenic, cybrid, hybrid and hybrid-cybrid lines for all six combinations between Col-0, C24, Ler-0 and Zu-0. Col-0/C24 (**A**), Col-0/Ler-0 (**B**), Col-0/Zu-0 (**C**), C24/Ler-0 (**D**), C24/Zu-0 (**E**) and Ler-0/Zu-0 (**F**). Black and grey dashed line represents median F1 seed size value for isogenic parents. Different letters display a significant statistical difference between groups, according to ANOVA Tukey-Kramer test with Bonferroni correction (α = 0.0001). Star represents sample sets displaying MPH (*) and BPH (**) compared to isogenic lines. The number n of seeds used for each line is indicated below each bean plot.

Differences in seed size between cybrids, hybrids and hybrids cybrids arises at 3 DAP

A time course experiment was set from 1 to 6 DAP to study the differences of seed size over time between lines. Three siliques at each timepoint were fixed in a solution of ethanol and acetic acid (9:1), progressively rehydrated and cleared in Hoyers solution before DIC microscropy imaging. Seed size was observed to diverge significantly between lines as early as 3 DAP in both Col-0 and Zu-0 groups (Figure 4.4A and 4.4C). Seed size at 3 DAP is represented as beanplots on figures 4.4B and 4.4D. All distributions appear normal at the exception of n[Zu/Col]cCol which displays a bimodal distribution. However this bimodal distribution appears to be an artefact resulting from the lower number of replicates of this sample (11) compared to the others (circa 20). The median value for this sample is displayed very close to the top peak, hence it is logical to assume that the lower peak is an artefact from the bean plotting of so few replicates. Similar seed size differences can be observed between cybrids, hybrids and hybrid-cybrids at 3 DAP and later developmental stages as was observed previously at seed maturity. The cybrid n[Col/Col]cZu is slightly smaller and the hybrid n[Col/Zu]cCol is slightly bigger than n[Col/Col]cCol, and the hybrid-cybrid n[Col/Zu]cZu is significantly smaller than all the other lines (figure 4.4B). Cybrid n[Zu/Zu]cCol and hybrid n[Zu/Col]cZu are not statistically different, but the hybrid-cybrid n[Zu/Col]cCol is statistically bigger than the two others (Figures 4.4D). Cybrid and hybrid lines of the Zu-0 group however were smaller than expected at that stage in comparison to controls: both cybrid and hybrid display a size closer to n[Col/Col]cCol than to n[Zu/Zu]cZu whereas their F1 seed size is closer to the one of n[Zu/Zu]cZu, and the hybrid-cybrid is not significantly different to n[Zu/Zu]cZu whereas its F1 seed size is significantly bigger.

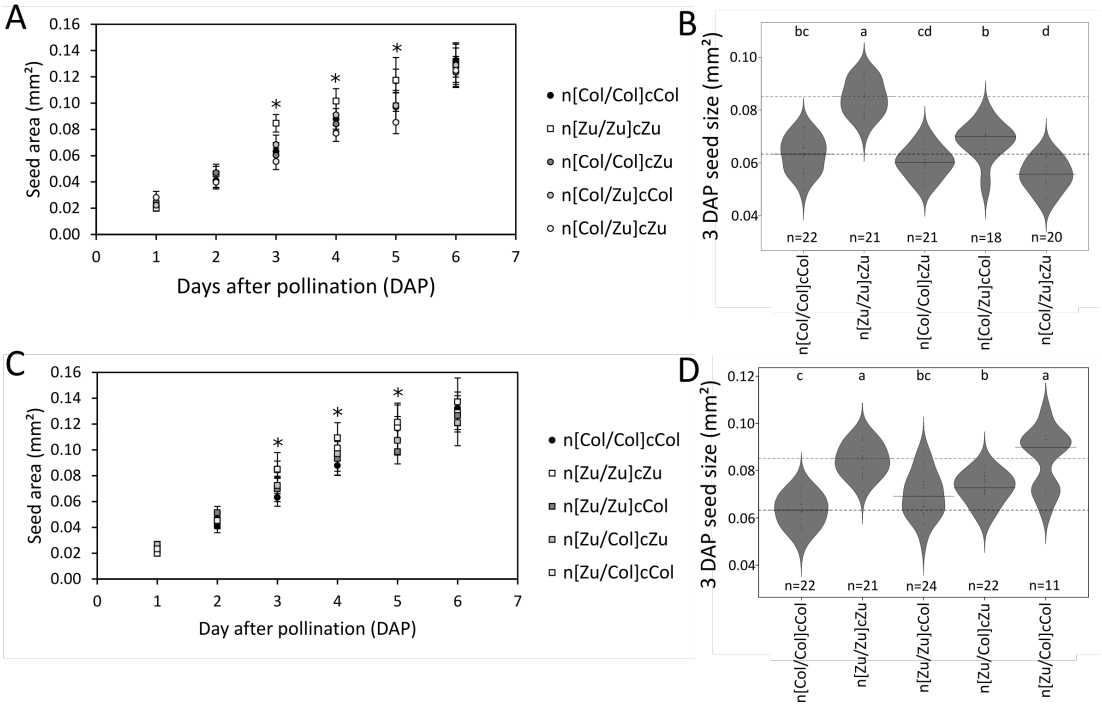


Figure 4.4: Time course of developing seeds.
Seed size of isogenic, cybrid, hybrid and hybrid-cybrid seeds from Col-0 and Zu-0 groups from 1 to 6 Days after pollination. **A and C**: Average seed size from 1 to 6 DAP of maternal Col-0 Cybrid, Hybrid and Hybrid-cybrid compared to Isogenic n[Col/Col]cCol parent (A) maternal Zu-0 Cybrid, Hybrid and Hybrid-cybrid compared to Isogenic n[Zu/Zu]cZu parent (C). Stars highlight timepoints at which the seed size of Cybrids, Hybrids and Hybrid-cybrids becomes statistically different from isogenic controls. **B and D**: Density distribution at 3 DAP of seed size maternal Col-0 Cybrid, Hybrid and Hybrid-cybrid compared to Isogenic n[Col/Col]cCol parent (B) and of maternal Zu-0 Cybrid, Hybrid and Hybrid-cybrid compared to Isogenic n[Zu/Zu]cZu parent (D). Statistical ANOVA test and post-hoc Tuckey test displayed as letters above. Different letters represent statistically different samples ($\alpha = 0.0001$). The number n of seeds used for each line is indicated below each bean plot.

Cybridity and hybridity on one hand and hybrid-cybridity of the other hand have different effects on whole seed transcriptome

In order to understand which transcriptomic changes in the seed could explain the observed differences in seed size, I performed an RNAseq analysis using RNA extracted from 3 DAP seeds and performed a paired-end sequencing analysis (20 million reads depth). RNA was extracted from two biological replicates from each of the eight crosses between Col-0 and Zu-0 cybrid lines. Resulting reads were cleaned using fastp and aligned to TAIR10.1 reference transcriptome using Tophat, giving a 97.2% alignment rate for n[Col/Col]cCol and n[Col/Col]cZu, 89.8% alignement for n[Zu/Zu]cZu and n[Zu/Zu]cCol, and a 89.5% alignment rate for hybrids and hybrid-

cybrids. I used Cuffdiff to calculate gene count tables per sample using the fragment per kilobase per million (FPKM) method and then determining differently expressed genes (DEGs) between samples using the geometric method and the negative binomial dispersion estimation method (Poisson Law). DEGs were calculated between samples as different subsets: isogenic subset (between n[Col/Col]cCol and n[Zu/Zu]cZu), cybridity subset (between cybrid n[Col/Col]cZu and its maternally-related isogenic n[Col/Col]cCol for the Col-0 group, and between n[Zu/Zu]cCol and n[Zu/Zu]cZu for the Zu-0 group), hybridity subset (between hybrid n[Col/Zu]cCol and its maternally-related isogenic n[Col/Col]cCol for the Col-0 group and between n[Zu/Col]Zu and n[Zu/Zu]cZu for the Zu-0 group), cybrid-hybridity subset (between hybrid-cybrid n[Col/Zu]cZu and its maternally-related cybrid n[Col/Col]cZu for the Col-0 group, and between n[Zu/Col]cCol and n[Zu/Zu]cCol for the Zu-0 group) and hybrid-cybridity subset (between hybrid-cybrid n[Col/Zu]cZu and its maternally-related hybrid n[Col/Zu]cCol for the Col-0 group and between n[Zu/Col]cCol and n[Zu/Col]cZu for the Zu-0 group). Cybridity and hybrid-cybridity subsets of DEGs are grouped under the name “cybridity-induced DEGs” and hybridity and cybrid hybridity subsets are grouped under the name “hybridity-induced DEGs”. True DEG were retained as associated to a q-value < 0.05 and a Log2 fold-change of at least 2.

A total of 440 DEGs were identified in the isogenic subset (275 downregulated and 175 upregulated DEGs). I identified 318 DEGs in the cybridity subset from the Col-0 group (216 downregulated and 102 upregulated) against 169 DEGs in the Zu-0 group (15 downregulated and 154 upregulated). In comparison, a smaller amount of DEGs were identified in the hybrid-cybridity subset suggesting a smaller effect of cybridity in hybrids than in isogenic lines: 50 DEGs were found for the Col-0 group (23 downregulated and 27 upregulated) and 109 DEGs for the Zu-0 group (40 downregulated and 69 upregulated). In opposition, more genes were found within the hybridity induced DEGs subsets for both groups (Figure 4.5A and 4.5C). I indeed identified 271 DEGs in the hybrid subset for the Col-0 group (126 downregulated and 145 upregulated) and 425 DEGs for the Zu-0 group (206 downregulated and 219 upregulated). A similar amount was found in the cybrid hybridity subsets: 248 DEGs for the Col-0 group (79 downregulated and 169 upregulated) and 320 for the Zu-0 group (243 downregulated and 77 upregulated). It thus seems that in general

hybridity induces more DEGs than cybridity, whether in the Col-0 or in the Zu-0 group. Figures 4.5A and 4.5C suggest an important overlap between both hybridity and cybrid hybridity clusters of DEGs, but also suggest a low overlap between cybridity and hybrid-cybridity clusters. Surprisingly however nearly no dysregulated genes can be observed from mitochondrial genome between samples. Eleven genes from plastidial genome are found dysregulated between Col-0 and Zu-0 isogenic cybrids, but a very low amount is found between cybrids, hybrids and hybrid-cybrids. This suggests a high impact of organelle genome swap on nuclear gene regulation, but a low impact on organelle gene regulation at that stage of development.

Gene ontology (GO) enrichment was performed on described above DEG subsets using Panther. The majority of found GO terms associated with all subsets are related to biotic and abiotic stress. This is especially true for both cybridity subsets. Cell-cell signalling GO terms were highly represented in hybridity-induced DEGs subsets for the Col-0 group (Figure 4.5B), but not for the Zu-0 group in which these DEG subsets were more enriched in development, secondary metabolism, circadian rhythm and abiotic stress GO terms (Figure 4.5D). Circadian rhythm alongside with transcription regulation and abiotic stress GO terms were found significantly enriched also within the hybrid-cybridity subset in both Col-0 and Zu-0 groups. Finally, the GO term response to hormone was found significantly enriched in both cybridity and hybrid-cybridity subsets from the Zu-0 group, in particular the upregulation of *TCH4* (Xyloglucan endotransglucosylase) and *ACS6* (1-aminocyclopropane-1-carboxylic acid synthase 6).

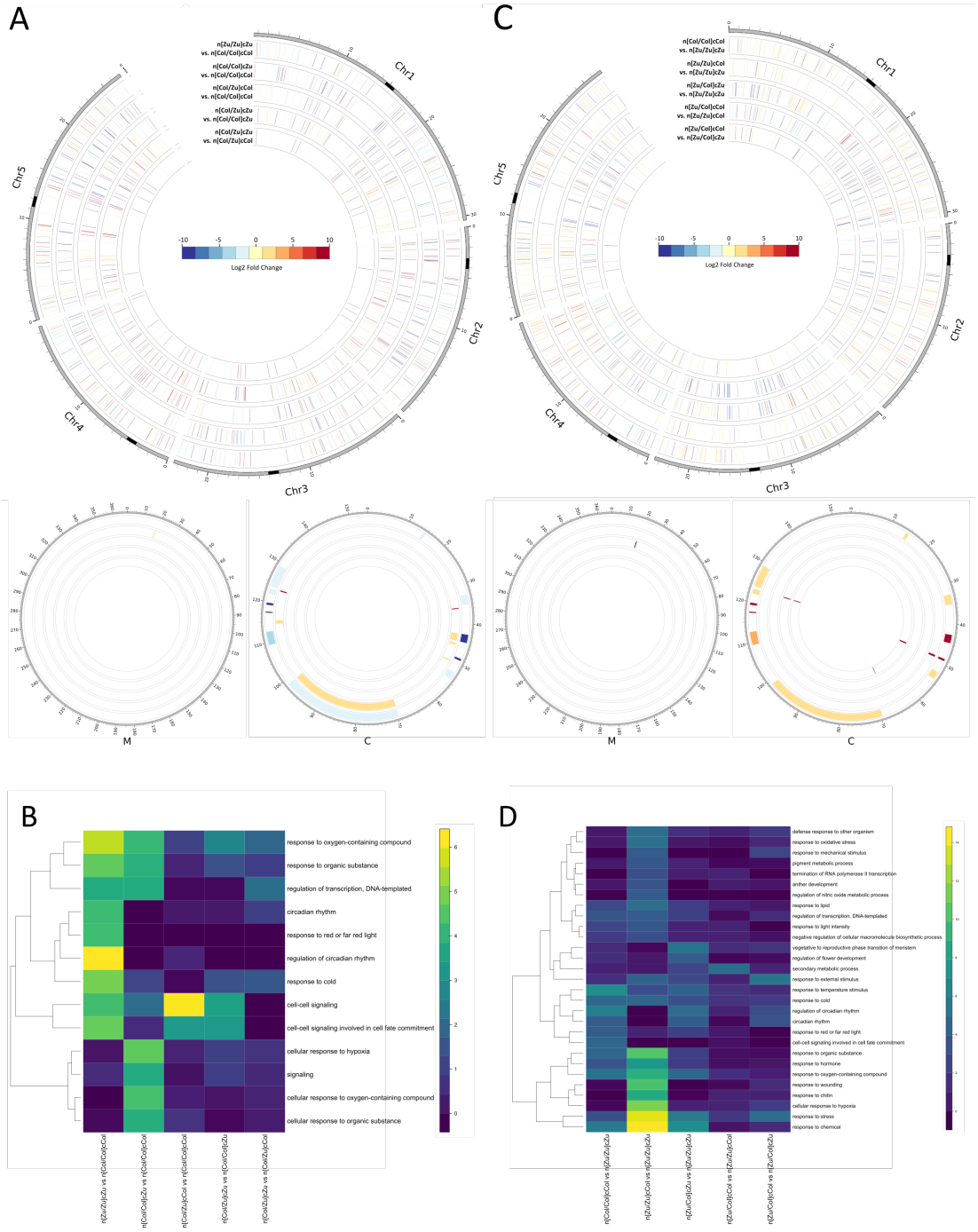


Figure 4.5: Differentially expressed genes (DEG).
Genomic mapping of DEGs and GO enrichment of each subset of Col-0 and Zu-0 groups. **A and C**: Circular genome representation of DEGs observed samples from Col-0 **(A)** and Zu-0 **(C)** groups. Each bar represent one gene at its position on its chromosome. Colour scale represent the Log2 Expression Fold Change for a DEG between samples. Nuclear genome is displayed on the top, Mitochondrial (M) and Chloroplastic (C) genomes are displayed on the bottom. Scale of nuclear genome is displayed as megabases and the one of mitochondrial and chloroplastic genomes as kilobases. Black bands on nuclear chromosome scales represent centromeres. **B and D:** Heatmap dendrogram of the –Log10 of significant Gene Ontology terms p-value for both Col-0 **(B)** and Zu-0 **(D)** groups.

The analysis of dysregulated pathways enrichment within each subsets tend to correlate our first assumptions on subset overlaps (Supplementary figure S4.2). Indeed, a significant representation of the ceramide metabolism pathway can be observed in association to an upregulation of the alkaline ceramidase-encoding *AT5G58980* in both hybridity-induced DEG subsets from the Col-0 group (Supplementary Figure S4.2B). Surprisingly, even though this pathway appears also dysregulated in the isogenic subset, it was not found significantly dysregulated within the Zu-0 group. Two other major pathways were found dysregulated within the Col-0 group. The cybridity subset is associated with a downregulation of genes involved in the glycosyl metabolism pathway especially the glycosyl hydrolase encoded by *AT5G20940* (Supplementary Figure S4.2A), whereas the hybrid-cybridity subset is rather associated with a downregulation of genes involved in the carbohydrate metabolism pathway represented by the phosphofructokinase encoded by *AT1G50390* (Supplementary Figure S4.2C). Different pathways were found dysregulated within the Zu-0 group more associated with cell wall organization like the gene *AT1G11590* encoding a pectin methyl-esterase inhibitor found downregulated in both hybridity-induced subsets and suggesting an enhanced cell wall loosening (Supplementary Figure S4.2D). Similarly, cell wall pathway was similarly dysregulated between n[Zu/Zu]cZu and n[Col/Col]cCol with the upregulation of *AT3G59850* encoding a the pectin lyase and the downregulation of *AT2G45220* encoding a pectin methyl-esterase inhibitor (supplementary figure S4.2E). These results suggest that hybridity tend to dysregulate similar pathways whether in isogenic or cybrid lines, whereas cybridity appears to dysregulate different pathways depending whether the genomic context is isogenic or hybrid. This suggests that either disruption of cytonuclear relationships or switching from maternal cytonuclear coupling to paternal cytonuclear coupling lead to different responses on global regulation of gene expression.

Next I isolated DEGs specifically induced by cybridity, or by hybridity, or induced by both cybridity and hybridity. I thus found 241 DEGs commonly dysregulated between samples in the Col-0 group, and 294 in the Zu-0 group. A high overlap was especially detected between both hybridity-induced DEG subsets with genes globally dysregulated in the same direction, same as for figures 4.5A and 4.5C. I found 13

downregulated and 80 upregulated DEGs in this overlap for the Col-0 group (Figure 4.6A and 4.6B) and 114 downregulated and 15 upregulated DEGs for the Zu-0 group, (Figure 4.6D and 4.6F). No or low overlap was found between both cybridity-induced DEG subsets except in the Zu-0 group where I found 1 downregulated and 17 upregulated DEGs (Figure 4.6D and 4.6E). These genes form a cluster mostly composed of genes involved in hormone response and cell wall plasticity such as *TCH4* and *ACS6* described above, but also in stress response GO terms with genes such as the chloroplast-encoded *ndhJ* encoding the NADH dehydrogenase subunit J. This cluster is displayed in the heatmap dendrogram were DEG are represented as Log2 Fold-Change (Log2FC) only from the overlap between cybridity, hybridity and hybrid-cybridity subsets from the Zu-0 group (Figure 4.6F). The exclusive overlap between hybridity-induced DEG subsets are not included in this figure. A similar dendrogram for the Col-0 subset is represented in the figure 4.6C. A significant overlap was also found between cybridity and hybridity subsets for both groups with genes globally dysregulated in the same direction. In the Col-0 group, 28 DEGs were downregulated and 7 DEGs were upregulated, whereas 6 DEGs were found downregulated and 24 DEGs were upregulated in the Zu-0 group. These clusters tend to also be upregulated within the isogenic subset, but not in either cybrid- hybridity nor hybrid-cybridity subsets (figure 4.6C and 4.6F), suggesting that cybridity or hybridity is sufficient to change significantly the regulation of these genes with no further dysregulation observed when cybridity and hybridity are combined within the same line. These upregulated clusters are mostly composed of genes involved in seed storage, lipid and starch metabolism (like the lipid transfer gene *AT4G12825* and the qua-quine starch gene *QQS*) in the Col-0 group and in hormone response (like the auxin responsive gene *BRU6),* photosynthesis (like the light harvesting complex subunit gene *LHB1B1* and the protochlorophyl oxidoreductase gene *PORA*) and oxidative stress response (like the thioredoxin superfamily genes *At2G15260* and *ROXY2*) in the Zu-0 group. The downregulated clusters are mostly composed of genes involved in self-incompatilibity (like the genes *AT1G26795* and *AT3G16960*), dormancy (like the genes *AT2G33830* and *DYL1*), seed storage (like the lipid transfer gene *AT2G37870*, the seed storage albumin gene *SESA1* and the raffinose synthase gene DIN10) in the Col-0 group and in self-incompatibility (like the self-

incompatibility-related gene *AT5G06040*) and stress responses (like the peroxidase family gene *AT5G06730*) in the Zu-0 group.

A small overlap was found between hybridity and hybrid-cybridity subsets. At the same time another significant overlap was found between hybrid-cybridity and cybrid-hybridity subsets both in Col-0 and Zu-0 groups. In the Col-0 group, this cluster is composed of 5 downregulated DEGs, including the embryo-expressed WUSHEL-related homeobox gene *WOX4* and the abscisic acid responsive glycosyl hydrolase-encoding ChiC*,* and 9 upregulated DEGs, including the gibberellin-related transcription factor *ddf2*. Interestingly, *ddf2* was found significantly downregulated in the cybridity and hybridity subsets in the Col-0 group. In the Zu-0 group, 10 DEGs were found downregulated, including the ethylene-responsive element binding factor *ERF13* involved in biotic stress responses, and 10 DEGs were upregulated, including the chloroplast-encoded *rps15* which encodes for the plastid ribosomal subunit S15 required for translation in the chloroplast. An important portion of these clusters specific to hybrid-cybrids is however composed of yet uncharacterised genes, which suggest the involvement of other new or incompletely described pathways. An additional cluster of genes is observable on each heatmap dendrogram but not within the Venn diagrams and is composed of genes downregulated in both cybridity and hybridity subsets while not being found significantly dysregulated in the isogenic subset. Interestingly this cluster tends to be found dysregulated in the opposite direction in both cybrid hybridity and hybrid-cybridity subsets. These clusters are mostly composed of genes involved in glucose metabolism (like the phosphoenolpyruvate carboxylase-encoding *PCK2*) and splicing (like *SMP1*) in the Col-0 group (figure 4.6C) and oxidative stress response (like the peroxidase superfamily gene *AT5G06730*) in the Zu-0 group (figure 4.6F).

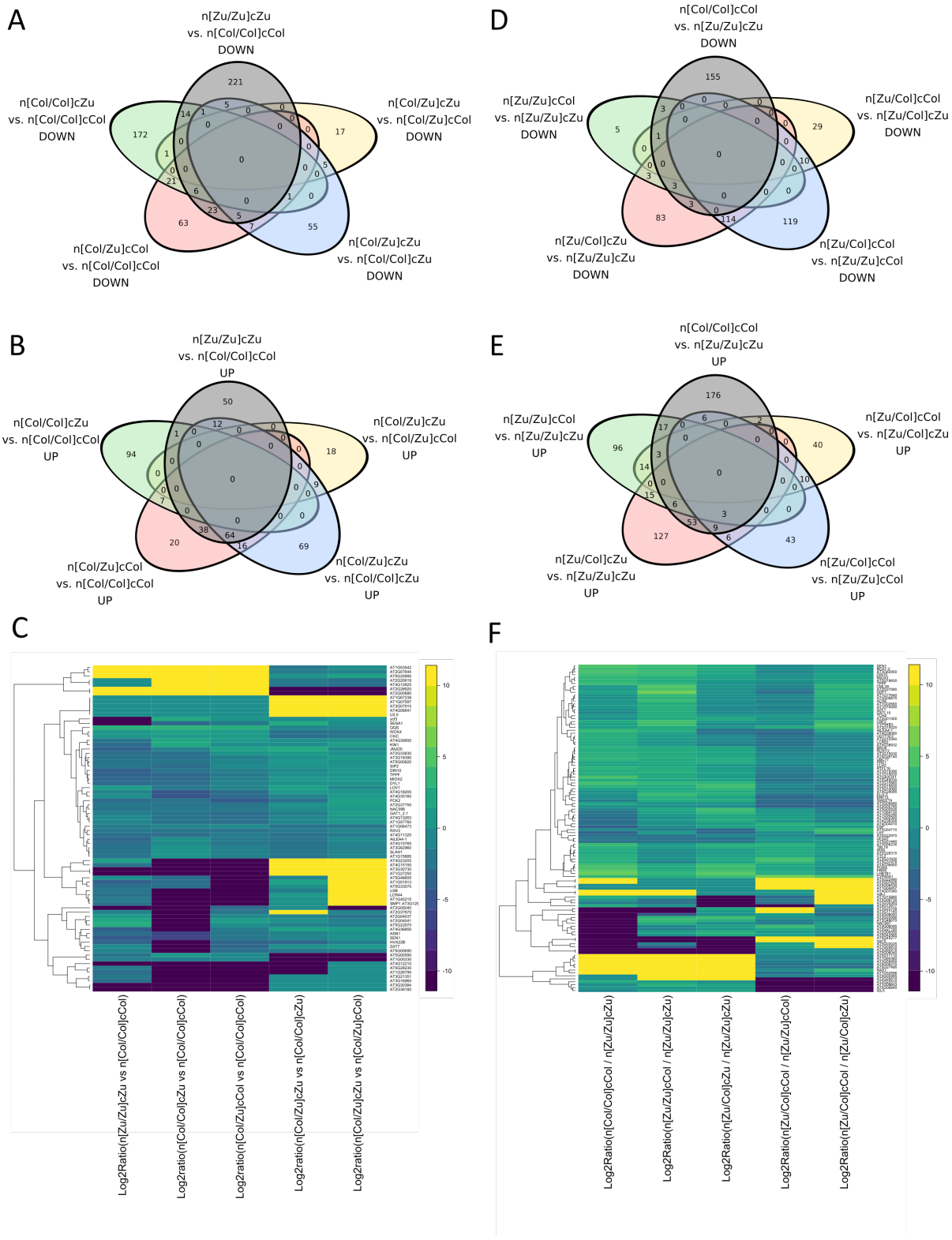


Figure 4.6: Overlaps of differentially expressed genes (DEG).
Overlaps and clusters of DEGs between the different subsets of Col-0 and Zu-0 groups. **A, B, D and E**: Venn diagram of Down- **(A and D)** and Up- **(B and E)** regulated DEGs between samples for the Col-0 group **(A and B)** and for the Zu-0 group **(D and E)**. **C and F:** Heatmap dendrogram of overlapping DEGs in maternal Col-0 **(A)** and maternal Zu-0 **(B)** cybrids, hybrids and hybrid-cybrids in comparison to isogenic controls.

Some gene targets especially involved in dysregulated pathways described above were verified by RT-qPCR on two different replicates from the same samples as the ones used for RNAseq. I included *AT5G20940* for the Col-0 group, *AT1G11590*, *TCH4* and *ACS6* for the Zu-0 group, and *AT1G50390* for both groups. In contrary to RNAseq results, I did not find any significant variation in *AT5G20940* in the Col-0 group and for *AT1G50390* expression in both groups, except a 2-fold decrease of *AT1G50390* expression between n[Zu/Col]cZu and n[Zu/Zu]cZu as well as between n[Zu/Col]cCol and n[Zu/Zu]cCol, which didn't appear in the the RNAseq data. There is also an upregulation tendency of *AT1G58980* in cybrids and hybrids from both group in comparison to control, and a downregulation tendency in both hybrid cybrids in comparison to both reciprocal hybrid-cybrid. However, a significant 33-fold upregulation of *AT1G58980* (p=0.013) was observed between n[Zu/Zu]cZu and n[Col/Col]cCol, correlating our RNAseq data (figure 4.6A and supplementary figure S4.3). The RT-qPCR of *AT1G11590* also showed a significant 2-fold decrease between n[Zu/Zu]cCol and n[Zu/Zu]cZu and a significant 1.5-fold increase between n[Zu/Col]cCol and n[Zu/Zu]cCol in contradiction with RNAseq data. The expression level of *TCH4* and *ACS6* was 4-fold and 13-fold upregulated, respectively, between n[Zu/Zu]cCol and n[Zu/Zu]cZu, as well as 2.5-fold upregulated between n[Zu/Col]cCol and n[Zu/Col]cZu, which correlates RNAseq data. A tendency for downregulation for *TCH4* and *ACS6* genes was also observed between n[Zu/Col]cZu and n[Zu/Zu]cZu as well as between n[Zu/Col]cCol and n[Zu/Zu]cCol, which wasn't found in RNAseq data.

Discussion

In most of the ecotype combinations shown by this chapter (Col-0/C24, Col-0/Ler-0, Col-0/Zu-0, C24/Ler-0 and Ler-0/Zu-0), it can be noted that F1 seed size of cybrids, hybrids and hybrid-cybrids was relatively close to the one of their nuclear maternal parent, correlating the previously published model highlighting the central role of the maternal seed coat and maternally biased endosperm tissues in seed size determination in *A. thaliana* (Li and Li 2015). Hybridity and cybridity tend to lead to either MPH or BPH if the maternal parent displays the best seed size of the two

parents, or to lead to either WPH or NH if the maternal parent displays the worst seed size of the two parents. Thus it seems that crosses with the best performing of both parents acting as a pollen receiver have a greater chances to lead to BPH F1 seed size. Despite the recent publications using the cytoswap technique to produce cybrids, also called “Cytoswaps” (Kuppu et al. 2015; Kelliher et al. 2017; Flood et al. 2020), this chapter is to our knowledge the first one reporting the effect a cumulated hybridity and cybridity in plant lines. Two major tendencies of seed size variation in cybrids, hybrids and hybrid-cybrids seem to show up in these case. 1) If both cybridity and hybridity lead to either BPH or NH, both effects seem to accumulate and lead respectively to a more intense BPH or NH in hybrid-cybrids. 2) If cybridity or hybridity leads to either MPH or WPH, both effects seem to balance each other and lead to a hybrid-cybrid seed size averaging the one of both cybrid and hybrid lines. It is interesting to see that this seed size variation between cybrids, hybrids and hybrid-cybrids in comparison to their maternal-of-origin isogenic line is already established as early at 3 DAP during the seed development as I show on figure 4.4 for the Col-0/Zu-0 combination. This correlates the early seed size heterosis establishment previously reported in Cvi-0 X Ler-0 hybrids by Alonso-Blanco et al. (1999). It is surprisingly to see in the Zu-0 group that although the seed size differences seems already established between the cybrid, hybrid and hybrid-cybrid lines, they seem to be slightly late in development compared to their isogenic n[Zu/Zu]cZu control: cybrid and hybrid seeds display MPH while final seed size has previously shown them to display BPH. From our data it is possible to hypothesize that the phenomenon of hybridization, whether it is cytonuclear or intra-nuclear, can lead to a slight delay in seed development which tend to disappear in later stages. Such a phenomenon has not yet been reported to our knowledge. This delay in size through seed development however seems to only happen for hybridized lines which will later display either MPH or BPH, and hasn’t been reported in previous cases to our knowledge so far.

Transcriptome analysis of whole seeds at 3 DAP demonstrates that cybridity can mimick hybridity to some levels by having a similar effect on a subset of the genes dysregulated as a consequence of hybridity. This appears to correlate the fact that cybrid n[Zu/Zu]cCol displays an equal or even slightly larger seed size than the hybrid n[Zu/Col]cZu. This effect can be more easily visualized with the cybrid n[Zu/Zu]cLer

and the hybrid n[Zu/Ler]cZu in the Zu-0 group from the Ler-0/Zu-0 combination (Figure 4.3F). The cumulative effect of cybridity and hybridity on seed size observed in hybrid-cybrid seeds appears to also be confirmed at the transcriptome level with a cluster of genes being dysregulated only in hybrid-cybrids when both cybridity and hybridity are combined.

Gene ontology analysis of the different sets of DEGs reveals a significant enrichment in biotic and abiotic stress responses pathways in all sets, but most importantly in isogenic cybrids from both Col-0 and Zu-0 group than in any other samples. Cytonuclear relationships thus appears to have a significant control over stress responses pathways, correlating previous data showing that the impact of plasmotype variation on plant fitness appeared more important under stressful conditions (Flood et al. 2020). In an analogous way of previous studies of stress responses from hybrids (Groszmann et al. 2015; Yang et al. 2015), it would be interesting to explore if hybrid-cybrids can obtain an even more important resistance to stressful conditions. It is interesting however to note that these pathways are more dysregulated in cybrids where nuclear genome and organelle genome are from different accessions than in hybrids or hybrid-cybrids in which organelles are from the same accession than at least one side of the nuclear genome. Thus despite the maternal cytonuclear genome coupling previously suggested in *A. thaliana* (Wolf and Hager 2006; Wolf 2009) and animals (Rand et al. 2004) and to which further evidence were found more recently in plants (Sloan et al. 2018; Li et al. 2019a), our data also suggest a strong relationship between paternal nuclear genome and organelle genomes in this self-pollinating species.

Apart from stress responses pathways, significantly dysregulated pathways give us some insights in the seed size differences observed in cybrids and hybrid-cybrids. Among these I find the Fructokinase (FRK)-encoding gene *AT1G50390* which assures the transformation of D-Fructose in D-fructose-6-phosphate. Despite the absence of confirmation of a downregulation of the gene between n[Col/Zu]cZu and n[Col/Zu]cCol which may be coming from some variation between replicates, this FRK was significantly downregulated (nearly 2-fold) between n[Zu/Col]cZu and n[Zu/Zu]cZu (supplementary figure S4.3B). The knockout of FRKs has been linked to

a reduction of sucrose storage in the seed, with a reduction of fatty acid metabolism leading to embryo arrest (Stein et al. 2017). The downregulation of this gene could thus correlate with the slightly reduced seed size of hybrid n[Zu/Col]cZu observed in comparison to cybrid n[Zu/Zu]cCol in which the expression of this particular FRK is not statistically different in comparison to the isogenic n[Zu/Zu]cZu (supplementary figure 4.3B). There is however a high redundancy of FRKs in *A. thaliana*, and as no other FRKs was found dysregulated in n[Zu/Col]cZu in our data this particular one would be expected to have an limited effect on its own.

We have seen that hybridity in comparison to cybridity has a much stronger effect on gene expression, suggesting that hybridity have a stronger effect on seed development than cybridity on their own. The most significantly dysregulated pathway found in both hybrids in the Col-0 group was an increase of alkaline ceramidase activity with a significant up-regulation of *AT5G58980*. Ceramide synthase overexpression was reported to enhance biomass thanks to an increase of cell division and growth in *A. thaliana* (Luttgeharm et al. 2015). Previous reports on alkaline ceramidase, especially TOD1, also revealed its importance in the regulation of turgor pressure necessary for pollen tube elongation and for guard cells opening in siliques (Chen et al. 2015). From our results a similar role of alkaline ceramidases in the control of turgor pressure in seeds could be speculated, especially during endosperm proliferation. An increase of this activity would then correlate with a decreased endosperm proliferation and seed size. It is important to note that RT-qPCR reveal little or no variation in *AT5G58980* expression pattern between samples of the same group in response to hybridization of cybridization (supplementary figure 4.3A and 4.3B). This might come from important variations of alkaline ceramidase gene expression between replicates of the same sample, but could also suggest a non-significant role of *AT5G58980* in seed size variation. Rather, the variation of *AT5G58980* expression would tend to correlate the important seed size difference between n[Col/Col]cCol and n[Zu/Zu]cZu controls and thus between Col and Zu sample groups, suggesting a strong maternal control of its expression.

In the other group cell wall plasticity seems to be the most dysregulated pathway, with two mainly representative genes: *TCH4* and *AT1G11590. TCH4* encodes a

xyloglucan endotransglucosylase, enzyme which modifies the composition of cell walls in response to auxin and brassinosteroids hormones and environmental cues, hence would be expected to promote growth (Xu et al. 1995; Van Sandt et al. 2007). In opposition, *AT1G11590* encodes a pectin methylesterase inhibitor. As pectin methylesterases have been shown to participate in cell wall loosening (Levesque-Tremblay et al. 2015), a reduction of pectin methylesterase inhibitors would correlate with an enhanced cell wall loosening further supporting plant growth. Finally hormone responsive genes such as *ACS6* seem to have some importance in the Zu-0 group samples. *ACS6* encodes the 1-aminocyclopropane-1-carboxylic acid synthase 6, which participates in ethylene production and is reported to promote iron storage during seed maturation (Sun et al. 2020). The impairment of ethylene biosynthesis was also linked with a reduction of seed size and number (Walton et al. 2012), hence an increase of ethylene biosynthesis as observed in our samples could probably be associated with an increased seed size.

Aside of these pathways it is also interesting to look closer at the previously identified cytonuclear retrograde and anterograde pathways. These data indeed show only one dysregulated genes associated with mitochondrial genome (*ccmB*, necessary for Cytochrome c biogenesis) and appearing dysregulated only in few sample comparisons, which is surprising in the light of the usual major role of mitochondria anterograde signalling in high plants (Yang et al. 2008; Berkowitz et al. 2016). In contrast two major chloroplast genes, *ndhJ* and *rps15* were found strongly upregulated in samples of the Zu-0 group, the first one being cybridity-specific and the second hybrid-cybrid specific. As *ndhJ* is related to oxidative stress, it would be interesting to investigate if it can act as an initiator in the plastid anterograde signalling pathway to the nucleus as chloroplastic are known to regulate genes of interest in the nucleus especially in conditions of oxidative stress (Nott et al. 2006; Wang et al. 2006). The same hypothesis could be used for *rps15* as it codes for a plastid ribosomal protein which, as well acting as part of the plastid ribosome, could possibly have a role to play in the nucleus from the many plastid ribosomal proteins previously identified in nucleus extracts under potentially stressful conditions. (Papadakis and Roubelakis-Angelakis 1999; Bae et al. 2003; Sakamoto and Takagi 2013; Goto et al. 2019). Other genes such as the WRKY family were also identified in

our data as dysregulated in different DEGs subsets, especially *WRKY40* upregulated in n[Zu/Zu]cCol vs n[Zu/Zu]cZu which was previously identified as a modulator of the expression of stress-responsive nuclear genes encoding mitochondrial and plastid proteins (Van Aken et al. 2013). It thus appear that a modification of the expression of genes involved in cytonuclear signalling itself in the cybrids, hybrids and hybrid-cybrids generated in this study is mostly related to stress related responses. This would not be expected to modify seed size determination per se, but might enhance cybridization effects on seed size in stressfull conditions.

Our transcriptomic data confirm our hypothesis of cumulative effects of hybridization and cybridization on gene expression, which is especially true for the Zu-0 group. A cluster of genes which is significantly dysregulated in n[Zu/Col]cCol as a consequence of a combination of hybridity and cybridity can be observed. The fact that this DEG cluster is larger in the Zu-0 group in comparison to the Col-0 group would suggests a more important synergistic effect of paternally inherited nuclear and organelle genomes when they come from the Col-0 accession than when they come from the Zu-0 accession. No particular pathway however appears enriched within this cluster, which contains a significant amount of uncharacterised genes. Thus a true effect of the combination of hybridity and cybridity could probably be found in genes dysregulated in either cybrids or hybrids and which combined effects would be exacerbated in hybrid-cybrids. It can be suggested from these results that the enhanced growth in n[Zu/Col]cCol would result mainly from a combination of several genes dysregulations involved in cell wall loosening and ethylene-induced growth. In opposite, the size decrease in n[Col/Zu]cZu could be due to the downregulation of genes involved in turgor pressure in the endosperm coupled with a downregulation of genes involved in seed storage. Furthermore, the data obtained in this chapter indicate a significant role of cytonuclear relationships in the determination of F1 seed size, especially in a hybrid context. I have shown that the disruption of maternal nuclear and organelle genome evolutionary coupling in *A. thaliana* in a hybrid context significantly exacerbates the effect of nuclear hybridity on seed size, whether it is a positive or a negative one. I also show that the effect of this maternal cytonuclear epistasis in hybrids has quite a different effect in comparison to the same epistasis in an isogenic background. This could arise from cytonuclear incompatibilities arising in

isogenic cybrids where nuclear and organelle genomes come from different lineages of *A. thaliana*, whereas these incompatibilities could be corrected in hybrid-cybrids with the inheritance from the same lineage of both paternally-inherited nuclear genomes and organelle genomes. These Dobzhansky-Muller incompatibilities (Orr 1995) could be at the origin of the large amount of stress response genes found dysregulated in cybrids, but in a much lesser extend in the other samples. However, as I show this newly paternal cytonuclear cooperation to enhance F1 seed size heterosis in hybrid-cybrids, it is important to remind that this heterosis effect can be enhanced in both beneficial and negative way depending of the crossing direction. Thus parental imprinting could possibly also have a role to play in the effects of this switch from maternal to paternal cytonuclear coupling on F1 seed size in enhancing the correction or the worsening of cytonuclear Dobzhansky-Muller incompatibilities depending on the crossing direction. The paternal-of-origin nucleotype thus appears to have a more important role to play in relation to the other two genomes than was previously suggested.

Conclusion

In this chapter, I have investigated the interactions between cybridity and hybridity and their reciprocal effects on F1 seed size in *A. thaliana*. Our data reveals that cybridity creates heterosis on a level comparable to hybridity and that the combination of hybridity and cybridity tend to exacerbate the effect of hybridity or cybridity alone on seed size determination in way depending on the populations involved and the direction of hybridization. Many genes are dysregulated by these hybridization mechanisms, but more genes appear dysregulated when F1 seed size is increased than when it is decreased as a consequence of hybridization. I report that seed size increase is especially associated with a modification of the expression of genes related to cell wall organization and to growth promoting hormones.

As a large number of genes involved in stress response are also found dysregulated as a response to the disruption of cytonuclear relationships and seemingly mostly impacted by a modification of stress response retrograde and anterograde

cytonuclear signalling pathways, it suggests that cybridization would results in seeds being more receptive to stressful conditions, thus impacting F1 seed size. If cybridity and hybridity perform better together than separate in term of resistance to stress as well as in terms of yield, this might prove highly beneficial in the elaboration of new cultivable strains of crops.

Chapter 5 : Plastid Ribosome Protein L5 is Essential for Post-globular Embryo Development in *Arabidopsis thaliana*

Introduction

The plastid is an essential organelle in plant cells acquired through a unique endosymbiosis event in the common ancestor of all *Archaeplastida*, including green plants (Viridiplantae), in which a non-plastid eukaryote absorbed a photosynthetic bacterium (Kishino et al. 1990; Moreira et al. 2000; Stiller 2007; Nowack and Weber 2018). Gene transfer processes have occurred between the original plastid genome (of prokaryote origin) and the nuclear genome following the evolution of this new photosynthetic eukaryotic cell leading most plastid-derived genes to relocate to the nucleus (Martin et al. 2002; McFadden 2014). This transfer is thought to have happened in parallel with the transfer of mitochondrial genes to the nucleus following the endosymbiosis event underlying eukaryotes. Since transferred genes include many of those responsible for the fundamental cellular and metabolic functions of the plastid their protein products needs to be targeted back to plastids by anterograde signalling to assume their function (Bräutigam et al. 2007). Around two thirds of the Plastid Ribosomal Protein (PRPs) genes in the model eudicot *Arabidopsis thaliana* have been transferred over to the nucleus for instance. Most of these genes have also been lost from the plastid genome with a small portion still remaining and thus potentially redundant with their nuclear duplicates (Allen 2018).

Many prokaryote ribosomal proteins (RPs) have been shown to be essential in *E. coli* (Shoji et al. 2011), most being homologs of cyanobacteria-derived plastid RPs (Yamaguchi and Subramanian 2000). Similarly, most of cyanobacteria-derived RPs have also been reported as essential for embryo development in at least one *Embryophyta* species (Table 1). Most essential PRPs are reported as necessary for embryogenesis in *A. thaliana* correlating with the essential functions of plastids in cellular metabolism before the start of photosynthetic activity, notably in lipid and starch biosynthesis (Neuhaus and Emes 2000). The knockout of these PRPs indeed does not allow the embryo to develop further than the globular stage. This phenotype highlights the necessity of plastid translation mechanism as early as globular stage and not before even with the maternal-to-zygote transition occurring as early as the zygote (Zhao et al. 2019).

Nuclear encoded PRPs have been studied in *A. thaliana* with the help of knockouts. The ones which mutations have been shown to lead to seed abortion have been considered as essential for embryo development, summarized in table 1. Essentiality of plastid-encoded PRPs has been shown via a knockout allele introduction in *Nicotiana tabacum* plastid genomes (biolistic chloroplast transformation) and considered essential after the apparition of leaf necrosis phenotype. Among them *PRPS12* is putatively considered as embryo lethal since the splicing of its mRNA is affected by the knockout of the embryo-lethal *AtCAF2* (Asakura and Barkan 2006), but its essentiality has not been directly proven so far.

Not all studied PRPs were reported to be essential for plant survival including PRPS1, -S17, -L24 and -L28. As the knockout of *PRPS17* and *PRPL24* was proven and there does not seem to be any redundancy of these proteins it seems that one can be confident in their non-essentiality. However *prps1-1* is shown to be only a knockdown so it could be argued that the viability of *prps1-1* seeds is due to this leak of expression. Despite being non-essential for embryogenesis, *PRPL28* however appears to be required for seed greening at later stages of embryogenesis since its knockout creates albino seeds which are able to geminate but die quickly after (Romani et al. 2012). A knockout of *PRPS17* was shown to reduce growth rate as well as leaf chlorophyll pigment (Woo et al. 2002). *PRPL24* (Nadine et al. 2012), *PRPS21* and *PRPL11* knockouts also lead to decreased plant size and reduced photosynthetic activity due to a decrease in the translation activity in plastids (Morita-Yamamuro et al. 2004; Pesaresi et al. 2006). *PRPL33* was reported to be required only in cold-stress conditions (Rogalski et al. 2008), and *PRPS15* and PRPL36 for full photosynthetic activity (Fleischmann et al. 2011).

PRP	**Essentiality in *E. coli* (Shoji et al. 2011)**	***A. thaliana* Locus ID**	**Essentiality post Globular Stage in *A. thaliana***	**Reference (*A. thaliana*)**
S1	Essential	AT5G30510	Non-essential	Romani et al. (2012)
S2	Essential	ATCG00160	Essential	Rogalski et al. (2008)
S3	Essential	ATCG00800	Essential	Fleischmann et al. (2011)
S4	Essential	ATCG00380	Essential	Rogalski et al. (2008)
S5	Essential	AT2G33800	Essential	Bryant et al. (2011); Muralla et al. (2011); Lloyd and Meinke (2012)

S7	Essential	ATCG00900/ ATCG01240	NA	NA
S8	Essential	ATCG00770	NA	NA
S9	Non-essential	AT1G74970	Essential	Hsu et al. (2010); Lloyd and Meinke (2012)
S10	Essential	AT3G13120	NA	NA
S11	Essential	ATCG00750	Essential	Muralla et al. (2011); Lloyd and Meinke (2012)
S12	Essential	ATCG00065/ ATCG01230	Putative essential	Asakura and Barkan (2006)
S13	Essential	AT5G14320	Essential	Bryant et al. (2011); Lloyd and Meinke (2012)
S14	Essential	ATCG00330	Essential	Jiang et al. (2018)
S15	NA	ATCG01120	Non-essential	Fleischmann et al. (2011)
S16	Essential	ATCG00050	Essential	Fleischmann et al. (2011)
S17	Non-essential	AT1G79850	Non-essential	Woo et al. 2002; Romani et al. (2012); Lloyd and Meinke (2012)
S18	Essential	ATCG00650	Essential	Rogalski et al. (2006)
S19	Essential	ATCG00820	NA	NA
S20	NA	AT3G15190	Essential	Romani et al. (2012)
S21	NA	AT3G27160	Non-essential	Morita-Yamamuro et al. (2004)
L1	NA	AT3G63490	Essential	Bryant et al. (2011); Romani et al. (2012); Lloyd and Meinke (2012)
L2	Essential	ATCG00830/ ATCG01310	NA	NA
L3	Essential	AT2G43030	NA	NA
L4	Essential	AT1G07320	Essential	Muralla et al. (2011); Lloyd and Meinke (2012)
L5	Essential	AT4G01310	Essential	This chapter
L6	Essential	AT1G05190	Essential	Hsu et al. (2010); Muralla et al. (2011); Lloyd and Meinke (2012)
L7	Essential	NA	NA	NA
L9	NA	AT3G44890	NA	NA
L10	Essential	AT5G13510	Essential	Bryant et al. (2011); Lloyd and Meinke (2012)
L11	NA	AT1G32990	Non-essential	Pesaresi et al. 2001; Lloyd and Meinke (2012)
L12	Essential	AT3G27850	NA	NA

L13	Essential	AT1G78630	Essential	Hsu et al. (2010); Muralla et al. (2011); Lloyd and Meinke (2012)
L14	Essential	ATCG00780	NA	NA
L15	Non-essential	AT3G25920	Essential	Bobik et al. (2018)
L16	Essential	ATCG00790	NA	NA
L17	Essential	AT3G54210	NA	NA
L18	Essential	AT1G48350	Essential	Bryant et al. (2011); Lloyd and Meinke (2012)
L19	Essential	NA	NA	NA
L20	Essential	ATCG00660	Essential	Rogalski et al. (2008)
L21	Non-essential	AT1G35680	Essential	Yin et al. (2012)
L22	Essential	ATCG00810	Essential	Fleischmann et al. (2011)
L23	Essential	ATCG01300/ ATCG00840	Essential	Fleischmann et al. (2011)
L24	Non-essential	AT5G54600	Non-essential	Nadine et al. (2012); Romani et al. (2012)
L25	NA	NA	NA	NA
L27	Non-essential	AT5G40950	Essential	Romani et al. (2012)
L28	Essential	AT2G33450	Essential for greening process and post germination	Romani et al. (2012)
L29	Non-essential	AT5G65220	NA	NA
L30	Non-essential	NA	NA	NA
L31	NA	AT1G75350	Essential	Hsu et al. (2010); Lloyd and Meinke (2012)
L32	NA	ATCG01020	Essential	Fleischmann et al. (2011)
L33	NA	ATCG00640	Non-essential but affects growth in response to cold stress	Rogalski et al. (2008)
L34	Non-essential	AT1G29070	NA	NA
L35	NA	AT2G24090	Essential	Romani et al. (2012)
L36	NA	ATCG00760/ AT5G20180	Non-essential	Fleischmann et al. (2011)

Table 5.1: *A. thaliana* Plastid ribosome subunits and functional homologs in *E.coli*.
Experimentally demonstrated essentiality of Plastid Ribosomal Proteins (PRPs) for embryogenesis and cell survival in *A. thaliana.* The essentiality of plastid ribosomal proteins encoded in the plastid genome has been demonstrated in *N. tabacum*. Each PRP is compared to its homologous ribosomal protein

(RP) in the cyanobacteria-related species *E. coli*. The essentiality of RPs for cell survival in *E. Coli* has been assessed by Shoji et al. (2011).

Some homologs of essential RPs in *E.coli* however still remain to be verified in green plants (PRPS2, -S4, -S7, -S8, -S10, -S11, -S12, -S19, -L2, -L3, -L16, -L17, -L19). However non-essentiality in *E. coli* does not necessarily mean non-essentiality in green plants: while its homolog in *E.coli* was shown as non-essential, *PRPL15* was revealed to be essential for embryo development in *A. thaliana* (Bobik et al. 2018).

In this chapter, I show that PRPL5 is required for post-globular stage embryo development in *A. thaliana* and I characterise the activity its N-terminal Chloroplast transit peptide (cTP) and C-terminal Nuclear localisation signal (NLS) which are completely absent from Chlorophyta and Charophyta homologs proteins.

Materials and methods

Plant material and genomic samples

Arabidopsis thaliana seeds were surface sterilized with Chlorine gas (3:1 bleach:Hydrochloric acid in a bell jar for one hour). Seeds were germinated on 0.5x Murashige and Skoog (MS) medium (Murashige & Skoog, 1962) containing 1% sucrose and 0.8% agar, and grown in a Percival Tissue Culture cabinet under a 16:8 hr light: dark (21 °C/18 °C) regime (Boyes et al. 2001) until they were transferred to soil (five parts Westland compost [Dungannon, N. Ireland]: 1 part perlite: 1 part vermiculite). Plants were grown in chambers under fluorescent lamps at 200 µmol m^{-2} s^{-1} with the same photoperiod.

Plant DNA extraction

Plant genomic DNA was extracted using 20 mg of rosette leaf which was grinded with glass beads and incubated in DNA extraction buffer [200 mM of Tris-HCl pH 7.5, 250 mM of NaCl, 25 mM of EDTA, 0.5% of SDS] for 10 min at 60°C. Samples were next mixed with equal volumes of ice cold isopropanol (1:1) and DNA was precipitated at -20°C for 10 min followed by a centrifugation. Precipitated DNA was washed once

with 70% ethanol, left over to dry, re-suspended in water and incubated at 60°C for 10 minutes.

Genotyping of *prpl5-1* was made using the following primers: Forward 5'-ATCCTCTCGAGGTAAGCGGT - 3', WT reverse 5'- TCTTCTCAGGTCGGTGTGGA - 3' and mutant reverse Lbb1.3 5'-ATTTTGCCGATTTCGGAAC-3'. Loss of 1057 bp in the PRPL5 sequence of *prpl5-2* was genotyped using the following primers: Forward 5'-AGGGGCTAAACGGAAAACTCC-3' and reverse 5'-CACGCGCTAGCTTTTCACG-3'. Presence of the WT allele was verified using the forward primer 5'-TCCAAGGTGTGAGTCCCAGT-3' with the reverse described above.

RNA extraction and PRPL5 *expression analysis*

RNA extraction was performed on 20 mg rosette leaf tissue using the ISOLATE II RNA Plant kit (Bioline, UK). DNase treatment was performed on 1µg of crude RNA using the DNase I amplification grade kit (Invitrogen, UK), and cDNAs were generated using the Superscript III First Strand Synthesis kit (Invitrogen, UK), following protocol supplied by the manufacturer. *PRPL5* expression was evaluated by RT-PCR using forward 5'-TGGGTTTAATCAACAACGAACGC-3' and reverse 5'-TCCAAACCCTTGTCGTTCTG -3' primers and RT-qPCR using forward 5'-TCAACGCCTCAAAACCGCTT-3' and reverse 5'-TCCAAACCCTTGTCGTTCTG-3' primers. *PRPL5* Expression was normalized with *EF-1α* (Wang et al. 2014) using forward 5'-TCACCCTTGGTGTCAAGCAGAT-3' and reverse 5'-CAGGGTTGTATCCGACCTTCTT-3' primers.

Thermal Asymmetric Interlaced PCR

Thermal Asymmetric Interlaced (TAIL) PCR was carried out according to (Liu et al. 2012). Three primers specific to the left border of pROK2 T-DNA were designed according to the publication specifications (the two first primers with overlapping sequences, and the third one at least 50bp downstream): LB1 (5'-ATTTTGCCGATTTCGGAACCACC-3'), LB2 (5'-ACCATCAAACAGGATTTTCGCCTGCT -3'), and LB3 (5'-CCGTCTCACTGGTGAAAAGAAAAACC -3') and the same degenerative

primers AD1 (5'-NTCGASTWTSGWGTT-3'), AD2 (5'-NGTCGASWGANAWGAA -3') and AD3 (5'-WGTGNAGWANCANAGA -3') have been used for all three reactions. Amplicons from the second reaction were extracted from a 2% agarose gel, and sent for Sanger sequencing to LGC Genomics (Ireland). The resulting sequences were aligned against the *Arabidopsis thaliana* genome (TAIR10.1) from the NCBI database.

Cas9-directed mutagenesis of Arabidopsis thaliana *Col-0 line.*

Cas9-directed mutagenesis was performed via the transformation of *A. thaliana* with a newly constructed p3-Cas9-mcherry plasmid vector (supplementary figure S5.2). Construction of the vector was made from the pORE03 backbone base (Coutu et al. 2007). First a *BsaI* site from the Bar resistance gene was removed by converting A to G in the recognition site, conserving the encoded arginine residue; this was performed by *BsaI* and *HpaI* digestion followed by ligation of annealed oligos for sequence replacement. The backbone was then digested with *EcoRI* and *SphI* and ligated to the corresponding pHEE401 (Wang et al. 2015) fragment (a kind gift from Dr. Qi-Jun Chen's lab, China Agricultural Universty, Beijing; Addgene plasmid #71286; http://n2t.net/addgene:71286; RRID:Addgene_71286). A positive selectable marker was then inserted between the *NotI* and *SpeI* sites in the form of an *At2S3* (seed specific) promoter-driven mCherry gene block synthetized by Integrated DNA Technology (Leuven, Belgium), according to the protocol previously described (Gao et al. 2016).

The double guide system for Cas9-directed mutagenesis of At4g01310 was designed according to pre-established protocol (Xing et al. 2014) in order to remove a 1067 bp fragment from the gDNA sequence (Fig. 1a). Both guides were designed using the CRISPR-P tool (Lei et al. 2014) and the obtained double pair of primers was used to amplify the double guide promotor cassette from the vector pCBC-DT1T2 (gift from Qi-Jun Chen, Addgene plasmid #50590, supplementary figure S5.1), as described by Xing et al. (2014):

- DT1-BsF: 5'-ATATATGGTCTCGATTGCAGCACAGGCGCGATAGTGGTT-3' and

DT1-F0: 5'-TGCAGCACAGGCGCGATAGTGGTTTTAGAGCTAGAAATAGC-3' (guide one)

- DT2-BsR: 5'-ATTATTGGTCTCGAAACCTTACCTCGAGAGGATTATCAA-3' and

DT2-R0: 5'-AACCTTACCTCGAGAGGATTATCAATCTCTTAGTCGACTCTAC-3' (guide 2)

The obtained amplicon was cloned into the p3-Cas9-mCheery vector using the Golden Gate reaction with *BsaI* restriction as described previously (Weber et al. 2011) with some modifications: one cycle of 5 min at 32 °C and 5 min at 16 °C for over 5 hours, followed by 5min at 50 °C and 10 min at 80 °C in a Applied Biosystems Veriti 96 well thermal cycler (Thermofisher Scientific, Paisley, UK).

The obtained vector was used to transform *E.coli* DHα electrocompetent cells using the electroporation method. Cells were plated on LB agar with 50 µg/ml kanamycin, and incubated overnight at 37 °C. The vector was purified from a positive colony saturated culture solution using a plasmid extraction kit (Bioline, Dublin, Ireland), and the Cas9 cassette sequence was confirmed by Sanger sequencing (LGC genomics, Berlin, Germany).

Electrocompetent cells of *Agrobacterium tumefasciens* strain GV3101 were transformed with the sequenced vector by the same method, and inoculated onto LB agar plates with 50 µg/ml kanamycin, 50 µg/ml rifampicin and 100 µg/ml gentamycin. Plates were incubated for two days at 28 °C.

Positive colonies were inoculated in LB broth with the same antibiotic concentrations for 2 days at 28 °C. Cells were then centrifuged and re-suspended in LB broth with 5% sucrose and 0.02% Silwett. This liquid culture was then used for *A. thaliana* flower dip method (Clough and Bent 1998), and seeds were harvested at the end of the plant life cycle.

Seeds were screened for transformants using the red fluorescence of the ERFP from the p3-Cas9-mcherry vector under green light, and positive T1 seedlings were grown on soil. Because no mutants were found within the T1 generation, seeds were harvested from T1 lines, screened for ERFP signal, and positive seeds were grown on soil. T2 seedlings were screened for Cas9-induced mutation by PCR, and positive bands where sent over for Sanger sequencing and aligned against *A. thaliana* genome to confirm the removal of the sequence of interest.

Cloning and transformation of A. thaliana

Genomic DNA sequences from *PRPL5* were amplified by Velocity DNA polymerase (Bioline, UK) using gateway-tailed primers and cloned first in pDONR221 and then in destination vector pB7YWG2 using the two-step gateway cloning system via BP and LR clonase reactions (Invitrogen, UK). Four different fragments were generated and cloned under the control of the 35S CamV promoter as fusion proteins with a C-terminal EYFP tag for fluorescence imaging: *PRPL5*, $_{\Delta 233\text{-}262}$*PRPL5*, $_{\Delta 1\text{-}41}$*PRPL5*, and $_{\Delta(1\text{-}41)+(233\text{-}262)}$*PRPL5*. A set of four gateway-tailed primers where used: two primers for the full PRPL5 sequence (Forward 5'-attB1-ATGGCGTCTCCTTCGCTTC-3' and Reverse 5'-ATCTCTTTCCTTTTCCTTTAGCATCAAAG-attP1-3') and two for the N-ter and C-ter truncated sequence (Forward 5'-attB1-ATGGCGTCTGGAACTGTACTGGTC-3' and Reverse 5'-ACCTGAAAGGCATTCCCATTAGAG-attP1-3'). One base was added to the reverse primers in 5' to keep *PRPL5* and *EYFP* in frame. *A. thaliana* Col-0 lines were transformed by floral dip, as described above. Resulting offspring were sprayed with 0.2 µg/ml of Basta to screen for transformants. Efficient transformation was verified by amplifying the YEFP sequence by PCR using forward 5'--3' and reverse 5'--3' primers and transformants with good fluorescence intensity were crossed with the chloroplast reporter line pt-ck (Nelson et al. 2007). Small pieces (1cm square) of leaf tissue from offspring plants were mounted on a microscopic slide in 5 µg/ml DAPI in PBS buffer and visualised with an Olympus BX51 epifluorescence microscope (Dublin, Ireland) with an UV source X-cite Series 120 Q (EXFO, Knightwood, UK). Images were captured with a Leica DFC7000 T camera (Leica microsystems, Ashbourne, Ireland). The same vectors and *A. thaliana* transformation method were used to transform *prpl5-1* and *prpl5-2* mutants for the rescue experiment.

Seed fixation and clearing

Siliques at a specific time point after pollination were extracted and fixed overnight in an ethanol and acetic acid solution (9:1) at 4°C. Resulting fixed siliques were then progressively rehydrated in solutions of decreasing ethanol concentration (10 min incubation per step) and rinsed in water. Siliques were cut open and seeds were mounted on a microscopic slide in Hoyers solution (15 ml distilled water, 3.75g gum

Arabic, 2.5 ml glycerine and 50g chloral hydrate) and left overnight at 4°C for clearing before being imaged by Differential interfering contrast (DIC) under an Olympus BX51 epifluorescence microscope (Dublin, Ireland). Images were captured with a Leica DFC7000 T camera (Leica microsystems, Ashbourne, Ireland).

Sample preparation for Transmission Electron Microscopy (TEM) Analysis

Seeds were harvested at 4 Day after pollination (DAP) from manually self-pollinated flowers from *prpl5-2* +/- line and were fixed with a first solutions [2% glutaraldehyde and 2% paraformaldehyde in 0.1M sodium cacodylate buffer pH 7.2] followed by a second solution [1% osmium tetroxide in 0.1M sodium cacodylate buffer pH 7.2] according to the procedure from NUI Galway Centre for Imaging. The seeds were then progressively dehydrated with increasing concentrations of Ethanol up to 100%, washed in acetone, and finally embedded in resin using the Agar Low Viscosity Resin kit (Agar Scientific, Stansted, UK) according to the manufacturer's protocol, and left for polymerisation at 65 °C for 48 hours. Samples were cut using an ultramicrotome, survey sections of 500 nm width were cut using a glass knife, and were stained with toluidine blue for microscope observation. From these, interesting regions were identified and trimmed to produce 70-90 nm sections using a diamond knife. Obtained sections were stained with uranyl acetate and lead citrate in the Leica EM AC20 automatic stainer (Leica microsystems, Ashbourne, Ireland) and allowed to air dry on a grid before visualisation on a Hitachi 7500 Transmission electron microscope (Hitachi, Daresbury, UK).

Identification of PRPL5 homolog sequences in Chlorophyta *and* Streptophyta.

An initial search was performed by using the PRPL5 sequence from *A. thaliana* as query for a BLAST search in Chlorophyta and Streptophyta (Mount 2007). Resulting sequences were compiled and aligned using MUSCLE (version 3.5, Edgar (2004)) on the phylogeny platform. Phylogeny was established using PhyML (version 3.0,

Guindon et al. (2010)), and the tree was build using TreeDyn (version 196, Chevenet et al. (2006)).

Results

PRPL5 is necessary for seed development in Arabidopsis thaliana

We observed a 23.5 ± 5.2% seed abortion mutant phenotype in the T-DNA line SALK_015079 not segregating with the mutations on *AT3G59380* reported in the line by the Nottingham Arabidopsis Stock Centre (figure 5.2A and 5.2B). The use of the TAIL-PCR technique on DNA extracted from a plant harbouring the phenotype revealed the presence of a so far unreported pROK2 T-DNA insertion in the 5'UTR region of *AT4G01310* (figure 5.1A), which codes for the PLASTID RIBOSOMAL PROTEIN L5 (PRPL5) according to previous proteomic characterisation (Zybailov et al. 2008; Ferro et al. 2010). The segregation of this mutation in the F1 offspring of a self-bred heterozygous parent followed a 1:2 ratio between wild type plants and heterozygous mutants (χ^2 test on 20 plants, p-value = 0.7903), with the total absence of any homozygous mutant offspring. The presence of the mutation correlated exclusively with the presence of the mutant phenotype. To confirm the causality of this insertion I performed a Cas9-directed mutagenesis on Wild-type (WT) Col-0 line using the double guide targeting method (Xing et al. 2014) with the recently produced p3-Cas9-mcherry vector to create a 1057 bp deletion from the *PRPL5* genomic sequence (figure 5.1A and 5.1B). Transformants were isolated from T2 generation seed ERFP fluorescence and the presence of a mutant allele of *PRPL5* was verified by Sanger sequencing. This verification further shows that the 1057 bp removal created a stop codon immediately after the 5' Cas9 cut (figure 5.1C). Offspring were then harvested and transgene-free lines were isolated by screening for non-fluorescent seeds. A 27.0 ± 4.5% seed abortion phenotype was observed from this purified line, agreeing with the segregation ratio previously observed (figure 5.2A and 5.2B). Again, no homozygous mutant could be isolated from the offspring and the presence of the mutation followed a 1:2 distribution for WT and heterozygous mutants (χ^2 test on 60

plants, p-value = 0.7589) and correlated with the presence of the phenotype. The two alleles are hereafter named *prpl5-1* (SALK line) and *prpl5-2* (Cas9-mutated line).

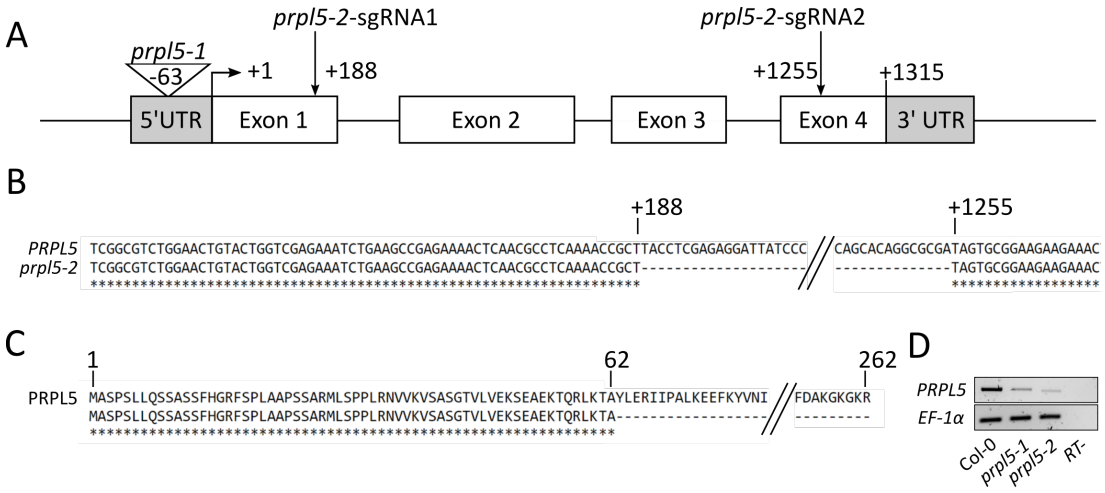


Figure 5.1: Mutant lines used in the chapter.
Mapping and expression of *PRPL5* in the two mutant lines *prpl5-1* and *prpl5-2* used in this chapter. **A.** Insertional mutant line *prpl5-1* and Cas9-generated mutant line *prpl5-2* used in this chapter. **B.** Genomic sequence alignment between *PRPL5* WT sequence and *prpl5-2* mutant sequence validated by Sanger Sequencing. **C.** Predicted amino-acid sequence of PRPL5 aligned with the predicted amino-acid sequence of PRPL5-2 mutant line. **D**. RT-PCR of *PRPL5* expression in both *prpl5-1* and *prpl5-2*.

Lack of PRPL5 in white seeds leads to post-globular embryo arrest

Seeds from self-pollinated *prpl5-2* siliques were extracted at different timepoints from 1 to 6 Days after pollination (DAP) and cleared. Normal embryogenesis was observed in every developing seed up to 3 DAP. After this time point, the embryos of 72.1% of seeds continue to develop normally (figure 5.2C). The other 27.0% of developing seeds remained white after this point (identified as homozygous mutants) and displayed a post globular stage embryo arrest phenotype (figure 5.2D). The remaining 0.9% of the seeds were unfertilized.

To determine whether embryo abortion is associated with aberrant plastid development, seeds from self-pollinated *prpl5-2* siliques were extracted at 4 DAP and processed for Transmission Electron Microscopy (TEM). Phenotypically normal plastids were observed in normally developing embryos (green seeds), including the presence of fully developed thylakoids and grana (figure 5.2C). In contrast, the plastid double envelope was still visible in post-globular arrested embryos, but the plastids were less than half the size of wild-type and appeared to completely lack any thylakoids or grana (figure 5.2D). These results show that a homozygous mutation of

PRPL5 leads to a defective plastid development causing embryo arrest at the globular stage and subsequent seed abortion.

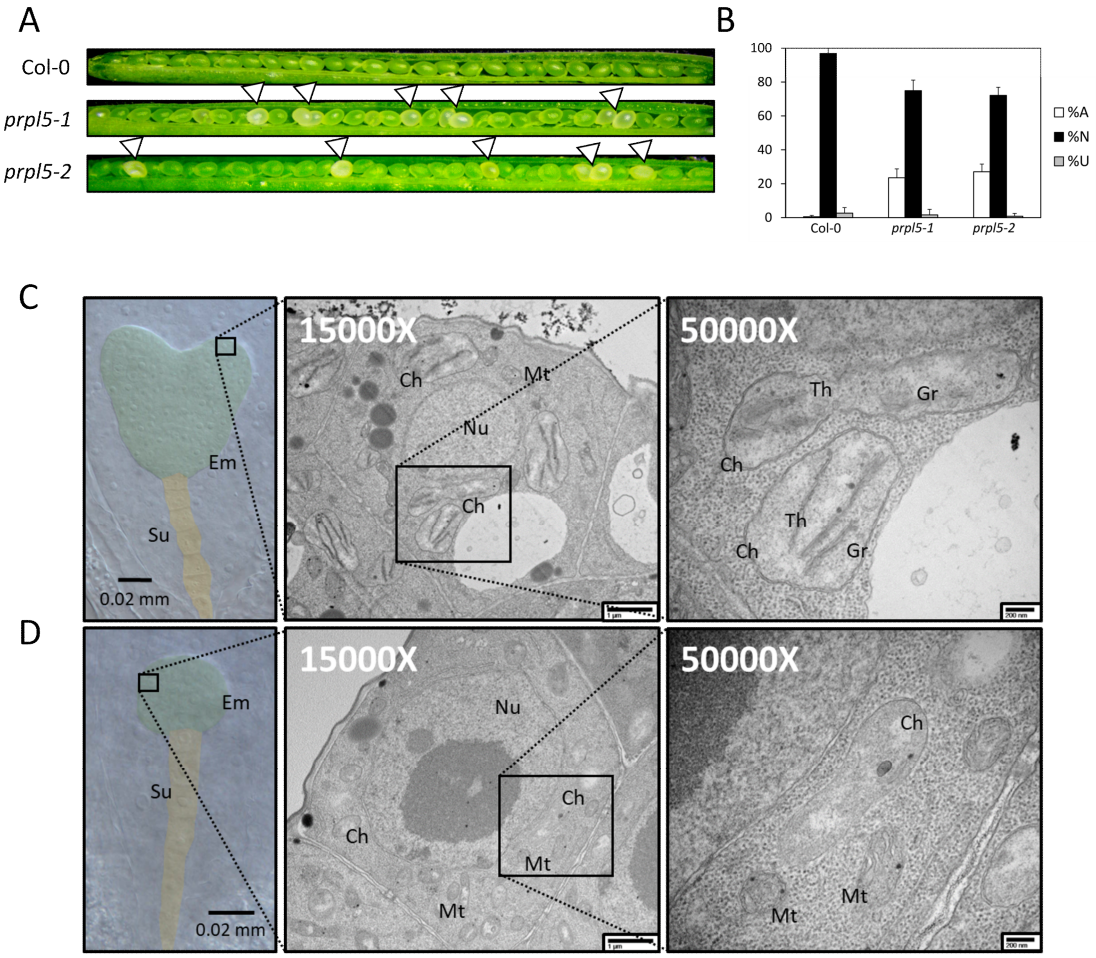


Figure 5.2: Heterozygous mutant of *AT4G01310* leads to 25% seed abortion phenotype.
Seed and plastid phenotype observed in both mutant lines in comparison to WT Col-0. **A**. Developing seeds in siliques at 7 DAP. Aborting seeds are white and indicated by arrows. **B**. Percentage of ANU (Aborted, Normal and Unfertilized) seeds for WT, *prp5-1* and *prpl5-2* lines. **C**. Green seed embryo from *prpl5*-2 at 4 DAP (heart stage) seen in false colours. The embryo has been highlighted in green while the suspensor has been highlighted in yellow. Chloroplasts as fully developed as shown on TEM pictures aside, with presence of thylakoid and grana in the chloroplast matrix. **D**. White seed embryo from *prpl5-2* at 4 DAP (arrested at globular stage) seen in the same false colours as above. Chloroplasts are under-developed as shown on TEM pictures aside, with no thylakoid nor grana being observed in the plastid matrix.

Em: Embryo; Su: Suspensor; Ch: Chloroplast; Nu: Nucleus; Mt: Mitochondria; Th: Thylakoid; Gr: Granum

PRPL5 N-terminal peptide sequence is necessary for plastid localisation, while C-terminal peptide sequence is not

As the nucleus-encoded *PRPL5* gene is essential for plastid development, its protein product must be targeted to the plastid by anterograde signalling early during embryogenesis. Hence, I sought to identify the molecular basis of its intracellular targeting. Bioinformatic analysis of the PRPL5 peptide sequence with ChloroP (Emanuelsson et al. 1999) predicted amino acids 1 to 39 to form a chloroplast transit peptide (cTP) with relative certainty (Score 0.575). In contrast, PROSITE (Sigrist et al. 2002) and cNLS mapper predictor (Kosugi et al. 2009) identified amino acids 245 to 262 to form a bipartite nuclear localisation signal with a middle level score of 3.7 for cNLS mapper (figure 5.3A). Interestingly, analysis of the Cryo-EM structure of the *Spinacea oleracea* plastid ribosome (Perez Boerema et al. 2018) indicates that the amino acids 1-41 at the N-terminus and 233 onwards at the C-terminus sequences are absent from the protein once incorporated into the plastid ribosome (figure 5.3A). This suggests that the nascent PRPL5 polypeptide contains multiple intracellular targeting sequences. To validate these targeting sequences and to determine their function, I used the cleavage sites depicted in PRPL5 from *S. oleracea*, very close to the one in *A. thaliana*, to delimit both N-terminal and C-terminal sequences and study their effect on PRPL5 cellular localisation.

To do so, four constructs were created from the *PRPL5* wild-type sequence and used to stably transform Arabidopsis Col-0 plants. The first construct, i.e. p35S:*PRPL5*:*EYFP*, was used as a positive control for protein localisation. The other three constructs were build to lack either the N-terminal peptide sequence, i.e. p35S:$_{\Delta 1\text{-}41}$*PRPL5*:*EYFP*, the C-terminal peptide sequence, i.e. p35S:$_{\Delta 233\text{-}262}$*PRPL5*:*EYFP*, or both, i.e. p35S:$_{\Delta(1\text{-}41)+(233\text{-}262)}$*PRPL5*:*EYFP*. These constructs were used to stably transform *A. thaliana* Col-0 and transformants with good fluorescence intensity were crossed with the chloroplast reporter line pt-ck (Nelson et al. 2007). Leaf tissue was mounted on microscopic slide with a 5 µg.L^{-1} DAPI solution for fluorescence imaging.

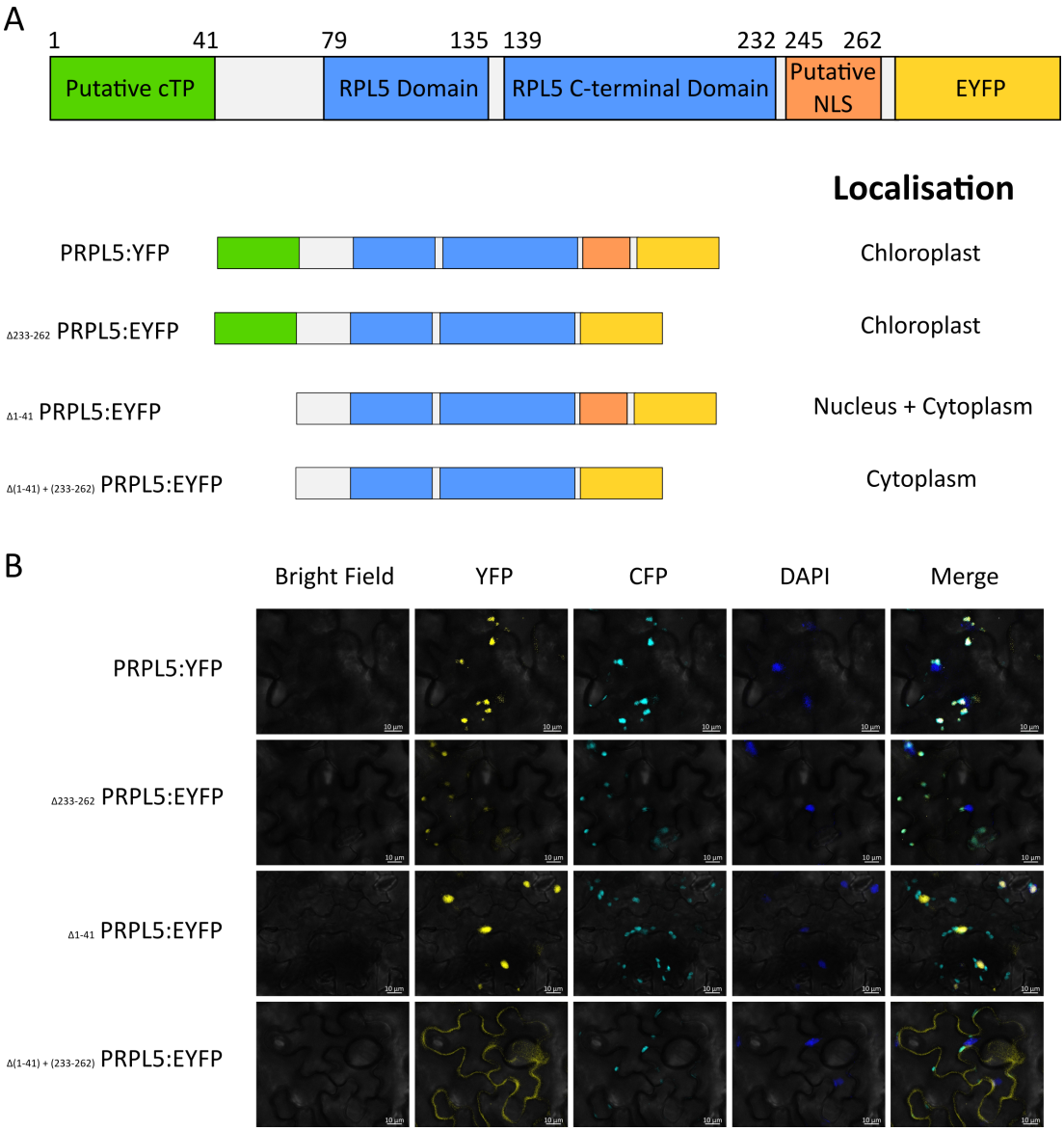


Figure 5.3: Subcellular localisation of PRPL5:YFP product in *A. thaliana* leaf epidermal cells. Subcellular localisation of four different PRPL5:YFP construct within A. thaliana leaf abaxial cells to verify the functionality of predicted N-ter and C-ter signalling peptides. **A**. C-terminal fusion protein of PRPL5 with EYFP (Venus). Description of the localisation of each fusion protein: PRPL5:EYFP (Full protein positive control), $_{\Delta 233\text{-}262}$ PRPL5:EYFP (C-terminal truncated protein), $_{\Delta 1\text{-}41}$ PRPL5:EYFP (N-terminal truncated protein, and $_{\Delta(1\text{-}41)+(233\text{-}262)}$ PRPL5:EYFP (N-ter and C-ter truncated negative control). **B**. EYFP visualisation of protein localisation. PRPL5:EYFP and $_{\Delta 233\text{-}262}$ PRPL5:EYFP co-localise with chloroplast reporting signal, $_{\Delta 1\text{-}41}$ PRPL5:EYFP co-localise with DAPI (nucleus), and $_{\Delta(1\text{-}41)+(233\text{-}262)}$ PRPL5:EYFP localises only in the cytoplasm. Nucleus localisation is displayed using DAPI and chloroplasts localisation is displayed using a CFP construct inherited from the *A. thaliana* pt-ck reporter line. White bar represents 10 µm.

We found that the first PRPL5:EYFP protein construct co-localised with chloroplasts, which is to be expected for a plastid protein (figure 5.3A) and this co-localisation remains effective with $_{\Delta 233\text{-}262}$PRPL5:EYFP lacking the C-terminal targeting sequence (figure 5.3B). However, subcellular localisation $_{\Delta 1\text{-}41}$PRPL5:EYFP lacking the N-terminal

sequence in did not co-localise with chloroplasts but rather localised within the cytosol, as well as in the nucleus (figure 5.3C). Finally, the $_{\Delta(1\text{-}41)+(233\text{-}262)}$PRPL5:EYFP construct lacking both N- and C-terminal peptide sequences localised exclusively in the cytosol (figure 5.3D). Therefore, these results suggest that both predicted targeting sequences are functional although signalisation of PRPL5 to the nucleus appears to only happen in case of an absence of the N-terminal cTP sequence.

The cTP of PRPL5 is essential for the complementation of prpl5-1 *and* prpl5-2 *mutant phenotypes*

The same transformation vectors were used in complementation assays for the 25% seed abortion phenotype observable in *prpl5-1* and *prpl5-2* heterozygotes to confirm that *PRPL5* is responsible for the observed abortion phenotype and to test the functions of the targeting peptides in relation to it (figure 5.4). Both *prpl5-1* and *prpl5-2* mutants were stably transformed with each of the four constructs by floral dipping as above and transformants were selected with Basta. Percentages of fertilized and unfertilized ovules and aborted and normal seeds were assessed for each transformant as previously reported to assess fertility and viability (Duszynska et al. 2019).

The expression of the transgene was also verified by both YFP fluorescence and RT-qPCR. Both PRPL5:EYFP and $_{\Delta233\text{-}262}$PRPL5:EYFP constructs were able to fully restore the WT phenotype in both mutant lines, whereas mutants transformed with $_{\Delta1\text{-}41}$PRPL5:EYFP and $_{\Delta(1\text{-}41)+(233\text{-}262)}$PRPL5:EYFP constructs still displayed 25% seed abortion (figure 5.4B) despite strong expression of the transgenes (figures 5.4C and 5.4D). Therefore, these constructs are unable to complement the mutation, as expected given their lack of plastid localisation. The genotyping of the T2 complemented generation further revealed the presence of homozygous mutants for both *prpl5-1* and *prpl5-2* complemented by PRPL5:EYFP and $_{\Delta233\text{-}262}$PRPL5:EYFP, showing the capability of both construct to rescue the PPRL5 function. Therefore, plastid signalling of PRPL5 with a cTP sequence is mandatory for *A. thaliana* embryogenesis, while nuclear signalling is not required despite being functional.

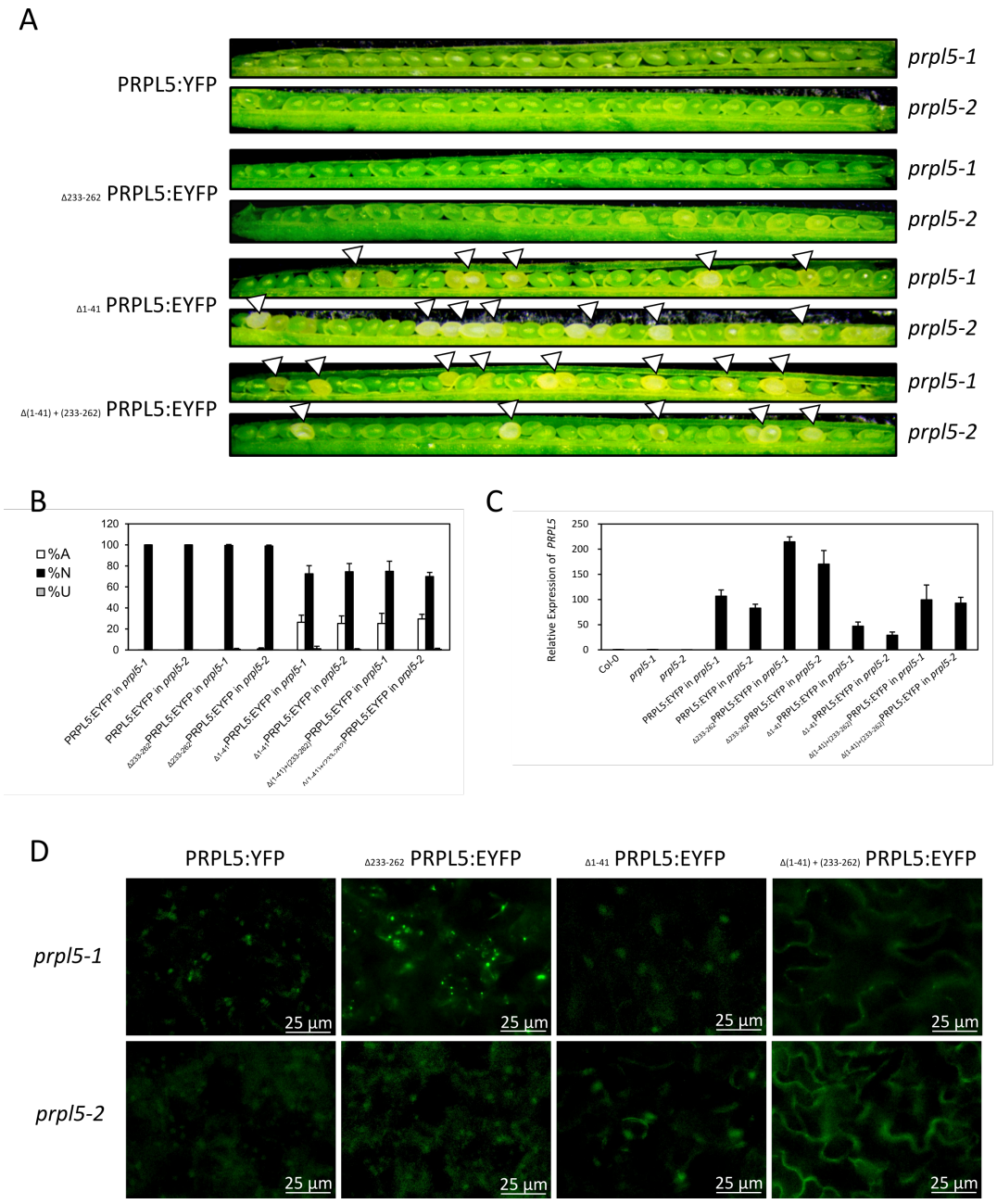


Figure 5.4: Rescue experiment of *prpl5-1* and *prpl5-2*.
Seed phenotype of mutant lines prpl5-1 and prpl5-2 rescued with different constructs of PRPL5:YFP harbouring no, one or both N-ter and C-ter signalling peptides. **A.** Developing seeds in 7 DAP siliques from transformed *prpl5-1* and *prpl5-2* lines. **B.** % of unfertilized ovules (U), and normal and aborted seeds (N, A), for transformed *prp5-1* and *prpl5-2* lines. **C**. RT-qPCR of p35S:*PRPL5:YFP* in leaves of transformed *prpl5-1* and *prpl5-2* plants. **D**. PRPL5:YFP fluorescence in rescued mutant lines.

PRPL5 transitioned from plastid-encoded to nucleus-encoded in the common ancestor of embryophytes.

We decided to investigate the origin on PRPL5 through the evolution of Viridiplantae to better understand the function of both N-terminal and C-terminal signalling

peptides. From the comparison between PRPL5 sequences from Chlorophytes to spermatophytes, it can be noted that the protein is exclusively nuclear encoded in every sequenced species of the *Embryophyta* clad (figure 5.5). Both N-terminal and C-terminal sequences of PRPL5 are quite well conserved in all seed plants (*Spermatophyta*) with circa 40% identity for the N-terminal sequence, and circa 60% identity for the C-terminal sequence. In comparison to *A. thaliana,* the sequences of *Zea mays* and *Ananas comosus* were found to be the most distant with only 28% sequence identity to *A. thaliana* for the N-terminal sequence and 50% for the C-terminal, suggesting some divergence in Monocots. The first and last 10 amino acids of the N-terminal sequence are the most conserved while a RKK_LK_HHF__K_KG motif in the C-terminal sequence is also extremely well conserved among all analysed sequences from *Embryophyta*.

All other groups from *Chlorophyta* and *Charophyta* (all *Streptophyta* species except land plants (Petersen et al. 2006)) have *PRPL5* exclusively encoded in their plastid genomes, indicating that the gene underwent transfer to the nuclear genome in the common ancestor of Embryophytes (indicated with a blue star on figure 5.5).

No conserved N-terminal nor C-terminal sequences comparable to those in *Embryophyta* can be detected in the *PRPL5* sequence from any *Chlorophyta* or *Charophyta* species. The protein sequences of PRPL5 found in *Lycophyta* and *Bryophyta* both possess an N-terminal and C-terminal signalling peptide, but these are not well conserved with the ones found in *Spermatophyta* species. I indeed found only 22% identity for the N-terminal sequence from *Physcomitrium patens* compared to *A. thaliana* and 18% identity for *Selaginella moellendorfii*. The C-terminal sequence harbours also only 20% and 24% identity for the same species. In addition, they appear to be longer in both cases (53-54 amino acids for both *P. patens* and *S. moellendorfii* vs 39 amino acids for *Spermatophyta* species), and the C-terminal motif lacks the *Spermatophyta* conserved motif. Despite these changes, ChloroP still predicts N-terminal sequences as cTPs with a similar certainty as for the *A. thaliana* PRPL5 sequence (scoring at 0.505 for *S. moellendorfii*, 0.581 for *P. patens*). The C-terminal sequence is also predicted by cNLS mapper predictor as being a bipartite NLS in *S. moellendorfii* and *P. patens* albeit with less certainty than for *A. thaliana*

(scores of 2.0 for both). Thus, the transition of PRPL5 from being plastid-encoded to nuclear-encoded after the divergence of Embryophytes from Charophytes can be associated with the co-apparition of both plastid and nuclear localization signals.

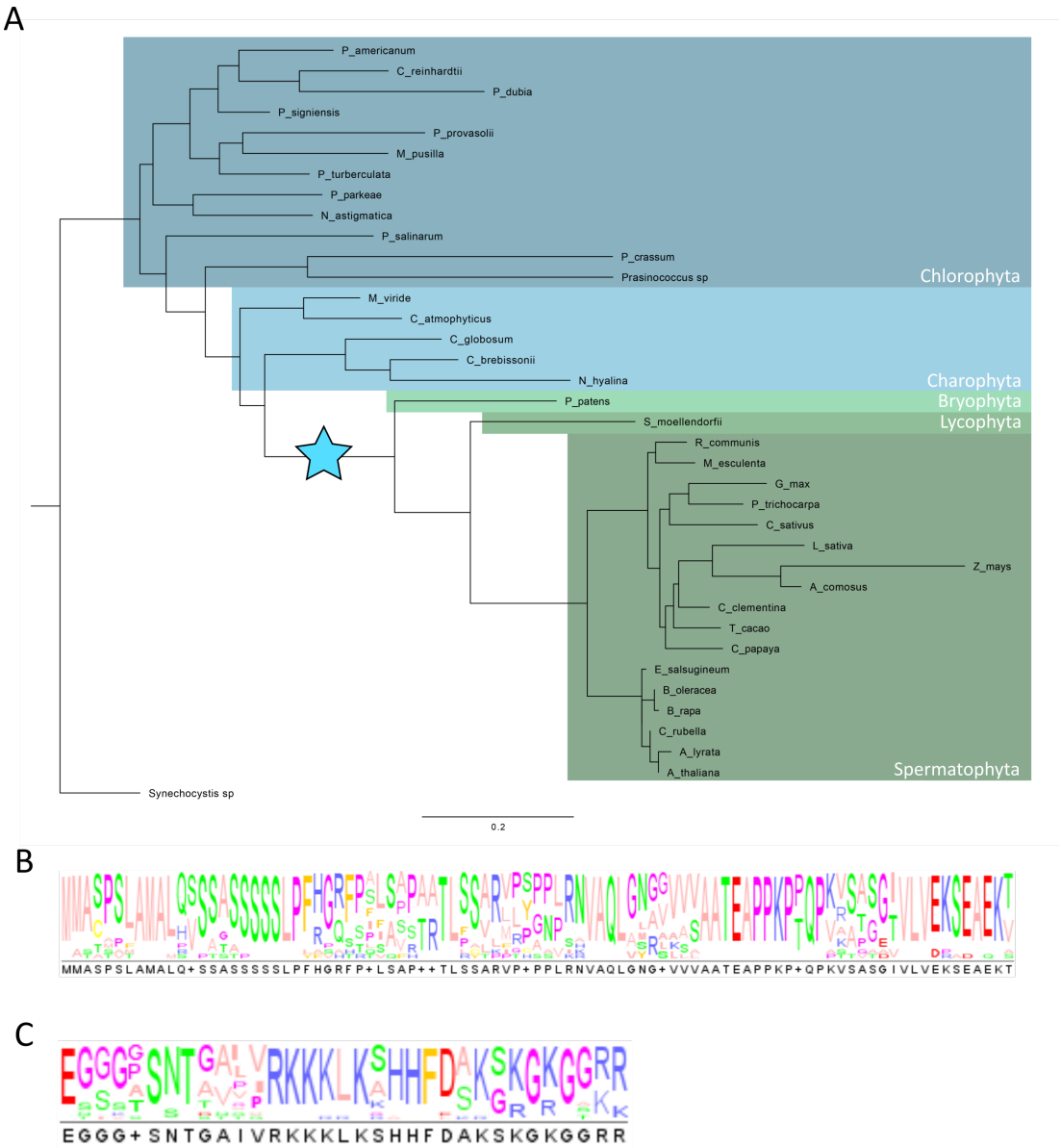


Figure 5.5: Evolution of PRPL5 in *Chlorophyta* and *Embryophyta*.
Study of PRPL5 evolution in Viridiplantae and conservation of N-ter and C-ter signalling peptides across species **A**. Phylogenetic tree of the PRPL5 sequence in Viridiplantae. The tree is rooted on the cyanobacteria species *Synechocystis sp.* (RPLE). Blue star indicates the interval in which *PRPL5* gene underwent transfer to the nuclear genome and corresponding loss from the plastid genome. **B**. Consensus sequence of PRPL5 N-terminal signalling peptides among species displayed in the tree. **C**. Consensus sequence of PRPL5 C-terminal signalling peptides among species displayed in the tree.

Discussion

PRPL5 is required for plastid development

PRPL5 is not the first plastid ribosome protein shown to be embryo-defective in *A. thaliana* (Meinke (2020), cf. Table 1), but so far no verification of its requirement for embryo development had been established, although its bacterial homolog RPLE was reported as essential in *E.coli* (Shoji et al. 2011). As previously observed with most embryo-defective PRPs, a lack of functional PRPL5 does not allow any embryo development past globular stage and leads embryo cells to proliferate ectopically while cotyledons initiation does not occur. As I show by TEM imaging, no thylakoids develop in plastids from either *prpl5-1* or *prpl5-2* homozygous mutants, correlating with the smaller and underdeveloped plastids phenotype reported in knockouts mutants of *PRPL21* (Yin et al. 2012). This plastid requirement for normal embryogenesis is also observed at different protein pathways levels. The knockout of the plastid Glycyl t-RNA transferase *EDD1* also leads to embryo failure, showing the importance of protein translation in plastids at such an early stage (Uwer et al. 1998). The same phenotype can be observed with the knockout of *DLC* (Bellaoui et al. 2003) and the Stromal Processing Peptidase gene *SPP* (Trösch and Jarvis 2011) intervening in the cleavage of plastid transit peptides. More recently, two partially redundant nuclear encoded Chloroplast proteins for Growth and Fertility (genes *CGF1* and *CGF2*) were also reported as embryo-defective as well as being required for normal female gametogenesis (Zhu et al. 2020). This underlines the importance of anterograde protein signalling for the functioning of chloroplasts and the role of chloroplast during early embryogenesis.

An explanation of the requirement of a well-functioning plastid at such an early stage of development prior to any photosynthetic activity seem to be found more particularly in the essentiality of plastid-derived lipid and starch biosynthesis pathways (Neuhaus and Emes 2000). One the genes related to this pathway, namely *accD* coding for a plastid acetyl-CoA Carboxylase and also essential for embryogenesis (Kode et al. 2005), is indeed still located within the plastid genome in *A. thaliana* while all the other gene related to the same pathway have been already transferred to the nuclear genome through evolution. This plastid localisation of the only copy of *accD*

in *A. thaliana* is contrasting with other species such as *Z. mays* (Bryant et al. 2011) and different species from the *Campanulaceae* linage (Rousseau-Gueutin et al. 2013) in which *accD* homolog has undergone transferral to the nuclear genome. In comparison to *A. thaliana*, impairment of plastid translation machinery does not lead to an embryo defect phenotype in *Z. mays* but merely to an impairment of the greening process (Asakura and Barkan 2006; Bryant et al. 2011). This suggest that control of the lipid and starch biosynthesis pathways has been completely transferred over to the nucleus in these species and does not need plastid-related translation anymore to function, while in *A. thaliana* a part of this control is still shared between nucleus and plastid which jeopardizes its functionality in a case of plastid malfunction.

PRPL5 might play a core role in the cohesion of the plastid ribosome

It has to be underlined that, even though most PRPs are required for embryo development, there are also many exceptions. One could therefore hypothesise that some PRPs are more important than others for plastid ribosome activity depending of their position and role within the ribosome structure. As such, the necessity of PRPL5 can be assessed from different interactions it seems to have in the whole complex. Firstly PRPL5 is known to form a heterodimer with PRPL11, this heterodimer binding to the 5S RNA in order to form one of the core elements of the large ribosome subunit (Steitz et al. 1988; Pelava et al. 2016). Secondly, the recent Cryo-EM structure of the spinach chloroplast ribosome shows that PRPL5 binds to RPL31, forming what is called the central protuberance of the 50S ribosomal subunit. PRPL5 further seems to be involved in a bridge between small and large ribosomal subunits via an interaction with PRPS13, interaction strengthened by the association of the Ribosomal pY factor (Ahmed et al. 2016; Bieri et al. 2017; Perez Boerema et al. 2018). Despite not being the only bridge structure observed between both ribosomal subunits, it however appears to be required for plastid ribosome normal function, as mutations of *PRPS13* have also been reported as embryo-defective (Bryant et al. 2011; Lloyd and Meinke 2012). The maintenance of the large subunit central protuberance also seems to be required for ribosome normal activity as *PRPL31* was

also reported as essential for embryogenesis (Hsu et al. 2010). This suggest a central position for PRPL5 in the structural organisation of the plastid ribosomal complex.

Roles of N-terminal and C-terminal peptide sequences in PRPL5

This chapter has confirmed that the first 41 amino acids of PRPL5 act as a cTP matching the predictions made by ChloroP. The plastid localisation of the protein is being supported by mass spectrometry data from isolated chloroplasts (Kleffmann et al. 2004) and further reports added to the PPDB (Zybailov et al. 2008) and AT_CHLORO proteomic databases (Ferro et al. 2010). This N-terminal sequence is well conserved among land plants and seems to be exclusively responsible for PRPL5 localisation within chloroplasts, The presence of the C-terminal peptide in PRPL5 leads to a localisation of the protein within the nucleus, correlating PROSITE and cNLS mapper perdictions. However, I report that such nuclear localisation of PRPL5 only occurs when the N-terminal cTP is disrupted, suggesting that the cTP is dominant over the NLS one. Since truncated version of PRPL5 lacking the C-terminal NLS sequence can still complement both mutant lines used in this chapter it is however unclear which role the NLS peptide could play in PRPL5 function. Comparisons with other nuclear-encoded PRPs also suggests that a C-terminal NLS can be predicted in the majority of them. Only in PRPS1, -L4, -L17 and -L18 could no NLS be predicted. A recent study has demonstrated that the acquisition of a cTP on nascent plastid targeted proteins can usually be attributed to insertions or deletions in nearby genome sequences, followed by substitutions at a lower rate (Christian et al. 2020). Acquisition of a cTP by gene duplication is considered less likely, which would explain the weak alignment of cTP sequences from different nuclear-encoded PRPs within *A.thaliana* (Supplementary figure S5.5A). The same observation can be made with respect to the C-terminal NLS predicted in the majority of PRPs (Supplementary Fig. S5.5B). Hence, it can likely be infered that cTPs and NLSs of the different nuclear-encoded PRPs were obtained from different origins.

However, in comparison to the importance of a cTP for the function of such proteins, which would explain its strong conservation across plant lineages (Fig. 5B), the high

level of conservation of an NLS is puzzling (Fig. 5C). Such a high level of conservation would indicate a conservation of its function across plant lineages, but I could not identify any obvious phenotype in mutant lines rescued with a truncated version of PRPL5 lacking the C-terminal NLS under normal growth conditions. An interesting mechanism which could explain the role of PRPL5 C-terminal NLS would be the relocation of protein in the nucleus to perform a different function under stress conditions, possibly to be recruited by RNA polymerases to maintain genome integrity. This would be possible without the intervention of alternative splicing or post translational modifications, using a mechanism which has been named moonlighting and which uses close contacts between organelles and nucleus to allow such a transfer (Foyer et al. 2020; Krupinska et al. 2020). So far, PRPL5 has been identified in only one crude nuclear lamina protein fraction isolated from *A. thaliana* leaf-derived protoplast (Sakamoto and Takagi 2013) which could correlate with a nuclear re-location under particular stress conditions. Protoplasts indeed harbour an atypical and somewhat artificial cellular environment as cells are disconnected from their usual tissue. The process of protoplast isolation has also been reported to induce an oxidative burst and the activation of oxidative stress responses genes, with an intensity varying between plant species (Papadakis and Roubelakis-Angelakis 1999).

Other PRPs have also been detected in nuclear environments in published nuclear proteomes. For instance, Bae et al. (2003) reports the detection of PRPS5 and of a protein identified as a homolog of PRPL21 within a whole leaf tissue nuclear proteins extract. Both proteins were however identified as less present in the nuclear environment after application of cold stress. Similarly, Goto et al. (2019) reports the presence of PRPS5, -S17 and -L21 in the proteome extracted from cultured cells nuclei, while Sakamoto and Takagi (2013) reports the presence of 24 PRPs in the proteome of protoplast nuclei in addition to PRPL5, namely PRPS1, -S5, -S9, -S13, -S17, -S20, -L1, -L3, -L4, -L6, -L9, -L10, -L11, -L13, -L15, -L17, -L18, -L21, -L24, -L27, -L28, -L29, -L31, and -L35. Such findings may link plastid protein relocation to the nucleus in the particular case of intense stress conditions and might explain the reported impact of some PRPs mutations on plant development only in stressful conditions,

such as for PRPL33 (Rogalski et al. 2008). Hence, it can be considered that PRPL5 might relocate in the nucleus under such conditions. However, any function of PRPL5 in the nuclear compartment is yet to be determined.

Conclusion

We have shown in this chapter that a protein of the central protuberance of the plastid 50S ribosomal subunit, namely PRPL5, is required for embryogenesis past the globular stage in *A. thaliana* indicating that *PRPL5* is critical for the function of the plastid ribosome, and hence for translation. I confirmed the N-terminal and C-terminal ends of the protein to act as reciprocally a cTP and NLS as predicted and further confirmed that the cTP is critical for protein function while the NLS is not. Despite its functionality, the role and the conditions upon which this NLS signal is activated during plant development remains to be determined. The fact that this NLS signal is well conserved in all of its land plant homologues (especially in seed plants) indicates some functional significance. Hence, this chapter further raises the question of the role of NLSs in nuclear-encoded plastid ribosomal protein, which seem to be very well conserved in plant lineages, but paradoxically do not seem to be necessary for plant survival and development in normal growth conditions.

Chapter 6 : General conclusion and perspectives

With this thesis, I aimed at exploring the effects of inter-genomic relationships on seed development and final seed size determination in *A. thaliana*.

In chapter 1 I made a review of the current knowledge on the role parentally-biased genomes into the determination of seed size in *A. thaliana* and other spermatophytes. Seed size, which is majorly determined by the availability of resources coming from the maternal plant, was shown to be heavily impacted by an excess of either maternally- or paternally-inherited genomes in the endosperm. These parental excesses are generally correlated with reciprocally an under-proliferation (maternal excess triploids seeds) and an over-proliferation (paternal excess triploids seeds) of the endosperm which directly impact seed size. Interestingly, the proliferation of the syncytial endosperm and its subsequent cellularization was correlated with a dynamic evolution of PGR balances especially auxin identified as a major drive of endoreduplication and syncytium proliferation. Although most endosperm imprinted genes were reported to not affect seed size, it is interesting to note that key auxin regulating genes were identified as paternally imprinted, and seem to be at the source of the over-proliferation of the endosperm in paternal excess triploid seeds. On the opposite, most genes involved into PRC2-mediated silencing of maternal alleles of iPEGs were revealed as iMEGs themselves and necessary for the maintenance of normal endosperm proliferation.

However, although parental imprinting in endosperm is involved in parental excess-associated phenotypes this mechanism cannot be fully considered as a vector of seed size modification. Other mechanisms were revealed to impact seed size by acting on the embryo development such as gametophyte-inherited factors like in the *YODA* signalling cascade. The effect of cytonuclear interactions on seed development have not been intensively studied as of yet but more recent results suggest an important role of the coevolution of nuclear- and organelle-encoded proteins for embryo development. This highlights the importance of the inheritance of these organelle genomes which are solely maternally inherited in most species.

In chapter 2 I have studied the effect of parental imprinting under positive Darwinan selection (PDS) in the endosperm on seed size determination. I have demonstrated that some of them appear to have an impact on seed size determination in a heterotic

context in *A. thaliana,* which is contrasting with the previous reports of genomic imprinting scarcely affecting seed size. Twelve imprinted genes reported from previous studies as under PDS (10 iPEGs and 2 iMEGs) were tested in hybrid heterotic crosses by reciprocally crossing knockout lines of these genes with a wild-type line (WT) from a different population. Seven of these crosses (6 iPEGs and one iMEG) displayed a modification of the seed size heterosis effect in comparison to WT crosses. Although the one iMEG resulting in seed size modification displayed the effect as expected when its maternal allele was mutated, I obtained puzzling results for the identified iPEGs. Three iPEGs indeed displayed an effect when their paternal allele was mutated and two others when their maternal allele was mutated. The last iPEGs, namely *PKR2*, had an effect on F1 seed size heterosis when either of his alleles was knocked out in crosses between Col-0 and Ler-0. In crosses involving Col-0 and Zu-0 however it had an effect only when the maternal allele was mutated. *PKR2* had already been reported as being involved in the endosperm over-proliferation phenotype from paternal excess triploid seeds. I have shown that a mutation of *PKR2* in hybrid heterotic crosses results in a suppression of the F1 seed size heterosis effect while no significant effect on seed size can be observed when it is mutated in an isogenic context. I also confirmed that *PKR2* is expressed as an iPEG in heterotic crosses involving Col-0 and Ler-0 as well as Col-0 and Zu-0, further showing that *PKR2* effect on F1 seed size is independent of its imprinting in the endosperm.

Transcriptome analysis of Col-0/Zu-0 crosses involving a mutant allele of *PKR2* confirmed the absence of an effect of a paternal mutation of *PKR2* in Zu-0 X Col-0 from the small quantity of DEGs observed in comparison to a mutation of the maternal allele. Our data further show that the seed size modification effect in Col-0 X Zu-0 when the maternal allele of *PKR2* is mutated can be imputable to a dysregulation of genes controlling cell wall organisation in association with a dysregulation of genes involved in gibberellic acid biosynthesis and signalling, suggesting an intervention of the endosperm cellularization process in blocking endosperm proliferation earlier in hybrid crosses involving a mutated allele of *PKR2* thus reducing final seed size.

In accordance with previous results I propose that *PKR2* effect on seed size determination is accentuated by increasing levels of transcriptome dysregulation as can be observed from nuclear hybridization events, from genome imbalance or in lines displaying a disruption of chromatin remodelers activity such as *pkl* and *fis* mutants. *PKR2* can thus to be of high interest in the future to understand phenotypical repercussions of hybridization events on seed development as well as a good candidate gene for seed size and content improvement within the Brassicaceae lineage or in other high plant families by the means of horizontal transfer.

In chapter 3 I have examined the evolution of the Brassicaceae-specific chromatin remodeler *PKR2* in comparison to its ortholog *PKL*. I have shown that even though *PKR2* is evolving fast under PDS its ortholog *PKL* does not, which suggest a mechanism of neofunctionalization of *PKR2* in comparison to *PKL*. Differences of protein structure can already be observed between PKL and PKR2: PKR2 is missing a PHD doamin as well as a portion of the SNF2-related N-terminal domain (ATPase domain) which I further identify as a high disorder-associated loop structure. This loop in PKL appears to be in close vicinity with the C-terminal end of the protein but no particular interaction could be predicted. The disappearance of this structure in PKR2 correlates with a local decrease of disorder within the protein and the release of the protein C-terminal end as well as with a smoothening of the level of disorder within the protein C-terminal end structure. Seven sites in PKR2 were identified as under PDS: three in the vicinity or within Chromo domains, suggesting a positively selected change in the way these domain interact with methylated histone proteins, and the other four found within the SNF2-related N-terminal domain. Further investigation on these sites revealed that five of them are involved in the cohesion of both secondary and tertiary protein structures by the intermediary of hydrogen bonds, whereas the two others are located within loops which don’t possess any particular secondary structure. The fast evolution of PKR2 is thus associated with a global decrease of intrinsic disorder within the protein which was previously reported as being correlated with the selection and fixing of newly acquired functions within the protein.

The analysis of the transcriptome differences between *pkl-1*, *pkr2-2* and *pkr2-2;pkl-10* 4 DAP seeds reveals that PKL and PKR2 control the regulation of different sets of genes despite a significant redundancy on a common gene cluster. I also report that *PKL* and *PKR2* have an antagonistic effect on a specific cluster particularly enriched in genes encoding ubiquitination proteins which suggest an antagonistic role of PKL and PKR2 on post-translational protein modifications and degradation. Finally, enriched dysregulated pathways between mutants suggest a combination of different pathways including the regulation of cell wall organisation in a similar way as described in chapter 2, the control of the endosperm proliferation by CTP synthases and the biosynthesis of fatty acids to explain previously reported seed size variations between Col-0, *pkl-1* and *pkr2-2:pkl-1* lines. PKL and PKR2 thus appear to be an interesting pair of chromatin remodelers impacting seed size determination in *A. thaliana* with PKL probably acting on endosperm proliferation and cellularization processes and PKR2 probably acting mostly on endosperm cellularization in an antagonistic way. It is plausible from our results to assume that PKR2 is under a process of neofunctionalization following a duplication of PKL at the beginning of the Brassicaceae embranchment. However, the role of *PKR2* newly obtained functions on chromatin remodelling are not yet clear.

In chapter 4 I have reported that cybridity, i.e. the disruption of cytonuclear relationships by recombination of nuclear and organelle genomes from different populations, can impact seed size determination and create F1 seed size heterosis in an accession dependant manner. Interestingly, cybridity reveals to be additive or subtractive to heterosis depending on the populations involved. Hybridization and cybridization which are both displaying either best parent heterosis (BPH) or negative heterosis (NH) tend to result in a more important BPH or NH when combined, whereas a combination involving at least one parent displaying either MPH or WPH leads to a heterosis level averaging the ones displayed by hybridization and cybridization alone. This cumulative effect of hybridization and cybridization is further echoed by the apparition of a new cluster of DEGs specific to hybrid-cybrids (lines resulting from a combination of cybridity and hybridity). The discovery of auxin responding genes among this cluster in the hybrid-cybrid n[Zu/Col]cCol correlates the increased seed size found in this line in comparison to n[Zu/Zu]cZu, n[Zu/Zu]cCol and

n[Zu/Col]cZu. Overall, it appears that the variation of F1 seed size observed in cybrids, hybrids and hybrid-cybrids displaying BPH can be attributable to a modification of cell wall organization correlating results found in the preceding chapters. This suggests an intervention of a modification of the endosperm cellularization timely establishment as a cause to this seed size heterosis effect. The variation observed in lines displaying NH however seem rather associated with a decrease in sucrose storage and lipid biosynthesis which would be expected to result in a reduced energy storage accumulation in these seeds, correlating their reduced size.

Organelle genome interactions with nuclear genome are thus not to be ignored in order to understand the mechanisms of F1 seed size heterosis, as they appear to have a non-negligible part to play in the processes involved in seed size determination from seed development stages as early as 3 Days after pollination. Our data also highlights the fact that an important amount of genes involved in stress responses are dysregulated by both hybridity and cybridity, as previously reported. The seed size effects of hybridization and cybridization would thus be expected to be even more significant with plants grown under stressful conditions as can be the case for crops grown in fields with changing weather conditions and the presence of pathogens and pests. It would be very helpful to study the effect of cybridity in combination to hybridity on stress resistance in the process of selection of new crop varieties.

Finally, I have studied in chapter 5 the impact of cytonuclear relationships on seed development with the case study of the Plastid Ribosomal Protein (PRP) L5 (PRPL5). PRPL5 is a nuclear-encoded protein which structurally participates in the cohesion of the large subunit of the plastid ribosome which I report to be necessary for embryo development in *A. thaliana*. This reflects on the necessity of the plastid own translation machinery in the embryo around globular stage, likely due to the necessity of a specific plastid expression of genes involved in lipid metabolism such as *accD* which is critical at that stage. *A. thaliana* and Brassica species in general however appear to be exceptions on the matter as species such as *Z. mays* in which lipid metabolism genes have been completely transferred to the nuclear genome do not need a functioning plastid translation apparatus for embryo development.

We have further reported the presence of a functional Chloroplast transit peptide (cTP) on the N-terminal end of the protein as well as a functional nuclear localisation signal (NLS) on the C-terminal end. However, the NLS peptide only appears to be functional when the cTP is lacking from the protein. This, associated to our data showing that the cTP is necessary in PRPL5 for embryo development whereas the NLS is not, suggest a different function of PRPL5 in the nucleus which conditions might be related to stress and were not met in these experiments. NLS peptides can also be predicted on most of nuclear-encoded PRPs from *A. thaliana*, which shows that a nuclear role of plastid ribosomal proteins might be a common feature of PRPs. It further suggests yet undiscovered functions of PRPs in plants in relation to the nucleus such as a role as transcription factors which if confirmed would deepen our understanding of the evolving relationship between cytoplasmic and nuclear genomes.

Overall, this Ph.D. thesis allowed me to develop a better understanding of how inter-genomic relationships can impact seed development in the plant molecular model *A. thaliana*. Quite interestingly it would appear from these results that the mechanisms of genomic imprinting have a very low impact on how these relationships affect seed side even in hybrids, which is however not too surprising in regards to recently published literature on the subject and raises numerous questions concerning the role of genomic imprinting within the parental conflict model. These results especially suggest that despite the previously reported important role of the endosperm proliferation in the determination of final seed size (as argued in chapter 1), there appears to be other mechanisms at play which regulates this effect. I have indeed shown especially with *pkr2* isogenic mutants that even though these mutants display a reduced endosperm proliferation at early seed development stages this did not significantly impact final seed size. A modification of the timely establishment of endosperm cellularization after heart stage thus would appear to be a good explanation for the modification of final seed size in the lines used in this thesis as it is correlated by the transcriptomics data discussed above. From especially the chapters 2, 3 and 4, this seems to be a common feature in hybrids, cybrids as well as mutants modifying hybridity effects. It will be mostly interesting in the future to study the mechanisms of endosperm cellularization in the light of parental conflict in other

plant species, as I report it to be especially affected by the parental inheritance of the organelles and which of the maternally or paternally inherited nuclear genomes they have been coevolving with. Furthermore, it will be interesting to investigate how these mechanisms are impacted by stressful conditions which are very likely to happen often outside of a controlled growth chamber. The data discussed in this thesis especially sheds more light on the role of cytonuclear relashionships in stress responses, which retrograde signal transduction actors are still in need of extensive investigations today. A potential double action of some organelle encoded genes and organelle proteins indeed appears likely from the data argued in chapter 4 and 5 and could help scientists to better understand retrograde signalling cascades.

In this thesis I have highlighted the presence of yet under-studied molecular interactors and inter-genomic interactions and their non-negligible effects on F1 seed size heterosis which should allow more advances in plant breeding. Further studies on the evolution of these molecular actors across the plant kingdom should also reveal to be very informative in order to better understand the key elements of processes determining seed size and how they could be used at our advantage to speed up the selection of more vigorous, resistant and productive crops.

References

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular Phylogenetics and Evolution 29 (3):380-395. doi:10.1016/S1055-7903(03)00194-5

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46 (W1):W537-W544. doi:10.1093/nar/gky379

Ahmed T, Yin Z, Bhushan S (2016) Cryo-EM structure of the large subunit of the spinach chloroplast ribosome. Scientific Reports 6 (1):35793. doi:10.1038/srep35793

Ahrens J, Dos Santos HG, Siltberg-Liberles J (2016) The Nuanced Interplay of Intrinsic Disorder and Other Structural Properties Driving Protein Evolution. Molecular Biology and Evolution 33 (9):2248-2256. doi:10.1093/molbev/msw092

Ahrens JB, Nunez-Castilla J, Siltberg-Liberles J (2017) Evolution of intrinsic disorder in eukaryotic proteins. Cellular and Molecular Life Sciences 74 (17):3163-3174. doi:10.1007/s00018-017-2559-0

Aichinger E, Villar CBR, Farrona S, Reyes JC, Hennig L, Köhler C (2009) CHD3 Proteins and Polycomb Group Proteins Antagonistically Determine Cell Identity in Arabidopsis. PLOS Genetics 5 (8):e1000605. doi:10.1371/journal.pgen.1000605

Aichinger E, Villar CBR, Mambro RD, Sabatini S, Köhler C (2011) The CHD3 Chromatin Remodeler PICKLE and Polycomb Group Proteins Antagonistically Regulate Meristem Activity in the Arabidopsis Root. The Plant Cell 23 (3):1047-1060. doi:10.1105/tpc.111.083352

Allen JF (2018) Translating photosynthesis. Nature Plants 4 (4):199-200. doi:10.1038/s41477-018-0132-y

Allorent G, Lambert E, Lerbs-Mache S, Courtois F (2010) RNA isolation from developing Arabidopsis thaliana seeds suitable for gene expression analyses. 20:9

Alonso-Blanco C, Vries HB-d, Hanhart CJ, Koornneef M (1999) Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. Proceedings of the National Academy of Sciences 96 (8):4710-4717. doi:10.1073/pnas.96.8.4710

Alonso-Peral MM, Trigueros M, Sherman B, Ying H, Taylor JM, Peacock WJ, Dennis ES (2017) Patterns of gene expression in developing embryos of Arabidopsis hybrids. The Plant Journal 89 (5):927-939. doi:10.1111/tpj.13432

An Y-qC, Goettel W, Han Q, Bartels A, Liu Z, Xiao W (2017) Dynamic Changes of Genome-Wide DNA Methylation during Soybean Seed Development. Scientific Reports 7 (1):1-14. doi:10.1038/s41598-017-12510-4

Anderson ED, Owens JN (1999) Megagametophyte Development, Fertilization, and Cytoplasmic Inheritance in Taxus brevifolia. International Journal of Plant Sciences 160 (3):459-469. doi:10.1086/314143

Anderson SN, Zhou P, Higgins K, Brandvain Y, Springer NM (2020) Widespread imprinting of transposable elements and young genes in the maize endosperm. bioRxiv:2020.2004.2008.032573. doi:10.1101/2020.04.08.032573

Asakura Y, Barkan A (2006) Arabidopsis Orthologs of Maize Chloroplast Splicing Factors Promote Splicing of Orthologous and Species-Specific Group II Introns. Plant Physiology 142 (4):1656-1663. doi:10.1104/pp.106.088096

Asmussen MA, Arnold J, Avise JC (1987) Definition and Properties of Disequilibrium Statistics for Associations Between Nuclear and Cytoplasmic Genotypes. Genetics 115 (4):755-768

Azhagiri AK, Maliga P (2007a) DNA markers define plastid haplotypes in Arabidopsis thaliana. Current Genetics 51 (4):269-275. doi:10.1007/s00294-006-0118-6

Azhagiri AK, Maliga P (2007b) Exceptional paternal inheritance of plastids in Arabidopsis suggests that low-frequency leakage of plastids via pollen may be universal in plants. The Plant Journal 52 (5):817-823. doi:10.1111/j.1365-313X.2007.03278.x

Bae MS, Cho EJ, Choi E-Y, Park OK (2003) Analysis of the Arabidopsis nuclear proteome and its response to cold stress. The Plant Journal 36 (5):652-663. doi:https://doi.org/10.1046/j.1365-313X.2003.01907.x

Batista RA, Figueiredo DD, Santos-González J, Köhler C (2019a) Auxin regulates endosperm cellularization in Arabidopsis. Genes & Development 33 (7-8):466-476. doi:10.1101/gad.316554.118

Batista RA, Köhler C (2020) Genomic imprinting in plants—revisiting existing models. Genes & Development 34 (1-2):24-36. doi:10.1101/gad.332924.119

Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, Köhler C (2019b) The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. eLife 8. doi:10.7554/eLife.50541

Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, Lukowitz W (2009) Paternal Control of Embryonic Patterning in Arabidopsis thaliana. Science 323 (5920):1485-1488. doi:10.1126/science.1167784

Beauzamy L, Fourquin C, Dubrulle N, Boursiac Y, Boudaoud A, Ingram G (2016) Endosperm turgor pressure decreases during early Arabidopsis seed development. Development 143 (18):3295-3299. doi:10.1242/dev.137190

Bellaoui M, Keddie JS, Gruissem W (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. Plant Molecular Biology 53 (4):531-543. doi:10.1023/B:PLAN.0000019061.79773.06

Berger F (1999) Endosperm development. Current Opinion in Plant Biology 2 (1):28-32. doi:10.1016/S1369-5266(99)80006-5

Berkowitz O, De Clercq I, Van Breusegem F, Whelan J (2016) Interaction between hormonal and mitochondrial signalling during growth, development and in plant defence responses. Plant, Cell & Environment 39 (5):1127-1139. doi:10.1111/pce.12712

Bernardi J, Lanubile A, Li Q-B, Kumar D, Kladnik A, Cook SD, Ross JJ, Marocco A, Chourey PS (2012) Impaired Auxin Biosynthesis in the defective endosperm18 Mutant Is Due to Mutational Loss of Expression in the ZmYuc1 Gene Encoding

Endosperm-Specific YUCCA1 Protein in Maize. Plant Physiology 160 (3):1318-1328. doi:10.1104/pp.112.204743

Bieri P, Leibundgut M, Saurer M, Boehringer D, Ban N (2017) The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. The EMBO Journal 36 (4):475-486. doi:10.15252/embj.201695959

Bobik K, Fernandez JC, Hardin SR, Ernest B, Ganusova EE, Staton Margaret E, Burch-Smith TM (2018) The essential chloroplast ribosomal protein uL15c interacts with the chloroplast RNA helicase ISE2 and affects intercellular trafficking through plasmodesmata. New Phytologist 0 (0). doi:10.1111/nph.15427

Boisnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, Bancharel E, Dumas C, Haseloff J, Berger F (2001) Dynamic Analyses of the Expression of the HISTONE::YFP Fusion Protein in Arabidopsis Show That Syncytial Endosperm Is Divided in Mitotic Domains.16

Borges F, Parent J-S, van Ex F, Wolff P, Martínez G, Köhler C, Martienssen RA (2018) Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. Nature Genetics 50 (2):186-192. doi:10.1038/s41588-017-0032-5

Boussardon C, Martin-Magniette M-L, Godin B, Benamar A, Vittrant B, Citerne S, Mary-Huard T, Macherel D, Rajjou L, Budar F (2019) Novel Cytonuclear Combinations Modify Arabidopsis thaliana Seed Physiology and Vigor. Frontiers in Plant Science 10. doi:10.3389/fpls.2019.00032

Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Görlach J (2001) Growth Stage–Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants. The Plant Cell 13 (7):1499-1510. doi:10.1105/TPC.010011

Brandvain Y, Haig D (2018) Outbreeders pull harder in a parental tug-of-war. Proceedings of the National Academy of Sciences 115 (45):11354-11356. doi:10.1073/pnas.1816187115

Bräutigam K, Dietzel L, Pfannschmidt T (2007) Plastid-nucleus communication: anterograde and retrograde signalling inthe development and function of plastids. In: Bock R (ed) Cell and Molecular Biology of Plastids. Topics in Current Genetics. Springer, Berlin, Heidelberg, pp 409-455

Bremner PM, Eckersall RN, Scott RK (1963) The relative importance of embryo size and endosperm size in causing the effects associated with seed size in wheat. The Journal of Agricultural Science 61 (1):139-145. doi:10.1017/S0021859600013800

Bryant N, Lloyd J, Sweeney C, Myouga F, Meinke D (2011) Identification of Nuclear Genes Encoding Chloroplast-Localized Proteins Required for Embryo Development in Arabidopsis. Plant Physiology 155 (4):1678-1689. doi:10.1104/pp.110.168120

Burkart-Waco D, Ngo K, Dilkes B, Josefsson C, Comai L (2013) Early Disruption of Maternal–Zygotic Interaction and Activation of Defense-Like Responses in Arabidopsis Interspecific Crosses. The Plant Cell 25 (6):2037-2055. doi:10.1105/tpc.112.108258

Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear Genomic Interactions and Hybrid Breakdown. Annual Review of Ecology, Evolution, and Systematics 44 (1):281-302. doi:10.1146/annurev-ecolsys-110512-135758

Cao J, Li G, Qu D, Li X, Wang Y (2020) Into the Seed: Auxin Controls Seed Development and Grain Yield. International Journal of Molecular Sciences 21 (5):1662. doi:10.3390/ijms21051662

Carter B, Bishop B, Ho KK, Huang R, Jia W, Zhang H, Pascuzzi PE, Deal R, Ogas J (2018) The Chromatin Remodelers PKL and PIE1 Act in an Epigenetic Pathway that Determines H3K27me3 Homeostasis in Arabidopsis. The Plant Cell:tpc.00867.02017. doi:10.1105/tpc.17.00867

Carter B, Henderson JT, Svedin E, Fiers M, McCarthy K, Smith A, Guo C, Bishop B, Zhang H, Riksen T, Shockley A, Dilkes BP, Boutilier K, Ogas J (2016) Cross-Talk Between Sporophyte and Gametophyte Generations Is Promoted by CHD3 Chromatin Remodelers in Arabidopsis thaliana. Genetics 203 (2):817-829. doi:10.1534/genetics.115.180141

Chandler JW, Werr W (2019) Histology versus phylogeny: Viewing plant embryogenesis from an evo-devo perspective. In: Current Topics in Developmental Biology, vol 131. Elsevier, pp 545-564

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997) Fertilization-independent seed development in Arabidopsis thaliana. Proceedings of the National Academy of Sciences 94 (8):4223-4228

Chen C, Li T, Zhu S, Liu Z, Shi Z, Zheng X, Chen R, Huang J, Shen Y, Luo S, Wang L, Liu Q-Q, E Z (2018a) Characterization of Imprinted Genes in Rice Reveals Conservation of Regulation and Imprinting with Other Plant Species. Plant Physiology 177 (4):1754-1771. doi:10.1104/pp.17.01621

Chen L-Y, Shi D-Q, Zhang W-J, Tang Z-S, Liu J, Yang W-C (2015) The Arabidopsis alkaline ceramidase TOD1 is a key turgor pressure regulator in plant cells. Nature Communications 6 (1):6030. doi:10.1038/ncomms7030

Chen S, Zhou Y, Chen Y, Gu J (2018b) fastp: an ultra-fast all-in-one FASTQ preprocessor. bioRxiv:274100. doi:10.1101/274100

Chen ZJ (2013) Genomic and epigenetic insights into the molecular bases of heterosis. Nature Reviews Genetics 14 (7):471-482. doi:10.1038/nrg3503

Cheng X, Pan M, E Z, Zhou Y, Niu B, Chen C (2021) The maternally expressed polycomb group gene OsEMF2a is essential for endosperm cellularization and imprinting in rice. Plant Communications 2 (1):100092. doi:10.1016/j.xplc.2020.100092

Cheng ZJ, Zhao XY, Shao XX, Wang F, Zhou C, Liu YG, Zhang Y, Zhang XS (2014) Abscisic Acid Regulates Early Seed Development in Arabidopsis by ABI5-Mediated Transcription of SHORT HYPOCOTYL UNDER BLUE1. The Plant Cell 26 (3):1053-1068. doi:10.1105/tpc.113.121566

Chevenet F, Brun C, Bañuls A-L, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7 (1):439. doi:10.1186/1471-2105-7-439

Christian RW, Hewitt SL, Nelson G, Roalson EH, Dhingra A (2020) Plastid transit peptides—where do they come from and where do they all belong? Multi-genome and pan-genomic assessment of chloroplast transit peptide evolution. PeerJ 8:e9772. doi:10.7717/peerj.9772

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium - mediated transformation of Arabidopsis thaliana. The Plant Journal 16 (6):735-743. doi:10.1046/j.1365-313x.1998.00343.x

Coate JE, Schreyer WM, Kum D, Doyle JJ (2020) Robust Cytonuclear Coordination of Transcription in Nascent Arabidopsis thaliana Autopolyploids. Genes 11 (2):134. doi:10.3390/genes11020134

Costa LM, Marshall E, Tesfaye M, Silverstein KAT, Mori M, Umetsu Y, Otterbach SL, Papareddy R, Dickinson HG, Boutiller K, VandenBosch KA, Ohki S, Gutierrez-Marcos JF (2014) Central Cell–Derived Peptides Regulate Early Embryo Patterning in Flowering Plants. Science 344 (6180):168-172. doi:10.1126/science.1243005

Costa Liliana M, Yuan J, Rouster J, Paul W, Dickinson H, Gutierrez-Marcos Jose F (2012) Maternal Control of Nutrient Allocation in Plant Seeds by Genomic Imprinting. Current Biology 22 (2):160-165. doi:10.1016/j.cub.2011.11.059

Coutu C, Brandle J, Brown D, Brown K, Miki B, Simmonds J, Hegedus DD (2007) pORE: a modular binary vector series suited for both monocot and dicot plant transformation. Transgenic Research 16 (6):771-781. doi:10.1007/s11248-007-9066-2

Curtis MD, Grossniklaus U (2008) Molecular control of autonomous embryo and endosperm development. Sexual Plant Reproduction 21 (1):79-88. doi:10.1007/s00497-007-0061-9

da Silva PAWJ (1983) Lysine-ketoglutarate reductase activity in maize: Its possible role in lysine metabolism of developing endosperm. Phytochemistry 22 (12):2687-2689. doi:10.1016/S0031-9422(00)97673-8

Daumann M, Hickl D, Zimmer D, DeTar RA, Kunz H-H, Möhlmann T (2018) Characterization of filament-forming CTP synthases from Arabidopsis thaliana. The Plant Journal: For Cell and Molecular Biology 96 (2):316-328. doi:10.1111/tpj.14032

Day RC, Herridge RP, Ambrose BA, Macknight RC (2008) Transcriptome Analysis of Proliferating Arabidopsis Endosperm Reveals Biological Implications for the Control of Syncytial Division, Cytokinin Signaling, and Gene Expression Regulation. Plant Physiology 148 (4):1964-1984. doi:10.1104/pp.108.128108

Dilkes BP, Comai L (2004) A Differential Dosage Hypothesis for Parental Effects in Seed Development. The Plant Cell 16 (12):3174-3180. doi:10.1105/tpc.104.161230

Dilkes BP, Spielman M, Weizbauer R, Watson B, Burkart-Waco D, Scott RJ, Comai L (2008) The Maternally Expressed WRKY Transcription Factor TTG2 Controls Lethality in Interploidy Crosses of Arabidopsis. PLoS Biology 6 (12). doi:10.1371/journal.pbio.0060308

Duszynska D, McKeown PC, Juenger TE, Pietraszewska-Bogiel A, Geelen D, Spillane C (2013) Gamete fertility and ovule number variation in selfed reciprocal F1 hybrid triploid plants are heritable and display epigenetic parent-of-origin effects. New Phytologist 198 (1):71-81. doi:10.1111/nph.12147

Duszynska D, Vilhjalmsson B, Castillo Bravo R, Swamidatta S, Juenger TE, Donoghue MTA, Comte A, Nordborg M, Sharbel TF, Brychkova G, McKeown PC, Spillane C (2019) Transgenerational effects of inter-ploidy cross direction on reproduction and F2 seed development of Arabidopsis thaliana F1 hybrid triploids. Plant Reproduction 32 (3):275-289. doi:10.1007/s00497-019-00369-6

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32 (5):1792-1797. doi:10.1093/nar/gkh340

Emanuelsson O, Nielsen H, Heijne GV (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8 (5):978-984. doi:10.1110/ps.8.5.978

Erdmann RM, Satyaki PRV, Klosinska M, Gehring M (2017) A Small RNA Pathway Mediates Allelic Dosage in Endosperm. Cell Reports 21 (12):3364-3372. doi:10.1016/j.celrep.2017.11.078

Erdős G, Pajkos M, Dosztányi Z (2021) IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Research 49 (W1):W297-W303. doi:10.1093/nar/gkab408

Etterson JR, Keller SR, Galloway LF (2007) Epistatic and Cytonuclear Interactions Govern Outbreeding Depression in the Autotetraploid Campanulastrum Americanum. Evolution 61 (11):2671-2683. doi:10.1111/j.1558-5646.2007.00234.x

Ferro M, Brugière S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, Kieffer-Jaquinod S, Bruley C, Garin J, Joyard J, Masselon C, Rolland N (2010) AT_CHLORO, a Comprehensive Chloroplast Proteome Database with Subplastidial Localization and Curated Information on Envelope Proteins. Molecular & Cellular Proteomics : MCP 9 (6):1063-1084. doi:10.1074/mcp.M900325-MCP200

Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C (2016) Auxin production in the endosperm drives seed coat development in Arabidopsis. eLife 5:e20542. doi:10.7554/eLife.20542

Figueiredo DD, Batista RA, Roszak PJ, Köhler C (2015) Auxin production couples endosperm development to fertilization. Nature Plants 1 (12):1-6. doi:10.1038/nplants.2015.184

Figueiredo DD, Köhler C (2018) Auxin: a molecular trigger of seed development. Genes & Development 32 (7-8):479-490. doi:10.1101/gad.312546.118

Flanagan JF, Mi L-Z, Chruszcz M, Cymborowski M, Clines KL, Kim Y, Minor W, Rastinejad F, Khorasanizadeh S (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature 438 (7071):1181. doi:10.1038/nature04290

Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schöttler MA, Bock R (2011) Nonessential Plastid-Encoded Ribosomal Proteins in Tobacco: A Developmental Role for Plastid Translation and Implications for Reductive Genome Evolution. The Plant Cell 23 (9):3137-3155. doi:10.1105/tpc.111.088906

Flood PJ, Theeuwen TPJM, Schneeberger K, Keizer P, Kruijer W, Severing E, Kouklas E, Hageman JA, Wijfjes R, Calvo-Baltanas V, Becker FFM, Schnabel SK, Willems LAJ, Ligterink W, Arkel Jv, Mumm R, Gualberto JM, Savage L, Kramer DM, Keurentjes JJB, Eeuwijk Fv, Koornneef M, Harbinson J, Aarts MGM, Wijnker E (2020) Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. Nature Plants:1-9. doi:10.1038/s41477-019-0575-9

Florez-Rueda AM, Paris M, Schmidt A, Widmer A, Grossniklaus U, Städler T (2016) Genomic Imprinting in the Endosperm Is Systematically Perturbed in Abortive Hybrid Tomato Seeds. Molecular Biology and Evolution:msw175. doi:10.1093/molbev/msw175

Fort A, Tuteja R, Braud M, McKeown PC, Spillane C (2017) Parental-genome dosage effects on the transcriptome of F1 hybrid triploid embryos of Arabidopsis thaliana. The Plant Journal 92 (6):1044-1058. doi:10.1111/tpj.13740

Foyer CH, Baker A, Wright M, Sparkes IA, Mhamdi A, Schippers JHM, Van Breusegem F (2020) On the move: redox-dependent protein relocation in plants. Journal of Experimental Botany 71 (2):620-631. doi:10.1093/jxb/erz330

Gao X, Chen J, Dai X, Zhang D, Zhao Y (2016) An Effective Strategy for Reliably Isolating Heritable and Cas9-Free Arabidopsis Mutants Generated by CRISPR/Cas9-Mediated Genome Editing. Plant Physiology 171 (3):1794-1800. doi:10.1104/pp.16.00663

Gehring M, Missirian V, Henikoff S (2011) Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. PLOS ONE 6 (8):e23687. doi:10.1371/journal.pone.0023687

Gerald JNF, Hui PS, Berger F (2009) Polycomb group-dependent imprinting of the actin regulator AtFH5 regulates morphogenesis in Arabidopsis thaliana. Development 136 (20):3399-3404. doi:10.1242/dev.036921

Gibson TA, Goldberg DS (2009) Questioning the Ubiquity of Neofunctionalization. PLOS Computational Biology 5 (1):e1000252. doi:10.1371/journal.pcbi.1000252

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40 (D1):D1178-D1186. doi:10.1093/nar/gkr944

Goto C, Hashizume S, Fukao Y, Hara-Nishimura I, Tamura K (2019) Comprehensive nuclear proteome of Arabidopsis obtained by sequential extraction. Nucleus 10 (SI2):81-92. doi:10.1080/19491034.2019.1603093

Grossniklaus U, Schneitz K (1998) The molecular and genetic basis of ovule and megagametophyte development. Seminars in Cell & Developmental Biology 9 (2):227-238. doi:10.1006/scdb.1997.0214

Grossniklaus U, Vielle-Calzada J-P, Hoeppner MA, Gagliano WB (1998) Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis. Science 280 (5362):446-450. doi:10.1126/science.280.5362.446

Groszmann M, Gonzalez-Bayon R, Lyons RL, Greaves IK, Kazan K, Peacock WJ, Dennis ES (2015) Hormone-regulated defense and stress response networks contribute to heterosis in Arabidopsis F1 hybrids. Proceedings of the National Academy of Sciences 112 (46):E6397-E6406. doi:10.1073/pnas.1519926112

GUIGNARD L (1899) Sur les antherozoides et la double copulation sexuelle chez les vegetaux angiosperms. CR Acad Sci Paris 128:864-871

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology 59 (3):307-321. doi:10.1093/sysbio/syq010

Guitton A-E, Page DR, Chambrier P, Lionnet C, Faure J-E, Grossniklaus U, Berger F (2004) Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Development 131 (12):2971-2981. doi:10.1242/dev.01168

Haig D (2000) The Kinship Theory of Genomic Imprinting. Annual Review of Ecology and Systematics 31 (1):9-32. doi:10.1146/annurev.ecolsys.31.1.9

Haig D (2014) Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. Heredity 113 (2):96-103. doi:10.1038/hdy.2013.97

Haig D, Westoby M (1989) Parent-Specific Gene Expression and the Triploid Endosperm. The American Naturalist 134 (1):147-155. doi:10.1086/284971

Hatorangan MR, Laenen B, Steige KA, Slotte T, Köhler C (2016) Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. The Plant Cell 28 (8):1815-1827. doi:10.1105/tpc.16.00304

Havird JC, Forsythe ES, Williams AM, Werren JH, Dowling DK, Sloan DB (2019) Selfish Mitonuclear Conflict. Current Biology 29 (11):R496-R511. doi:10.1016/j.cub.2019.03.020

He H, Yokoi S, Tezuka T (2020) A high maternal genome excess causes severe seed abortion leading to ovary abscission in Nicotiana interploidy-interspecific crosses. Plant Direct 4 (8):e00257. doi:10.1002/pld3.257

He S, Sun Y, Yang Q, Zhang X, Huang Q, Zhao P, Sun M, Liu J, Qian W, Qin G, Gu H, Qu L-J (2017) A Novel Imprinted Gene NUWA Controls Mitochondrial Function in Early Seed Development in Arabidopsis. PLOS Genetics 13 (1):e1006553. doi:10.1371/journal.pgen.1006553

He T, Sa J, Zhong P-S, Cui Y (2014) Statistical Dissection of Cyto-Nuclear Epistasis Subject to Genomic Imprinting in Line Crosses. PLOS ONE 9 (3):e91702. doi:10.1371/journal.pone.0091702

Hermon P, Srilunchang K-o, Zou J, Dresselhaus T, Danilevskaya ON (2007) Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. Plant Molecular Biology 64 (4):387-395. doi:10.1007/s11103-007-9160-0

Hickl D, Scheuring D, Möhlmann T (2021) CTP Synthase 2 From Arabidopsis thaliana Is Required for Complete Embryo Development. Frontiers in Plant Science 12. doi:10.3389/fpls.2021.652434

Ho KK, Zhang H, Golden BL, Ogas J (2013) PICKLE is a CHD subfamily II ATP-dependent chromatin remodeling factor. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1829 (2):199-210. doi:10.1016/j.bbagrm.2012.10.011

Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, Fischer RL (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. Proceedings of the National Academy of Sciences 108 (5):1755-1762. doi:10.1073/pnas.1019273108

Hsu S-C, F. Belmonte M, J. Harada J, Inoue K (2010) Indispensable Roles of Plastids in Arabidopsis thaliana Embryogenesis. Current Genomics 11 (5):338-349. doi:10.2174/138920210791616716

Huang F, Zhu Q-h, Zhu A, Wu X, Xie L, Wu X, Helliwell C, Chaudhury A, Finnegan EJ, Luo M (2017) Mutants in the imprinted PICKLE RELATED 2 gene suppress seed

abortion of fertilization independent seed class mutants and paternal excess interploidy crosses in Arabidopsis. The Plant Journal 90 (2):383-395. doi:10.1111/tpj.13500

Ingram GC (2020) Family plot: the impact of the endosperm and other extra-embryonic seed tissues on angiosperm zygotic embryogenesis. F1000Research 9:18. doi:10.12688/f1000research.21527.1

Initiative TAG (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408 (6814):796-815. doi:10.1038/35048692

Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, Shimizu B-i, Onishi A, Miyagawa H, Katoh E (2013) Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. Nature Genetics 45 (6):707-711. doi:10.1038/ng.2612

Jacobs SA, Khorasanizadeh S (2002) Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. Science 295 (5562):2080-2083. doi:10.1126/science.1069473

Jahnke S, Scholten S (2009) Epigenetic Resetting of a Gene Imprinted in Plant Embryos. Current Biology 19 (19):1677-1681. doi:10.1016/j.cub.2009.08.053

Jiang H, Moreno-Romero J, Santos-González J, Jaeger GD, Gevaert K, Slijke EVD, Köhler C (2017) Ectopic application of the repressive histone modification H3K9me2 establishes post-zygotic reproductive isolation in Arabidopsis thaliana. Genes & Development. doi:10.1101/gad.299347.117

Jiang T, Zhang J, Rong L, Feng Y, Wang Q, Song Q, Zhang L, Ouyang M (2018) ECD1 functions as an RNA-editing trans-factor of rps14-149 in plastids and is required for early chloroplast development in seedlings. Journal of Experimental Botany 69 (12):3037-3051. doi:10.1093/jxb/ery139

Jing Y, Guo Q, Lin R (2019) The Chromatin-Remodeling Factor PICKLE Antagonizes Polycomb Repression of FT to Promote Flowering1. Plant Physiology 181 (2):656-668. doi:10.1104/pp.19.00596

Josefsson C, Dilkes B, Comai L (2006) Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. Current Biology 16 (13):1322-1328. doi:10.1016/j.cub.2006.05.045

Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ (2013) Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. eLife 2:e00776. doi:10.7554/eLife.00776

Joyard J, Ferro M, Masselon C, Seigneurin-Berny D, Salvi D, Garin J, Rolland N (2010) Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Progress in Lipid Research 49 (2):128-158. doi:10.1016/j.plipres.2009.10.003

Jullien PE, Berger F (2010) Parental Genome Dosage Imbalance Deregulates Imprinting in Arabidopsis. PLOS Genetics 6 (3):e1000885. doi:10.1371/journal.pgen.1000885

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis

D (2021) Highly accurate protein structure prediction with AlphaFold. Nature:1-11. doi:10.1038/s41586-021-03819-2

Kang X, Li W, Zhou Y, Ni M (2013) A WRKY Transcription Factor Recruits the SYG1-Like Protein SHB1 to Activate Gene Expression and Seed Cavity Enlargement. PLOS Genetics 9 (3):e1003347. doi:10.1371/journal.pgen.1003347

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10 (6):845-858. doi:10.1038/nprot.2015.053

Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, Green J, Chen Z, McCuiston J, Wang W, Liebler T, Bullock P, Martin B (2017) MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature 542 (7639):105-109. doi:10.1038/nature20827

Kelliher T, Starr D, Su X, Tang G, Chen Z, Carter J, Wittich PE, Dong S, Green J, Burch E, McCuiston J, Gu W, Sun Y, Strebe T, Roberts J, Bate NJ, Que Q (2019) One-step genome editing of elite crop germplasm during haploid induction. Nature Biotechnology 37 (3):287. doi:10.1038/s41587-019-0038-x

Kermicle JL (1970) Dependence of the R-Mottled Aleurone Phenotype in Maize on Mode of Sexual Transmission. Genetics 66 (1):69-85

Killian JK, Nolan CM, Stewart N, Munday BL, Andersen NA, Nicol S, Jirtle RL (2001) Monotreme IGF2 expression and ancestral origin of genomic imprinting. Journal of Experimental Zoology 291 (2):205-212. doi:10.1002/jez.1070

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14 (4):R36. doi:10.1186/gb-2013-14-4-r36

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T (2004) One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. Science 303 (5657):521-523. doi:10.1126/science.1089835

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA Polycomb Gene in the Arabidopsis Endosperm. The Plant Cell 11 (10):1945-1952. doi:10.1105/tpc.11.10.1945

Kinser TJ, Smith RD, Lawrence AH, Cooley AM, Vallejo-Marin M, Conradi-Smith GD, Puzey JR (2018) Mechanisms driving endosperm-based hybrid incompatibilities: insights from hybrid monkeyflowers. bioRxiv:461939. doi:10.1101/461939

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution 31 (2):151-160. doi:10.1007/BF02109483

Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB, Fischer RL (1999) Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. Proceedings of the National Academy of Sciences 96 (7):4186-4191. doi:10.1073/pnas.96.7.4186

Kleffmann T, Russenberger D, Zychlinski Av, Christopher W, Sjölander K, Gruissem W, Baginsky S (2004) The Arabidopsis thaliana Chloroplast Proteome Reveals Pathway Abundance and Novel Protein Functions. Current Biology 14 (5):354-362. doi:10.1016/j.cub.2004.02.039

Kode V, Mudd EA, Iamtham S, Day A (2005) The tobacco plastid accD gene is essential and is required for leaf development. The Plant Journal 44 (2):237-244. doi:10.1111/j.1365-313X.2005.02533.x

Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W (2003a) Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. The EMBO Journal 22 (18):4804-4814. doi:10.1093/emboj/cdg444

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003b) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes & Development 17 (12):1540-1553. doi:10.1101/gad.257403

Köhler C, Makarevich G (2006) Epigenetic mechanisms governing seed development in plants. EMBO reports 7 (12):1223-1227. doi:10.1038/sj.embor.7400854

Köhler C, Moreno-Romero J, León GDTD, Yadav VK, Santos-González J (2018) Epigenetic signatures associated with imprinted paternally-expressed genes in the Arabidopsis endosperm. bioRxiv:423137. doi:10.1101/423137

Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nature Genetics 37 (1):28-30. doi:10.1038/ng1495

Köhler C, Wolff P, Spillane C (2012) Epigenetic Mechanisms Underlying Genomic Imprinting in Plants. Annual Review of Plant Biology 63 (1):331-352. doi:10.1146/annurev-arplant-042811-105514

Kosugi S, Hasebe M, Tomita M, Yanagawa H (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proceedings of the National Academy of Sciences 106 (25):10171-10176. doi:10.1073/pnas.0900604106

Krupinska K, Blanco NE, Oetke S, Zottini M (2020) Genome communication in plants mediated by organelle–n-ucleus-located proteins. Philosophical Transactions of the Royal Society B: Biological Sciences 375 (1801):20190397. doi:10.1098/rstb.2019.0397

Kuppu S, Tan EH, Nguyen H, Rodgers A, Comai L, Chan SWL, Britt AB (2015) Point Mutations in Centromeric Histone Induce Post-zygotic Incompatibility and Uniparental Inheritance. PLOS Genetics 11 (9):e1005494. doi:10.1371/journal.pgen.1005494

Kutschera U, Niklas KJ (2005) Endosymbiosis, cell evolution, and speciation. Theory in Biosciences 124 (1):1-24. doi:10.1016/j.thbio.2005.04.001

Lafon-Placette C, Hatorangan MR, Steige KA, Cornille A, Lascoux M, Slotte T, Köhler C (2018) Paternally expressed imprinted genes associate with hybridization barriers in Capsella. Nature Plants 4 (6):352-357. doi:10.1038/s41477-018-0161-6

Latta RG, Linhart YB, Mitton JB (2001) Cytonuclear Disequilibrium and Genetic Drift in a Natural Population of Ponderosa Pine. Genetics 158 (2):843-850

Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen L-L (2014) CRISPR-P: A Web Tool for Synthetic Single-Guide RNA Design of CRISPR-System in Plants. Molecular Plant 7 (9):1494-1496. doi:10.1093/mp/ssu044

Lermontova I, Koroleva O, Rutten T, Fuchs J, Schubert V, Moraes I, Koszegi D, Schubert I (2011) Knockdown of CENH3 in Arabidopsis reduces mitotic

divisions and causes sterility by disturbed meiotic chromosome segregation. The Plant Journal 68 (1):40-50. doi:10.1111/j.1365-313X.2011.04664.x

Levesque-Tremblay G, Müller K, Mansfield SD, Haughn GW (2015) HIGHLY METHYL ESTERIFIED SEEDS Is a Pectin Methyl Esterase Involved in Embryo Development. Plant Physiology 167 (3):725-737. doi:10.1104/pp.114.255604

Li C, Sun X, Conover JL, Zhang Z, Wang J, Wang X, Deng X, Wang H, Liu B, Wendel JF, Gong L (2019a) Cytonuclear Coevolution following Homoploid Hybrid Speciation in Aegilops tauschii. Molecular Biology and Evolution 36 (2):341-349. doi:10.1093/molbev/msy215

Li J, Nie X, Tan JLH, Berger F (2013) Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. Proceedings of the National Academy of Sciences 110 (38):15479-15484. doi:10.1073/pnas.1305175110

Li N, Li Y (2015) Maternal control of seed size in plants. Journal of Experimental Botany 66 (4):1087-1097. doi:10.1093/jxb/eru549

Li N, Xu R, Li Y (2019b) Molecular Networks of Seed Size Control in Plants. Annual Review of Plant Biology 70 (1):435-463. doi:10.1146/annurev-arplant-050718-095851

Li S, Castillo-González C, Yu B, Zhang X (2017) The functions of plant small RNAs in development and in stress responses. The Plant Journal 90 (4):654-670. doi:10.1111/tpj.13444

Lin B-Y (1984) Ploidy Barrier to Endosperm Development in Maize. Genetics 107 (1):103-115

Linder B, Cabot RA, Schwickert T, Rupp RAW (2004) The SNF2 domain protein family in higher vertebrates displays dynamic expression patterns in Xenopus laevis embryos. Gene 326:59-66. doi:10.1016/j.gene.2003.09.053

Linksvayer TA, Wade MJ, Gordon DM (2006) GENETIC CASTE DETERMINATION IN HARVESTER ANTS: POSSIBLE ORIGIN AND MAINTENANCE BY CYTO-NUCLEAR EPISTASIS. Ecology 87 (9):2185-2193. doi:10.1890/0012-9658(2006)87[2185:GCDIHA]2.0.CO;2

Liu S, Yeh C-T, Tang HM, Nettleton D, Schnable PS (2012) Gene Mapping via Bulked Segregant RNA-Seq (BSR-Seq). PLOS ONE 7 (5):e36406. doi:10.1371/journal.pone.0036406

Lloyd J, Meinke D (2012) A Comprehensive Dataset of Genes with a Loss-of-Function Mutant Phenotype in Arabidopsis. Plant Physiology 158 (3):1115-1129. doi:10.1104/pp.111.192393

Locascio A, Roig-Villanova I, Bernardi J, Varotto S (2014) Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. Frontiers in Plant Science 5. doi:10.3389/fpls.2014.00412

Lukowitz W, Roeder A, Parmenter D, Somerville C (2004) A MAPKK Kinase Gene Regulates Extra-Embryonic Cell Fate in Arabidopsis. Cell 116 (1):109-119. doi:10.1016/S0092-8674(03)01067-5

Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM (1999) Genes controlling fertilization-independent seed development in Arabidopsis thaliana. Proceedings of the National Academy of Sciences 96 (1):296-301. doi:10.1073/pnas.96.1.296

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A (2011) A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting

Primarily Occurs in the Endosperm. PLoS Genetics 7 (6). doi:10.1371/journal.pgen.1002125

Lur HS, Setter TL (1993) Role of Auxin in Maize Endosperm Development (Timing of Nuclear DNA Endoreduplication, Zein Expression, and Cytokinin). Plant Physiology 103 (1):273-280. doi:10.1104/pp.103.1.273

Luttgeharm KD, Chen M, Mehra A, Cahoon RE, Markham JE, Cahoon EB (2015) Overexpression of Arabidopsis Ceramide Synthases Differentially Affects Growth, Sphingolipid Metabolism, Programmed Cell Death, and Mycotoxin Resistance. Plant Physiology 169 (2):1108-1117. doi:10.1104/pp.15.00987

magnus.nordborg@gmi.oeaw.ac.at GCEa, Consortium G (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell 166 (2):481-491. doi:10.1016/j.cell.2016.05.063

Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, Grossniklaus U, Köhler C (2006) Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO reports 7 (9):947-952. doi:10.1038/sj.embor.7400760

Makarevich G, Villar CBR, Erilova A, Köhler C (2008) Mechanism of PHERES1 imprinting in Arabidopsis. Journal of Cell Science 121 (6):906-912. doi:10.1242/jcs.023077

Makinde AI, Oyekale KO, Daramola DS (2020) Impact of seed size on the seedling vigour, dry matter yield and oil content of jatropha (Jatropha curcas L.). Journal of Agricultural Science (Toronto) 12 (3):197-203

Mao Y, Gabel A, Nakel T, Viehöver P, Baum T, Tekleyohans DG, Vo D, Grosse I, Groß-Hardt R (2020) Selective egg cell polyspermy bypasses the triploid block. eLife 9:e52976. doi:10.7554/eLife.52976

Marín-Rodríguez MC, Orchard J, Seymour GB (2002) Pectate lyases, cell wall degradation and fruit softening. Journal of Experimental Botany 53 (377):2115-2119. doi:10.1093/jxb/erf089

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proceedings of the National Academy of Sciences 99 (19):12246-12251. doi:10.1073/pnas.182432999

Martinez G, Wolff P, Wang Z, Moreno-Romero J, Santos-González J, Conze LL, DeFraia C, Slotkin RK, Köhler C (2018) Paternal easiRNAs regulate parental genome dosage in Arabidopsis. Nature Genetics 50 (2):193-198. doi:10.1038/s41588-017-0033-4

McFadden GI (2014) Origin and Evolution of Plastids and Photosynthesis in Eukaryotes. Cold Spring Harbor Perspectives in Biology 6 (4):a016105. doi:10.1101/cshperspect.a016105

McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37 (1):179-183. doi:10.1016/0092-8674(84)90313-1

McKeown P, Spillane C (2020) An Overview of Current Research in Plant Epigenetic and Epigenomic Phenomena. In: Spillane C, McKeown P (eds) Plant Epigenetics and Epigenomics : Methods and Protocols. Methods in Molecular Biology. Springer US, New York, NY, pp 3-13

McKeown PC, Fort A, Duszynska D, Sulpice R, Spillane C (2013) Emerging molecular mechanisms for biotechnological harnessing of heterosis in crops. Trends in Biotechnology 31 (10):549-551. doi:10.1016/j.tibtech.2013.06.008

McKeown PC, Laouielle-Duprat S, Prins P, Wolff P, Schmid MW, Donoghue MT, Fort A, Duszynska D, Comte A, Lao NT, Wennblom TJ, Smant G, Köhler C, Grossniklaus U, Spillane C (2011) Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thalianaseeds. BMC Plant Biology 11 (1):113. doi:10.1186/1471-2229-11-113

Meinke DW (2020) Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. New Phytologist 226 (2):306-325. doi:10.1111/nph.16071

Meng L-S, Xu M-K, Li D, Zhou M-M, Jiang J-H (2017) Soluble Sugar Accumulation Can Influence Seed Size via AN3–YDA Gene Cascade. Journal of Agricultural and Food Chemistry 65 (20):4121-4132. doi:10.1021/acs.jafc.7b00228

Meng L-S, Xu M-K, Wan W, Wang J-Y (2018) Integration of Environmental and Developmental (or Metabolic) Control of Seed Mass by Sugar and Ethylene Metabolisms in Arabidopsis. Journal of Agricultural and Food Chemistry 66 (13):3477-3488. doi:10.1021/acs.jafc.7b05992

Mi H, Muruganujan A, Casagrande JT, Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. Nature Protocols 8 (8):1551-1566. doi:10.1038/nprot.2013.092

Moison M, Roux F, Quadrado M, Duval R, Ekovich M, Lê D-H, Verzaux M, Budar F (2010) Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in Arabidopsis thaliana. The Plant Journal 63 (5):728-738. doi:10.1111/j.1365-313X.2010.04275.x

Moreira D, Guyader HL, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. Nature 405 (6782):69-72. doi:10.1038/35011054

Moreno-Pérez AJ, Venegas-Calerón M, Vaistij FE, Salas JJ, Larson TR, Garcés R, Graham IA, Martínez-Force E (2012) Reduced expression of FatA thioesterases in Arabidopsis affects the oil content and fatty acid composition of the seeds. Planta 235 (3):629-639. doi:10.1007/s00425-011-1534-5

Morita-Yamamuro C, Tsutsui T, Tanaka A, Yamaguchi J (2004) Knock-out of the Plastid Ribosomal Protein S21 Causes Impaired Photosynthesis and Sugar-Response during Germination and Seedling Development in Arabidopsis thaliana. Plant and Cell Physiology 45 (6):781-788. doi:10.1093/pcp/pch093

Mosca R, Pache RA, Aloy P (2012) The Role of Structural Disorder in the Rewiring of Protein Interactions through Evolution *. Molecular & Cellular Proteomics 11 (7):M111.014969-M014111.014969-014968. doi:10.1074/mcp.M111.014969

Mount DW (2007) Using the Basic Local Alignment Search Tool (BLAST). Cold Spring Harbor Protocols 2007 (7):pdb.top17. doi:10.1101/pdb.top17

Mozgova I, Hennig L (2015) The Polycomb Group Protein Regulatory Network. Annual Review of Plant Biology 66 (1):269-296. doi:10.1146/annurev-arplant-043014-115627

Muir G, Ruiz-Duarte P, Hohmann N, Mable BK, Novikova P, Schmickl R, Guggisberg A, Koch MA (2015) Exogenous selection rather than cytonuclear incompatibilities shapes asymmetrical fitness of reciprocal Arabidopsis hybrids. Ecology and Evolution 5 (8):1734-1745. doi:10.1002/ece3.1474

Muralla R, Lloyd J, Meinke D (2011) Molecular Foundations of Reproductive Lethality in Arabidopsis thaliana. PLOS ONE 6 (12):e28398. doi:10.1371/journal.pone.0028398

Murawska M, Kunert N, van Vugt J, Längst G, Kremmer E, Logie C, Brehm A (2008) dCHD3, a Novel ATP-Dependent Chromatin Remodeler Associated with Sites of Active Transcription. Molecular and Cellular Biology 28 (8):2745-2757. doi:10.1128/MCB.01839-07

Musielak TJ, Bayer M (2014) YODA signalling in the early Arabidopsis embryo. Biochemical Society Transactions 42 (2):408-412. doi:10.1042/BST20130230

Muyle A, Zemp N, Fruchard C, Cegan R, Vrana J, Deschamps C, Tavares R, Hobza R, Picard F, Widmer A, Marais GAB (2018) Genomic imprinting mediates dosage compensation in a young plant XY system. Nature Plants 4 (9):677-680. doi:10.1038/s41477-018-0221-y

Nadine T, Magdalena W, Wolfram T, Eugenia M, A. SM, Ralph B (2012) The plastid-specific ribosomal proteins of Arabidopsis thaliana can be divided into non-essential proteins and genuine ribosomal proteins. The Plant Journal 69 (2):302-316. doi:10.1111/j.1365-313X.2011.04791.x

Nawaschin S (1898) Resultate einer Revision der Befruchtungsvorg"ange bei Lilium Martagon und Fritillaria tenella.7

Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. The Plant Journal 51 (6):1126-1136. doi:10.1111/j.1365-313X.2007.03212.x

Neuhaus HE, Emes MJ (2000) Nonphotosynthetic Metabolism in Plastids. Annual Review of Plant Physiology and Plant Molecular Biology 51 (1):111-140. doi:10.1146/annurev.arplant.51.1.111

Ng S, De Clercq I, Van Aken O, Law SR, Ivanova A, Willems P, Giraud E, Van Breusegem F, Whelan J (2014) Anterograde and Retrograde Regulation of Nuclear Genes Encoding Mitochondrial Proteins during Growth, Development, and Stress. Molecular Plant 7 (7):1075-1093. doi:10.1093/mp/ssu037

Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, Kouzarides T, Murzin AG, Murzina NV, Laue ED (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature 416 (6876):103-107. doi:10.1038/nature722

Nishiyama I, Inomata N (1966) EMBRYOLOGICAL STUDIES ON CROSS-INCOMPATIBILITY BETWEEN 2x AND 4x IN *BRASSICA*. 遺伝學雜誌 41 (1):27-42. doi:10.1266/jjg.41.27

Nott A, Jung H-S, Koussevitzky S, Chory J (2006) Plastid-to-Nucleus Retrograde Signaling. Annual Review of Plant Biology 57 (1):739-759. doi:10.1146/annurev.arplant.57.032905.105310

Nowack ECM, Weber APM (2018) Genomics-Informed Insights into Endosymbiotic Organelle Evolution in Photosynthetic Eukaryotes. Annual Review of Plant Biology 69 (1):51-84. doi:10.1146/annurev-arplant-042817-040209

Ogas J, Cheng J-C, Sung ZR, Somerville C (1997) Cellular Differentiation Regulated by Gibberellin in the Arabidopsis thaliana pickle Mutant. Science 277 (5322):91-94. doi:10.1126/science.277.5322.91

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative

development in Arabidopsis. Proceedings of the National Academy of Sciences 96 (24):13839-13844. doi:10.1073/pnas.96.24.13839

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL (1999) Mutations in FIE, a WD Polycomb Group Gene, Allow Endosperm Development without Fertilization. The Plant Cell 11 (3):407-415. doi:10.1105/tpc.11.3.407

Olsen O-A (2020) The Modular Control of Cereal Endosperm Development. Trends in Plant Science 25 (3):279-290. doi:10.1016/j.tplants.2019.12.003

Oñate-Sánchez L, Vicente-Carbajosa J (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. BMC Research Notes 1 (1):93. doi:10.1186/1756-0500-1-93

Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L (2015) Networks controlling seed size in Arabidopsis. Plant Reproduction 28 (1):17-32. doi:10.1007/s00497-015-0255-5

Orr HA (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139 (4):1805-1813. doi:10.1093/genetics/139.4.1805

Păcurar DI, Păcurar ML, Street N, Bussell JD, Pop TI, Gutierrez L, Bellini C (2012) A collection of INDEL markers for map-based cloning in seven Arabidopsis accessions. Journal of Experimental Botany 63 (7):2491-2501. doi:10.1093/jxb/err422

Paolo D, Rotasperti L, Schnittger A, Masiero S, Colombo L, Mizzotti C (2021) The Arabidopsis MADS-Domain Transcription Factor SEEDSTICK Controls Seed Size via Direct Activation of E2Fa. Plants (Basel, Switzerland) 10 (2). doi:10.3390/plants10020192

Papadakis AK, Roubelakis-Angelakis KA (1999) The Generation of Active Oxygen Species Differs in Tobacco and Grapevine Mesophyll Protoplasts. Plant Physiology 121 (1):197-206. doi:10.1104/pp.121.1.197

Pechmann S, Frydman J (2014) Interplay between Chaperones and Protein Disorder Promotes the Evolution of Protein Networks. PLOS Computational Biology 10 (6):e1003674. doi:10.1371/journal.pcbi.1003674

Pelava A, Schneider C, Watkins NJ (2016) The importance of ribosome production, and the 5S RNP–MDM2 pathway, in health and disease. Biochemical Society Transactions 44 (4):1086-1090. doi:10.1042/BST20160106

Perez Boerema A, Aibara S, Paul B, Tobiasson V, Kimanius D, Forsberg BO, Wallden K, Lindahl E, Amunts A (2018) Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. Nature Plants 4 (4):212-217. doi:10.1038/s41477-018-0129-6

Pesaresi P, Masiero S, Eubel H, Braun H-P, Bhushan S, Glaser E, Salamini F, Leister D (2006) Nuclear Photosynthetic Gene Expression Is Synergistically Modulated by Rates of Protein Synthesis in Chloroplasts and Mitochondria. The Plant Cell 18 (4):970-991. doi:10.1105/tpc.105.039073

Petersen J, Teich R, Becker B, Cerff R, Brinkmann H (2006) The GapA/B Gene Duplication Marks the Origin of Streptophyta (Charophytes and Land Plants). Molecular Biology and Evolution 23 (6):1109-1118. doi:10.1093/molbev/msj123

Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M (2014) Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife 3:e03198. doi:10.7554/eLife.03198

Pignatta D, Novitzky K, Satyaki PRV, Gehring M (2018) A variably imprinted epiallele impacts seed development. bioRxiv:339036. doi:10.1101/339036

Raghavan V (2003) Some reflections on double fertilization, from its discovery to the present. New Phytologist 159 (3):565-583. doi:10.1046/j.1469-8137.2003.00846.x

Ramsey AJ, McCauley DE, Mandel JR (2019) Heteroplasmy and Patterns of Cytonuclear Linkage Disequilibrium in Wild Carrot. Integrative and Comparative Biology 59 (4):1005-1015. doi:10.1093/icb/icz102

Rand DM, Haney RA, Fry AJ (2004) Cytonuclear coevolution: the genomics of cooperation. Trends in Ecology & Evolution 19 (12):645-653. doi:10.1016/j.tree.2004.10.003

Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464 (7288):615-618. doi:10.1038/nature08842

Ravi M, Marimuthu MPA, Tan EH, Maheshwari S, Henry IM, Marin-Rodriguez B, Urtecho G, Tan J, Thornhill K, Zhu F, Panoli A, Sundaresan V, Britt AB, Comai L, Chan SWL (2014) A haploid genetics toolbox for Arabidopsis thaliana. Nature Communications 5 (1):1-8. doi:10.1038/ncomms6334

Rebernig CA, Lafon-Placette C, Hatorangan MR, Slotte T, Köhler C (2015) Non-reciprocal Interspecies Hybridization Barriers in the Capsella Genus Are Established in the Endosperm. PLOS Genetics 11 (6):e1005295. doi:10.1371/journal.pgen.1005295

Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nature Reviews Genetics 2 (1):21-32. doi:10.1038/35047554

Riefler M, Novak O, Strnad M, Schmülling T (2006) Arabidopsis Cytokinin Receptor Mutants Reveal Functions in Shoot Growth, Leaf Senescence, Seed Size, Germination, Root Development, and Cytokinin Metabolism. The Plant Cell 18 (1):40-54. doi:10.1105/tpc.105.037796

Rodermel SR, Abbott MS, Bogorad L (1988) Nuclear-organelle interactions: Nuclear antisense gene inhibits ribulose bisphosphate carboxylase enzyme levels in transformed tobacco plants. Cell 55 (4):673-681. doi:10.1016/0092-8674(88)90226-7

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D (2013) Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proceedings of the National Academy of Sciences 110 (19):7934-7939. doi:10.1073/pnas.1306164110

Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Research 34 (16):4537-4545. doi:10.1093/nar/gkl634

Rogalski M, Schöttler MA, Thiele W, Schulze WX, Bock R (2008) Rpl33, a Nonessential Plastid-Encoded Ribosomal Protein in Tobacco, Is Required under Cold Stress Conditions. The Plant Cell 20 (8):2221-2237. doi:10.1105/tpc.108.060392

Romani I, Tadini L, Rossi F, Masiero S, Pribil M, Jahns P, Kater M, Leister D, Pesaresi P (2012) Versatile roles of Arabidopsis plastid ribosomal proteins in plant

growth and development. The Plant Journal 72 (6):922-934. doi:10.1111/tpj.12000

Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN (2013) Potential Functional Replacement of the Plastidic Acetyl-CoA Carboxylase Subunit (accD) Gene by Recent Transfers to the Nucleus in Some Angiosperm Lineages. Plant Physiology 161 (4):1918-1929. doi:10.1104/pp.113.214528

Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, Durand S, Villoutreix R, Martin-Magniette M-L, Camilleri C, Budar F (2016) Cytonuclear interactions affect adaptive traits of the annual plant Arabidopsis thaliana in the field. Proceedings of the National Academy of Sciences 113 (13):3687-3692. doi:10.1073/pnas.1520687113

Ryder P, McKeown PC, Fort A, Spillane C (2014) Epigenetics and Heterosis in Crop Plants. In: Alvarez-Venegas R, De la Peña C, Casas-Mollano JA (eds) Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications: Transcriptional Regulation and Chromatin Remodelling in Plants. Springer International Publishing, Cham, pp 13-31

Sackton TB, Haney RA, Rand DM (2003) Cytonuclear Coadaptation in Drosophila: Disruption of Cytochrome C Oxidase Activity in Backcross Genotypes. Evolution 57 (10):2315-2325. doi:10.1111/j.0014-3820.2003.tb00243.x

Sakamoto Y, Takagi S (2013) LITTLE NUCLEI 1 and 4 Regulate Nuclear Morphology in Arabidopsis thaliana. Plant and Cell Physiology 54 (4):622-633. doi:10.1093/pcp/pct031

Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH (2008) Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. Ecology Letters 11 (10):1082-1091. doi:10.1111/j.1461-0248.2008.01224.x

Samel A, Cuomo A, Bonaldi T, Ehrenhofer-Murray AE (2012) Methylation of CenH3 arginine 37 regulates kinetochore integrity and chromosome segregation. Proceedings of the National Academy of Sciences 109 (23):9029-9034. doi:10.1073/pnas.1120968109

Sandve SR, Rohlfs RV, Hvidsten TR (2018) Subfunctionalization versus neofunctionalization after whole-genome duplication. Nature Genetics 50 (7):908-909. doi:10.1038/s41588-018-0162-4

Schatlowski N, Wolff P, Santos-González J, Schoft V, Siretskiy A, Scott R, Tamaru H, Köhler C (2014) Hypomethylated Pollen Bypasses the Interploidy Hybridization Barrier in Arabidopsis[C][W]. The Plant Cell 26 (9):3556-3568. doi:10.1105/tpc.114.130120

Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B (2011) Protein disorder—a breakthrough invention of evolution? Current Opinion in Structural Biology 21 (3):412-418. doi:10.1016/j.sbi.2011.03.014

Schnable PS, Springer NM (2013) Progress Toward Understanding Heterosis in Crop Plants. Annual Review of Plant Biology 64 (1):71-88. doi:10.1146/annurev-arplant-042110-103827

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends in Plant Science 3 (5):175-180. doi:10.1016/S1360-1385(98)01235-7

Schnepf E, Brown RM (1971) On Relationships between Endosymbiosis and the Origin of Plastids and Mitochondria. In: Reinert J, Ursprung H (eds) Origin and

Continuity of Cell Organelles. Results and Problems in Cell Differentiation. Springer, Berlin, Heidelberg, pp 299-322

Schon MA, Nodine M (2017) Widespread contamination of Arabidopsis embryo and endosperm transcriptome datasets. The Plant Cell:tpc.00845.02016. doi:10.1105/tpc.16.00845

Scott RJ, Spielman M, Bailey J, Dickinson HG (1998) Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125 (17):3329-3341

Sergushichev AA, Loboda AA, Jha AK, Vincent EE, Driggers EM, Jones RG, Pearce EJ, Artyomov MN (2016) GAM: a web-service for integrated transcriptional and metabolic network analysis. Nucleic Acids Research 44 (W1):W194-W200. doi:10.1093/nar/gkw266

Sew YS, Ströher E, Fenske R, Millar AH (2016) Loss of Mitochondrial Malate Dehydrogenase Activity Alters Seed Metabolism Impairing Seed Maturation and Post-Germination Growth in Arabidopsis1[OPEN]. 171:15

Shoji S, Dambacher CM, Shajani Z, Williamson JR, Schultz PG (2011) Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes. Journal of Molecular Biology 413 (4):751-761. doi:10.1016/j.jmb.2011.09.004

Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) PROSITE: A documented database using patterns and profiles as motif descriptors. Briefings in Bioinformatics 3 (3):265-274. doi:10.1093/bib/3.3.265

Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC (2018) Cytonuclear integration and co-evolution. Nature Reviews Genetics 19 (10):635-648. doi:10.1038/s41576-018-0035-9

Sobanski J, Giavalisco P, Fischer A, Kreiner JM, Walther D, Schöttler MA, Pellizzer T, Golczyk H, Obata T, Bock R, Sears BB, Greiner S (2019) Chloroplast competition is controlled by lipid biosynthesis in evening primroses. Proceedings of the National Academy of Sciences 116 (12):5665-5674. doi:10.1073/pnas.1811661116

Soltani A, Kumar A, Mergoum M, Pirseyedi SM, Hegstad JB, Mazaheri M, Kianian SF (2016) Novel nuclear-cytoplasmic interaction in wheat (Triticum aestivum) induces vigorous plants. Functional & Integrative Genomics 16 (2):171-182. doi:10.1007/s10142-016-0475-2

Song MJ, Potter B, Doyle JJ, Coate JE (2020) Gene Balance Predicts Transcriptional Responses Immediately Following Ploidy Change in Arabidopsis thaliana. The Plant Cell. doi:10.1105/tpc.19.00832

Sørensen MB, Chaudhury AM, Robert H, Bancharel E, Berger F (2001) Polycomb group genes control pattern formation in plant seed. Current Biology 11 (4):277-281. doi:10.1016/S0960-9822(01)00072-0

Spillane C, MacDougall C, Stock C, Köhler C, Vielle-Calzada J-P, Nunes SM, Grossniklaus U, Goodrich J (2000) Interaction of the Arabidopsis Polycomb group proteins FIE and MEA mediates their common phenotypes. Current Biology 10 (23):1535-1538. doi:10.1016/S0960-9822(00)00839-3

Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo J-M, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U (2007) Positive darwinian selection at the imprinted MEDEA locus in plants. Nature 448 (7151):349-352. doi:10.1038/nature05984

Stein O, Avin-Wittenberg T, Krahnert I, Zemach H, Bogol V, Daron O, Aloni R, Fernie AR, Granot D (2017) Arabidopsis Fructokinases Are Important for Seed Oil Accumulation and Vascular Development. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.02047

Steitz JA, Berg C, Hendrick JP, La Branche-Chabot H, Metspalu A, Rinke J, Yario T (1988) A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. Journal of Cell Biology 106 (3):545-556. doi:10.1083/jcb.106.3.545

Stiller JW (2007) Plastid endosymbiosis, genome evolution and the origin of green plants. Trends in Plant Science 12 (9):391-396. doi:10.1016/j.tplants.2007.08.002

Stoute AI, Varenko V, King GJ, Scott RJ, Kurup S (2012) Parental genome imbalance in Brassica oleracea causes asymmetric triploid block. The Plant Journal 71 (3):503-516. doi:10.1111/j.1365-313X.2012.05015.x

Sun Y, Li JQ, Yan JY, Yuan JJ, Li GX, Wu YR, Xu JM, Huang RF, Harberd NP, Ding ZJ, Zheng SJ (2020) Ethylene promotes seed iron storage during Arabidopsis seed maturation via ERF95 transcription factor. Journal of Integrative Plant Biology 62 (8):1193-1212. doi:https://doi.org/10.1111/jipb.12986

Sundaresan V (2005) Control of seed size in plants. Proceedings of the National Academy of Sciences 102 (50):17887-17888. doi:10.1073/pnas.0509021102

Tang Z, Yang Z, Hu Z, Zhang D, Lu X, Jia B, Deng D, Xu C (2013) Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. Molecular Breeding 31 (1):1-14. doi:10.1007/s11032-012-9762-3

Tejos RI, Mercado AV, Meisel LA (2010) Analysis of chlorophyll fluorescence reveals stage specific patterns of chloroplast-containing cells during Arabidopsis embryogenesis. Biological Research 43 (1):99-111. doi:10.4067/S0716-97602010000100012

Tiwari S, Spielman M, Schulz R, Oakey RJ, Kelsey G, Salazar A, Zhang K, Pennell R, Scott RJ (2010) Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. BMC Plant Biology 10 (1):72. doi:10.1186/1471-2229-10-72

Toda E, Ohnishi Y, Okamoto T (2018) An imbalanced parental genome ratio affects the development of rice zygotes. Journal of Experimental Botany 69 (10):2609-2619. doi:10.1093/jxb/ery094

Tonosaki K, Sekine D, Ohnishi T, Ono A, Furuumi H, Kurata N, Kinoshita T (2018) Overcoming the species hybridization barrier by ploidy manipulation in the genus Oryza. The Plant Journal: For Cell and Molecular Biology 93 (3):534-544. doi:10.1111/tpj.13803

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. Nature Biotechnology 28 (5):511-515. doi:10.1038/nbt.1621

Trivers RL (1974) Parent-Offspring Conflict. Integrative and Comparative Biology 14 (1):249-264. doi:10.1093/icb/14.1.249

Trösch R, Jarvis P (2011) The Stromal Processing Peptidase of Chloroplasts is Essential in Arabidopsis, with Knockout Mutations Causing Embryo Arrest after the 16-Cell Stage. PLOS ONE 6 (8):e23039. doi:10.1371/journal.pone.0023039

Tuteja R, McKeown PC, Ryan P, Morgan CC, Donoghue MTA, Downing T, O'Connell MJ, Spillane C (2019) Paternally expressed imprinted genes under positive Darwinian selection in Arabidopsis thaliana. Molecular Biology and Evolution. doi:10.1093/molbev/msz063

Ueda M, Aichinger E, Gong W, Groot E, Verstraeten I, Vu LD, Smet ID, Higashiyama T, Umeda M, Laux T (2017) Transcriptional integration of paternal and maternal factors in the Arabidopsis zygote. Genes & Development 31 (6):617-627. doi:10.1101/gad.292409.116

Upadhyaya HD, Kumar S, Gowda CLL, Singh S (2006) Two major genes for seed size in chickpea (Cicer arietinum L.). Euphytica 147 (3):311-315. doi:10.1007/s10681-005-9013-3

Uwer U, Willmitzer L, Altmann T (1998) Inactivation of a Glycyl-tRNA Synthetase Leads to an Arrest in Plant Embryo Development. The Plant Cell 10 (8):1277-1294. doi:10.1105/tpc.10.8.1277

Van Aken O, Zhang B, Law S, Narsai R, Whelan J (2013) AtWRKY40 and AtWRKY63 Modulate the Expression of Stress-Responsive Nuclear Genes Encoding Mitochondrial and Chloroplast Proteins. Plant Physiology 162 (1):254-271. doi:10.1104/pp.113.215996

Van Sandt VST, Suslov D, Verbelen J-P, Vissenberg K (2007) Xyloglucan Endotransglucosylase Activity Loosens a Plant Cell Wall. Annals of Botany 100 (7):1467-1473. doi:10.1093/aob/mcm248

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

Wade MJ, Goodnight CJ (2007) Cyto-Nuclear Epistasis: Two-Locus Random Genetic Drift in Hermaphroditic and Dioecious Species. Evolution 60 (4):643-659. doi:10.1111/j.0014-3820.2006.tb01146.x

Walbot V, Evans MMS (2003) Unique features of the plant life cycle and their consequences. Nature Reviews Genetics 4 (5):369-379. doi:10.1038/nrg1064

Walker EL (1998) Paramutation of the r1 Locus of Maize Is Associated With Increased Cytosine Methylation. Genetics 148 (4):1973-1981

Walton LJ, Kurepin LV, Yeung EC, Shah S, Emery RJN, Reid DM, Pharis RP (2012) Ethylene involvement in silique and seed development of canola, Brassica napus L. Plant Physiology and Biochemistry 58:142-150. doi:10.1016/j.plaphy.2012.06.016

Wang H, Wang J, Jiang J, Chen S, Guan Z, Liao Y, Chen F (2014) Reference genes for normalizing transcription in diploid and tetraploid Arabidopsis. Scientific Reports 4 (1):6781. doi:10.1038/srep06781

Wang P, Duan W, Takabayashi A, Endo T, Shikanai T, Ye J-Y, Mi H (2006) Chloroplastic NAD(P)H Dehydrogenase in Tobacco Leaves Functions in Alleviation of Oxidative Damage Caused by Temperature Stress. Plant Physiology 141 (2):465-474. doi:10.1104/pp.105.070490

Wang Y, Jiang H, Wang G (2020) PHERES1 Controls Endosperm Gene Imprinting and Seed Development. Trends in Plant Science 25 (6):517-519. doi:10.1016/j.tplants.2020.03.004

Wang Z-P, Xing H-L, Dong L, Zhang H-Y, Han C-Y, Wang X-C, Chen Q-J (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous

mutants for multiple target genes in Arabidopsis in a single generation. Genome Biology 16:144. doi:10.1186/s13059-015-0715-0

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C-T, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM (2011) Parent-of-Origin Effects on Gene Expression and DNA Methylation in the Maize Endosperm. The Plant Cell 23 (12):4221-4233. doi:10.1105/tpc.111.092668

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLOS ONE 6 (2):e16765. doi:10.1371/journal.pone.0016765

Weng M-L, Ruhlman TA, Jansen RK (2016) Plastid–Nuclear Interaction and Accelerated Coevolution in Plastid Ribosomal Genes in Geraniaceae. Genome Biology and Evolution 8 (6):1824-1838. doi:10.1093/gbe/evw115

Willi Y (2013) The Battle of the Sexes over Seed Size: Support for Both Kinship Genomic Imprinting and Interlocus Contest Evolution. The American Naturalist 181 (6):787-798. doi:10.1086/670196

Williams AM, Friso G, van Wijk KJ, Sloan DB (2019) Extreme variation in rates of evolution in the plastid Clp protease complex. The Plant Journal 98 (2):243-259. doi:10.1111/tpj.14208

Wolf JB (2009) Cytonuclear Interactions Can Favor the Evolution of Genomic Imprinting. Evolution 63 (5):1364-1371. doi:10.1111/j.1558-5646.2009.00632.x

Wolf JB (2013) Evolution of genomic imprinting as a coordinator of coadapted gene expression. Proceedings of the National Academy of Sciences 110 (13):5085-5090. doi:10.1073/pnas.1205686110

Wolf JB, Brandvain Y (2014) Gene interactions in the evolution of genomic imprinting. Heredity 113 (2):129-137. doi:10.1038/hdy.2014.7

Wolf JB, Hager R (2006) A Maternal–Offspring Coadaptation Theory for the Evolution of Genomic Imprinting. PLoS Biology 4 (12):e380. doi:10.1371/journal.pbio.0040380

Wolff P, Jiang H, Wang G, Santos-González J, Köhler C (2015) Paternally expressed imprinted genes establish postzygotic hybridization barriers in Arabidopsis thaliana. eLife 4:e10074. doi:10.7554/eLife.10074

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C, Nordborg M, Rehmsmeier M, Köhler C (2011) High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLoS Genetics 7 (6). doi:10.1371/journal.pgen.1002126

Woo HR, Chang-Hyo G, Joon-Hyun P, Teyssendier dlSB, Jin-Hee K, Youn-Il P, Gil NH (2002) Extended leaf longevity in the ore4-1 mutant of Arabidopsis with a reduced expression of a plastid ribosomal protein gene. The Plant Journal 31 (3):331-340. doi:10.1046/j.1365-313X.2002.01355.x

Woodson JD, Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. Nature Reviews Genetics 9 (5):383-395. doi:10.1038/nrg2348

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442 (7098):86. doi:10.1038/nature04815

Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q, Yadegari R, Wang X, Ni Z (2013) Dynamic Expression of Imprinted Genes Associates with Maternally Controlled Nutrient Allocation during Maize Endosperm Development. The Plant Cell 25 (9):3212-3227. doi:10.1105/tpc.113.115592

Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, Wang X-C, Chen Q-J (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant Biology 14:327. doi:10.1186/s12870-014-0327-y

Xu W, Purugganan MM, Polisensky DH, Antosiewicz DM, Fry SC, Braam J (1995) Arabidopsis TCH4, regulated by hormones and the environment, encodes a xyloglucan endotransglycosylase. The Plant Cell 7 (10):1555-1567. doi:10.1105/tpc.7.10.1555

Yamaguchi K, Subramanian AR (2000) The Plastid Ribosomal Proteins IDENTIFICATION OF ALL THE PROTEINS IN THE 50 S SUBUNIT OF AN ORGANELLE RIBOSOME (CHLOROPLAST). Journal of Biological Chemistry 275 (37):28466-28482. doi:10.1074/jbc.M005012200

Yamaguchi S (2006) Gibberellin Biosynthesis in Arabidopsis. Phytochemistry Reviews 5 (1):39-47. doi:10.1007/s11101-005-4248-0

Yang J, Zhang M, Yu J (2008) Mitochondrial retrograde regulation tuning fork in nuclear genes expressions of higher plants. Journal of Genetics and Genomics 35 (2):65-71. doi:10.1016/S1673-8527(08)60010-7

Yang L, Li B, Zheng X-y, Li J, Yang M, Dong X, He G, An C, Deng XW (2015) Salicylic acid biosynthesis is enhanced and contributes to increased biotrophic pathogen resistance in Arabidopsis hybrids. Nature Communications 6 (1):7309. doi:10.1038/ncomms8309

Yang R, He L, Huang H, Zhu J-K, Lozano-Duran R, Zhang H (2020) RNA-directed DNA methylation has an important developmental function in Arabidopsis that is masked by the chromatin remodeler PICKLE. Journal of Integrative Plant Biology 00 (00):1-6. doi:10.1111/jipb.12979

Yin T, Pan G, Liu H, Wu J, Li Y, Zhao Z, Fu T, Zhou Y (2012) The chloroplast ribosomal protein L21 gene is essential for plastid development and embryogenesis in Arabidopsis. Planta 235 (5):907-921. doi:10.1007/s00425-011-1547-0

Yoon J, Cho L-H, Kim S-R, Tun W, Peng X, Pasriga R, Moon S, Hong W-J, Ji H, Jung K-H, Jeon J-S, An G (2021) CTP synthase is essential for early endosperm development by regulating nuclei spacing. Plant Biotechnology Journal n/a (n/a). doi:https://doi.org/10.1111/pbi.13644

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Maroof MAS (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proceedings of the National Academy of Sciences 94 (17):9226-9231. doi:10.1073/pnas.94.17.9226

Yuan J, Chen S, Jiao W, Wang L, Wang L, Ye W, Lu J, Hong D, You S, Cheng Z, Yang D-L, Chen ZJ (2017) Both maternally and paternally imprinted genes regulate seed development in rice. New Phytologist:n/a-n/a. doi:10.1111/nph.14510

Zhang B, Li C, Li Y, Yu H (2020a) Mobile TERMINAL FLOWER1 determines seed size in Arabidopsis. Nature Plants:1-12. doi:10.1038/s41477-020-0749-5

Zhang H, Bishop B, Ringenberg W, Muir WM, Ogas J (2012) The CHD3 Remodeler PICKLE Associates with Genes Enriched for Trimethylation of Histone H3 Lysine 27. Plant Physiology 159 (1):418-432. doi:10.1104/pp.112.194878

Zhang H, Rider SD, Henderson JT, Fountain M, Chuang K, Kandachar V, Simons A, Edenberg HJ, Romero-Severson J, Muir WM, Ogas J (2008) The CHD3 Remodeler PICKLE Promotes Trimethylation of Histone H3 Lysine 27. Journal of Biological Chemistry 283 (33):22637-22648. doi:10.1074/jbc.M802129200

Zhang X-F, Tong J-H, Bai A-N, Liu C-M, Xiao L-T, Xue H-W (2020b) Phytohormone dynamics in developing endosperm influence rice grain shape and quality. Journal of Integrative Plant Biology n/a (n/a). doi:10.1111/jipb.12927

Zhang Y, Yao W, Wang F, Su Y, Zhang D, Hu S, Zhang X (2020c) AGC protein kinase AGC1-4 mediates seed size in Arabidopsis. Plant Cell Reports 39 (6):825-837. doi:10.1007/s00299-020-02533-z

Zhao P, Zhou X, Shen K, Liu Z, Cheng T, Liu D, Cheng Y, Peng X, Sun M-x (2019) Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. Developmental Cell 49 (6):882-893.e885. doi:10.1016/j.devcel.2019.04.016

Zhu R-M, Chai S, Zhang Z-Z, Ma C-L, Zhang Y, Li S (2020) Arabidopsis Chloroplast protein for Growth and Fertility1 (CGF1) and CGF2 are essential for chloroplast development and female gametogenesis. BMC Plant Biology 20 (1):172. doi:10.1186/s12870-020-02393-5

Zhu X, Tang G, Granier F, Bouchez D, Galili G (2001) A T-DNA Insertion Knockout of the Bifunctional Lysine-Ketoglutarate Reductase/Saccharopine Dehydrogenase Gene Elevates Lysine Levels in Arabidopsis Seeds. Plant Physiology 126 (4):1539-1545. doi:10.1104/pp.126.4.1539

Zimorski V, Ku C, Martin WF, Gould SB (2014) Endosymbiotic theory for organelle origins. Current Opinion in Microbiology 22:38-48. doi:10.1016/j.mib.2014.09.008

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, Wijk KJv (2008) Sorting Signals, N-Terminal Modifications and Abundance of the Chloroplast Proteome. PLOS ONE 3 (4):e1994. doi:10.1371/journal.pone.0001994

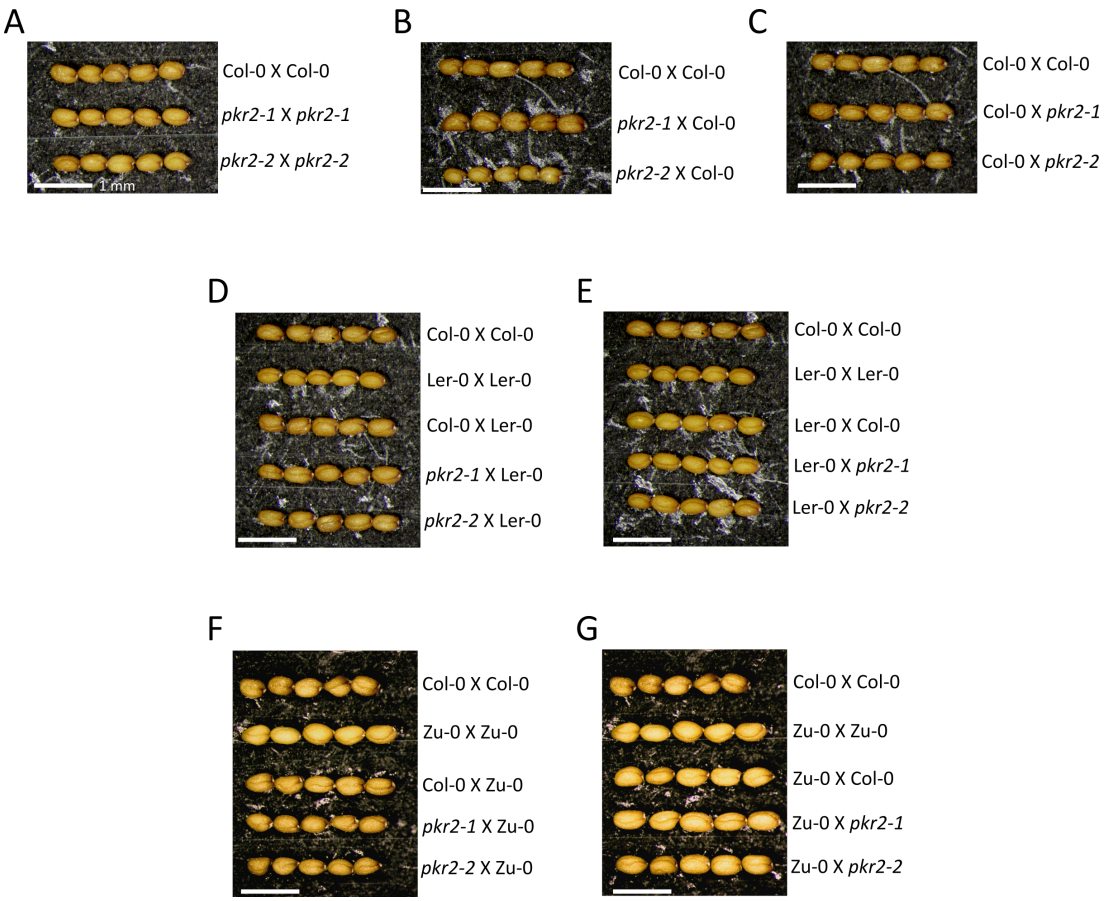
Supplementary material

Line	Identifier	Target gene	iMEG/ iPEG	PDS (lineage and/or site)	Primer	Sequence
pds9	SALK_022743	At1g70560	iPEG	Site	Forward	5'-TCTTTTCCATTACAACGTGGG -3'
					Reverse	5'-CGTCAAGACCATGACATCATG -3'
pds11	SAIL_221_C02	At2g40520	iPEG	Lineage and site	Forward	5'-TGTGTGCTTGAAGCTGAAATG-3'
					Reverse	5'-CTCTCCAGGCAGTGTAAGCAC-3'
pds14	SALK_201746	At5g28300	iPEG	Lineage and site	Forward	5'-GAAACTTTCTTCTGCCTTGGG-3'
					Reverse	5'-TTTCTTCATATGCATATGGACATG-3'
pds15	SALK_141443	At1g20910	iPEG	Lineage and site	Forward	5'-CTCTACAGGCACAGGGAACTG-3'
					Reverse	5'-TTCGTACTCCAAAAGTGCCTG-3'
pds17	SALK_115303	At4g31900	iPEG	Lineage	Forward	5'-TTCTGATTTTTCAACCGATGG-3'
					Reverse	5'-GGGGAGGAGTATCTGGTGAAG-3'
pds19	SALK_006274	At1g61330	iPEG	Lineage	Forward	5'-GACACATGGAAGCTATCTCGC-3'
					Reverse	5'-CTGTTCGAGTTTAGGGCACTC-3'
pds20	SALK_112939	At1g17770	iPEG	Lineage and site	Forward	5'-CAGTTTCTGTCTCTGCTTCGG-3'
					Reverse	5'-CCGCAAAATTCTCCATTTCTC-3'
pds21	SALK_152090C	At5g54650	iMEG	Site	Forward	5'-TTTTCGATCAGGGTTGTTGAG-3'
					Reverse	5'-AAGAGCTCCAGATACTTGGGG-3'
pds22	SALK_029765	At5g53150	iPEG	Lineage and site	Forward	5'-GCTAAAATGGAAGGGAAGGTG-3'
					Reverse	5'-TGGCTACTTTCTGGGTCAACC-3'
pds24	SALK_036992	At3g26590	iMEG	Site	Forward	5'-TCTTCGTTTTACACATGAAAAAGC-3'
					Reverse	5'-AATTTACCATGATGCCGAATG-3'
pds29	SALK_130191	At5g43780	iPEG	Site	Forward	5'-CTTCTGATCGTCGTCGATAGC-3'
					Reverse	5'-CACCGTGTCAATAAATCTCCG-3'
pds38	SALK_000930	At1g57820	iPEG	Lineage	Forward	5'-CTTAAAATCGTGGGCCTTTTC-3'
					Forward	5'-TTTCCATGAATGGAAAGCATC-3'

Supplementary Table S2.1: List of knockout *pds* lines and genotyping primers used in chapter 2.

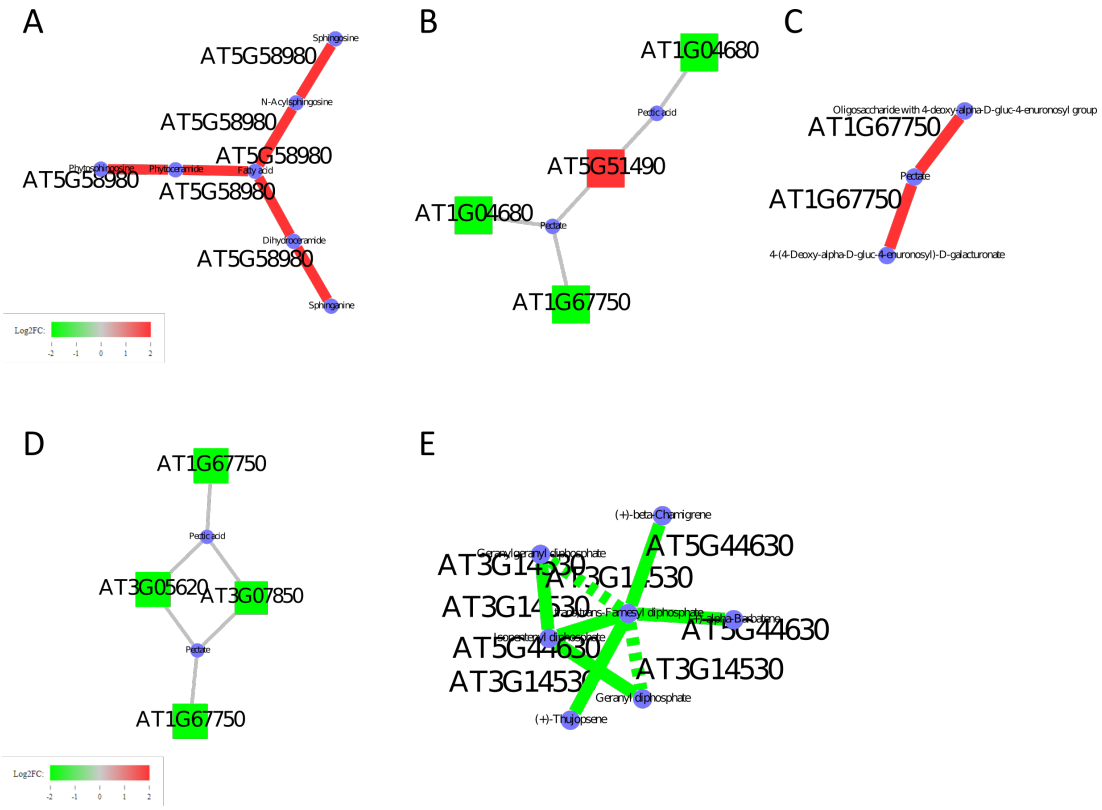
Gene	Sense	Sequence
PKR2	Forward	5'-CGGGCCAAAGCCAGATTATA-3'
	Reverse	5'-TTGGTGTCCCGTTGTCAGAA-3'
At1g04680	Forward	5’-CCAACACGCAGTGACGAATC-3’
	reverse	5’-AATTGGGGTCACATCGCCAA-3’
At1g07850	Forward	5’-GCATCGCCTCCAACATTCTTTT-3’
	Reverse	5’-GCTGATGTACCCTTAATGCCCT-3’
At5g51490	Forward	5’-TGTTGGTCGGAGATGGAATGAG-3’
	Reverse	5’-CTGGACCAGCTGTGTTCCTAAA-3’
At3g05620	Forward	5’-GGTGACCGGAATTTCATGCAG-3’
	Reverse	5’-AAGGCGGATTGGTCAGAATCA-3’
At3g14530	Forward	5’-CGGTCGAGATGATCCACACA-3’
	Reverse	5’-GTCATGTGCTCAAACGCCAA-3’
At5g44630	Forward	5’-CAATCTCCACCATCTTCCGAGT-3’
	Reverse	5’-TTGTCCCAAAGTGCACTGCTT-3’
At1g67750	Forward	5’-TTCTCCGACCTTCATTGCCTC-3’
	Reverse	5’-TTTCTCCCATTTCGGGTCGC-3’
At5g58980	Forward	5’-TCAATGCTCTTGTGGACGGT-3’
	Reverse	5’-CGTTCATGTGCCGGATTGTT-3’
EF-1α	Forward	5’-TCACCCTTGGTGTCAAGCAGAT-3’
	Reverse	5’-CAGGGTTGTATCCGACCTTCTT-3’

Supplementary table S2.2: Primers used for RT-qPCR experiments in chapter 2.



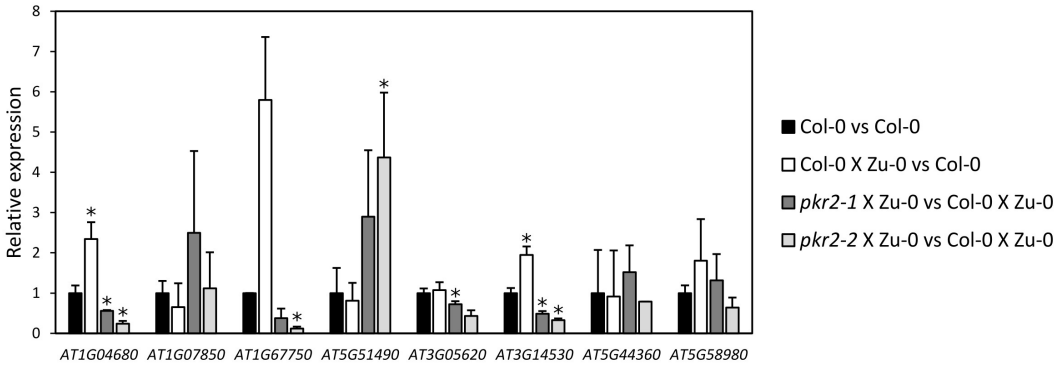
Supplementary figure S2.1: F1 Seed size heterosis

F1 seeds obtained from crosses between either Col-0, *pkr2-1* and *pk2-2* and reciprocally crosses with Ler-0 and Zu-0. **A**: Isogenic lines Col-0, pkr2-1 and pkr2-2. **B and C**: Maternal mutation (**B**) and paternal mutation (**C**) of *PKR2* in Col-0 background. **D and E**: Maternal mutation (**D**) and paternal mutation (**E**) of *PKR2* in reciprocal crosses between Col-0 and Ler-0. **F and G**: Maternal mutation (**F**) and paternal mutation (**G**) of *PKR2* in reciprocal crosses between Col-0 and Ler-0.



Supplementary Figure S2.2: Dysregulated pathways

Significantly dysregulated pathways between reciprocal Col-0, pkr2-2, Col-0 X Zu-0 and *pk2-2* X Zu-0. **A**: Alkalyne ceramidase pathway is upregulated in pkr2-2 X Zu-0 in comparison to Col-0, pkr2-2 and Col-0 X Zu-0. **B**: Regulation of cell wall plasticity is dysregulated in the direction of a tightening of cell walls in *pkr2-2* X Zu-0 compared to Col-0 X Zu-0. **C**: Regulation of cell wall plasticity is dysregulated in the direction of a loosening of cell walls in Col-0 X Zu-0 compared to Col-0. **D**: Downregulation of both pectin lyases and pectin methylesterase inhibitor in *pkr2-2* X Zu-0 in comparison to *pkr2-2*. **E**: Downregulation of terpenoid biosynthesis pathway in *pkr2-2* X Zu-0 in comparison to *pkr2-2*.



Supplementary Figure S2.3: RT-qPCR validation of RNAseq experiment

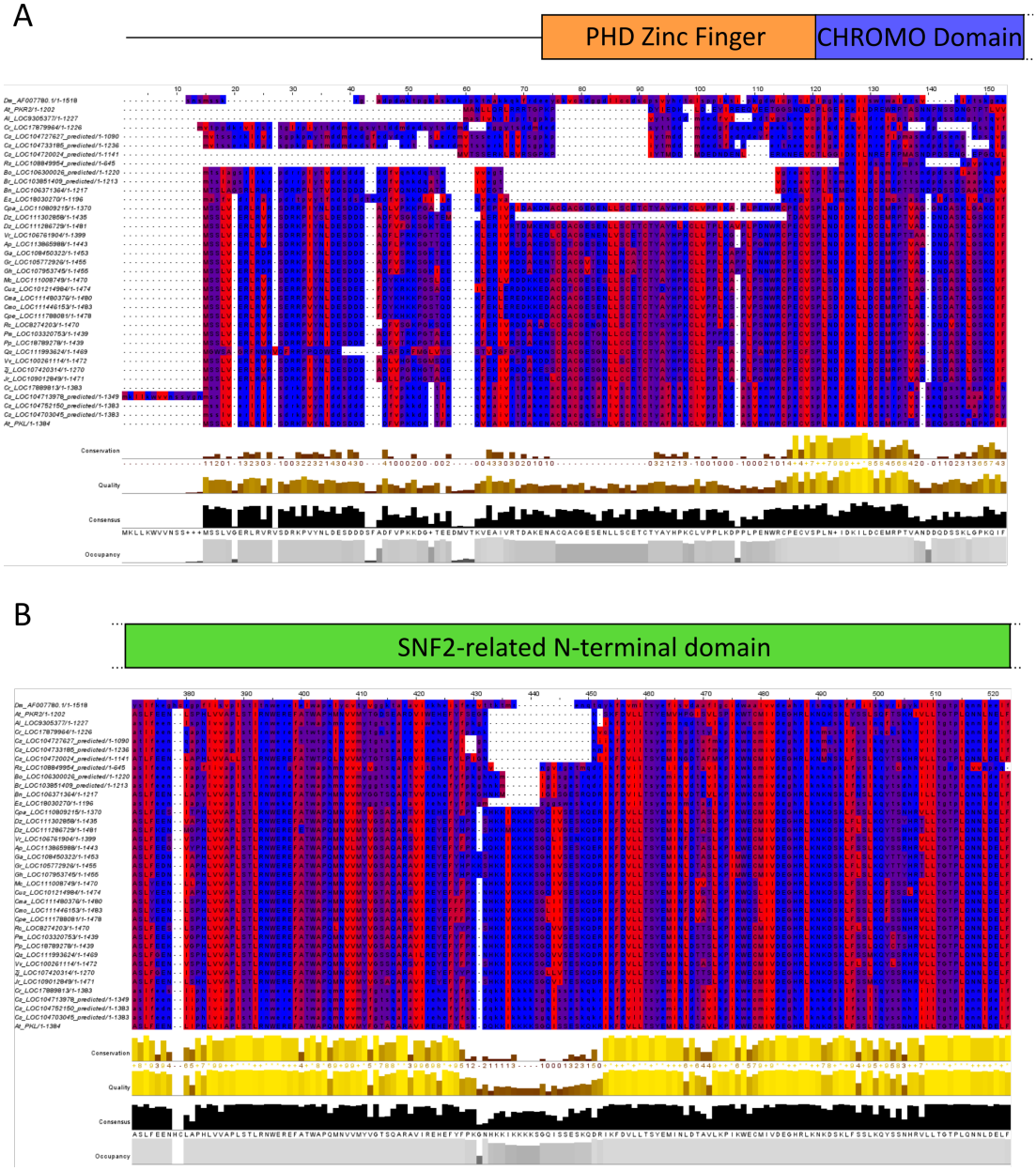
RT-qPCR verification of the relative expression of target genes of interest identified by RNAseq. Each gene expression is normalized with the expression of the housekeeping gene *EF-1α*. Stars indicate conditions for with gene expression fold change is statistically significant according to Student test (p-value < 0.05).

Line	Primer	Sequence
	Forward	5'-CGTAAAAGCTTCATTTGCGTC-3'
pkr2-1	Reverse (WT)	5'-TTGCTCACGACGTATGAGATG-3'
	Reverse (mutant)	5'-ATTTTGCCGATTTCGGAAC-3'
	Forward	5'-TTCTGATTTTTCAACCGATGG-3'
pkr2-2	Reverse (WT)	5'-GGGGAGGAGTATCTGGTGAAG-3'
	Reverse (mutant)	5'-ATTTTGCCGATTTCGGAAC-3'
	Forward	5'-AGCATTCTGTGCCCAACTGA-3'
pkl-1	Reverse (WT)	5'-CAATTCGCTCGTACTGCCATTAC-3'
	Reverse (mutant)	5'-CAATTCGCTCGTACTGCCATTAT-3'
	Forward	5'-TGCGGTTTGTTTCTCCTCTCG-3'
pkl-10	Reverse (WT)	5'-TGTTTCAGAGACCCAAAATCGTG-3'
	Reverse (mutant)	5'-ATATTGACCATCATACTCATTGC-3'

Supplementary table S3.1: Genotyping primers for *pkr2* and *pkl* mutants.

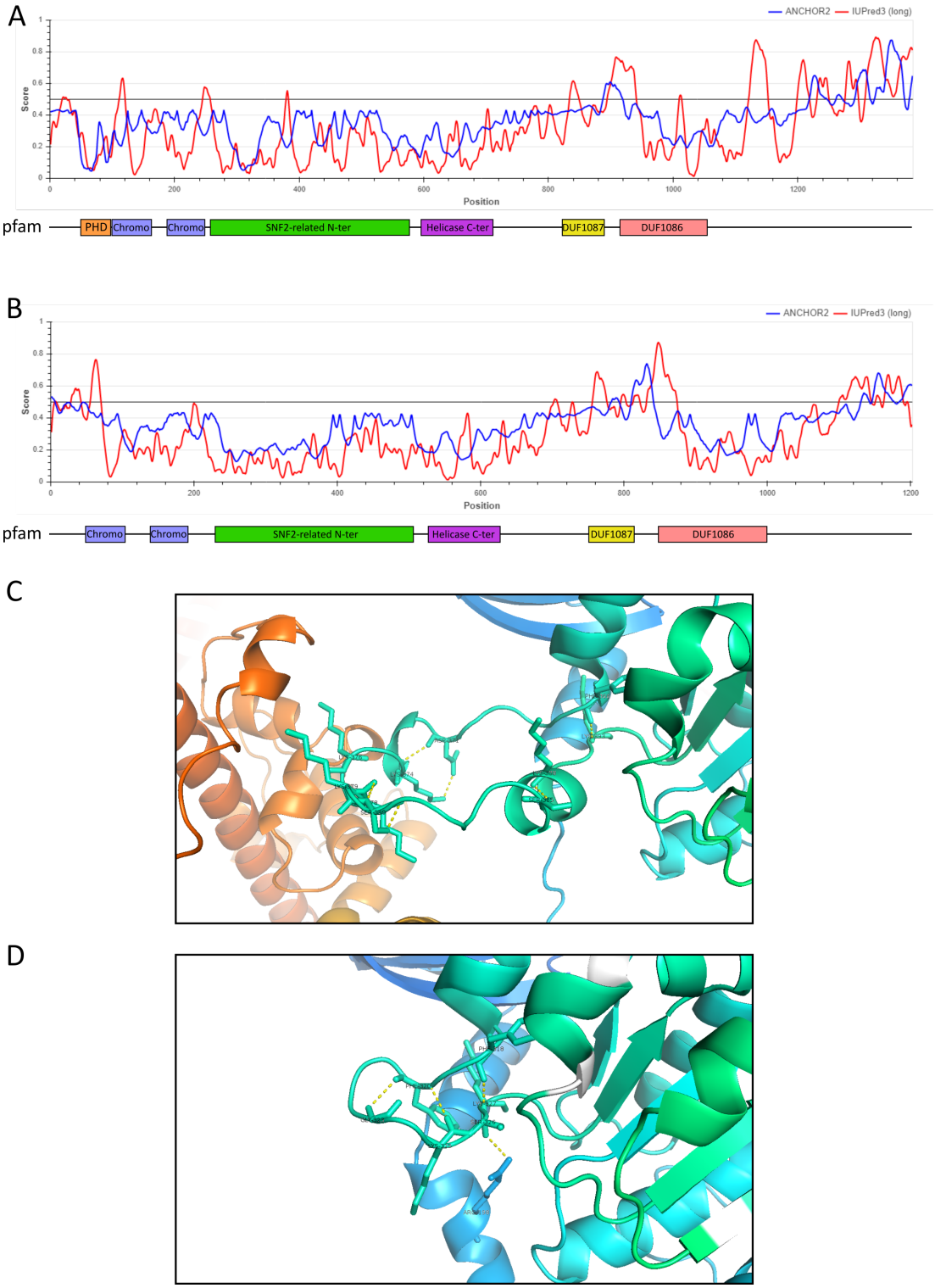
Gene	Sense	Sequence
At3g01270	Forward	5’-CACGAGCGTCATTACGACAAC-3’
	reverse	5’-CTTTTCCTTTTCCCTTGCCCC-3’
At3g09540	Forward	5’-GCTGGTGTCGAATCTCAGATACAT-3’
	Reverse	5’-CTAAGACACGATTTGGCTCCG-3’
At1g02810	Forward	5’-ACGGTGGTCACCGGAAATAG-3’
	Reverse	5’-AGAAGATTGAGAAATCTGCGCC-3’
At3g10720	Forward	5’-ACGCACTTAACCGTAACCTGA-3’
	Reverse	5’-CCCCTAAGTTCCGGCTAGTTTT-3’
LKR	Forward	5’-GGCAGAGCTGGTCTTGTTGA-3’
	Reverse	5’-TAATGGCAGTCCCTGGCTTG-3’
At2g34890	Forward	5’-TCGTGTGATCTCGAAAAGACCA-3’
	Reverse	5’-TTCTCGGGCATATTTTGCTGC-3’
At4g17470	Forward	5’-CAAAGGTCTGCTCTCCTTGTGA-3’
	Reverse	5’-GGAGTGGTTGATGAGGAACTGT-3’
At2g45220	Forward	5’-ATGGTGCAGCCAAACTCCAA-3’
	Reverse	5’-TAGTGTGAACGCGTGGGTTTT-3’
At1g67750	Forward	5’-TTCTCCGACCTTCATTGCCTC-3’
	Reverse	5’-TTTCTCCCATTTCGGGTCGC-3’
EF-1α	Forward	5’-TCACCCTTGGTGTCAAGCAGAT-3’
	Reverse	5’-CAGGGTTGTATCCGACCTTCTT-3’

Supplementary table S3.2: Primers used for RT-qPCR experiments in chapter 3.



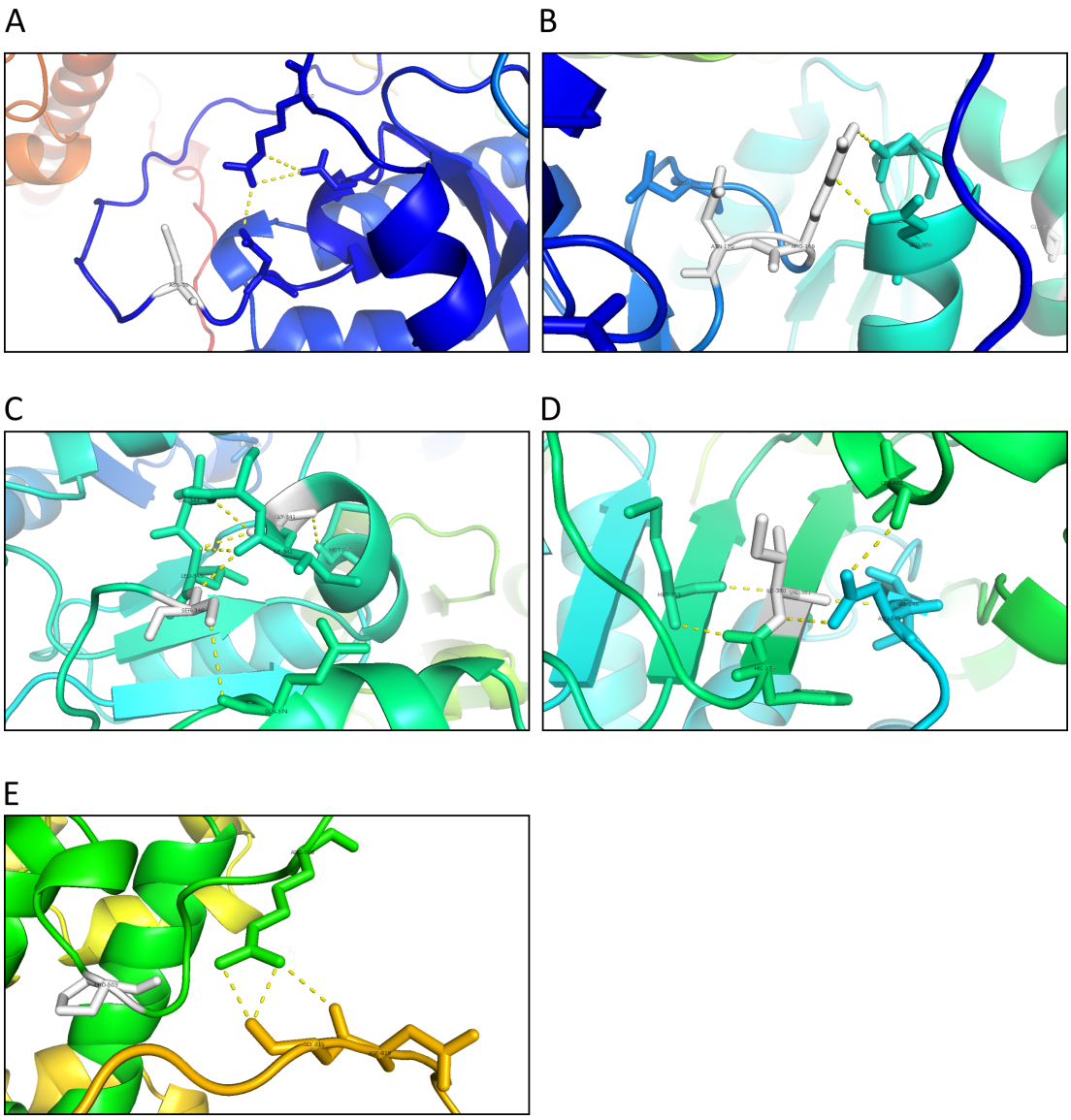
Supplementary figure S3.1: Alignement of PKL and PKR2 sequences

Alignment of PKL and PKR2 sequences from Embryophyta. **A**: PhD zinc finger from PKL is absent from PKR2 sequence. **B**: A portion of the DEAD-like helicase superfamily domain is missing from PKR2. Amino acids are coloured in accordance with their hydrophobicity. Blue: Hydrophilic amino acid; Red: Hydrophobic amino acid.



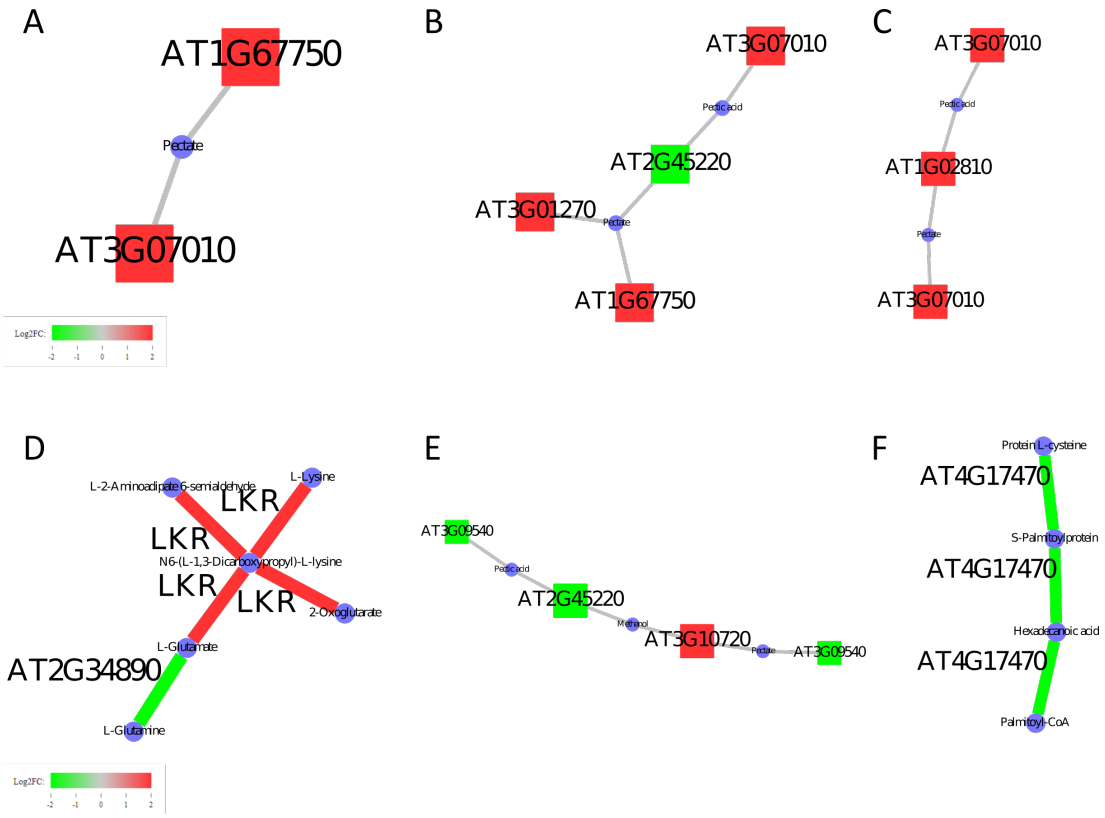
Supplementary figure S3.2: Protein intrinsic disorder

IUPRED3 and ANCHOR2 disorder prediction in PKL (**A**) and PKR2 (**B**) proteins. Loss of PHD domain in PKR2 correspond to the disappearance of a low disorder section, while the disappearance of a small loop section from the SNF2-related domain corresponds to the disappearance of a high disorder peak. The release of the C-terminal tail in PKR2, previously interacting with the SNF2-related domain small loop section in PKL seems to result in a smoothening of the disorder signal. **C and D**: Alphafold prediction of PKL (**C**) and PKR2 (**D**) NSF2-related N-terminal domain loop structure.



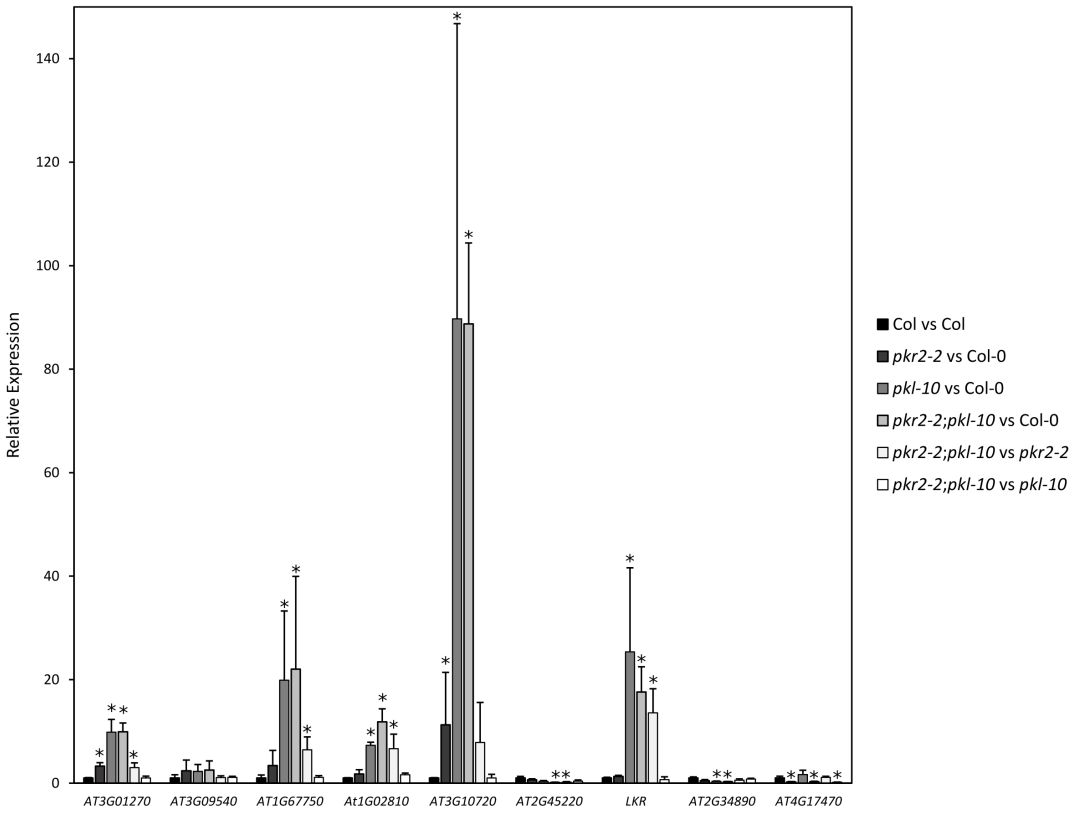
Supplementary figure S3.3: Amino acid positions under positive Darwinian selection in PKR2

Alphafold 3D prediction of sites under PDS in PKR2. **A**: N 39 is not involved in any 3D structure, but is located near an inter-loop hydrogen bond. **B**: R 169 is involved in a double hydrogen bond with D 306 and E 308 from the nearby alpha helix. N 170 does not seem to be involved in any non-covalent bond. **C**: G 341 is involved in hydrogen bonds with M 337, V 344 and L 345 which participating in an alpha helix structure. S 346 is at the start of the alpha helix and is involved in hydrogen bonds with I 342. It is also involved in a hydrogen bond with Q 314, maintaining cohesion between the two neighbouring alpha helices. **D**: I 380 is involved in a double hydrogen bond with N 245 and V 246 and participates in the cohesion of four neighbouring beta sheets. **E**: P 503 is not involved in any 3D structure, but is located near an inter-loop hydrogen triple hydrogen bond between R 506 and D 818 and G819.



Supplementary figure S3.4: Dysregulated pathways

Significantly dysregulated pathways between reciprocal Col-0, *pkr2-2*, *pkl-1* and *pkr2-2;pkl-10*. **A**: Increase of cell wall loosening activity in *pkr2-2* in comparison to Col-0. **B**: Increase of cell wall loosening activity in *pkl-1* in comparison to Col-0. **C**: Up-regulation of pectin lyases and pectin methylesterase inhibition activity in *pkr2-2;pkl-10* in comparison to Col-0. **D**: Decrease in CTP synthase activity and increase in Lysine-ketoglutarate reductase activity in *pkr2-2;pkl-10* in comparison to Col-0. **E**: Decrease of cell wall loosening activity in *pkr2-2;pkl-10* in comparison to *pkr2-2*. **F**: Reduction of thioesterase activity in *pkr2-2;pkl-10* in comparison to *pkl-1*.



Supplementary figure S3.5: RT-qPCR validation of RNAseq experiment

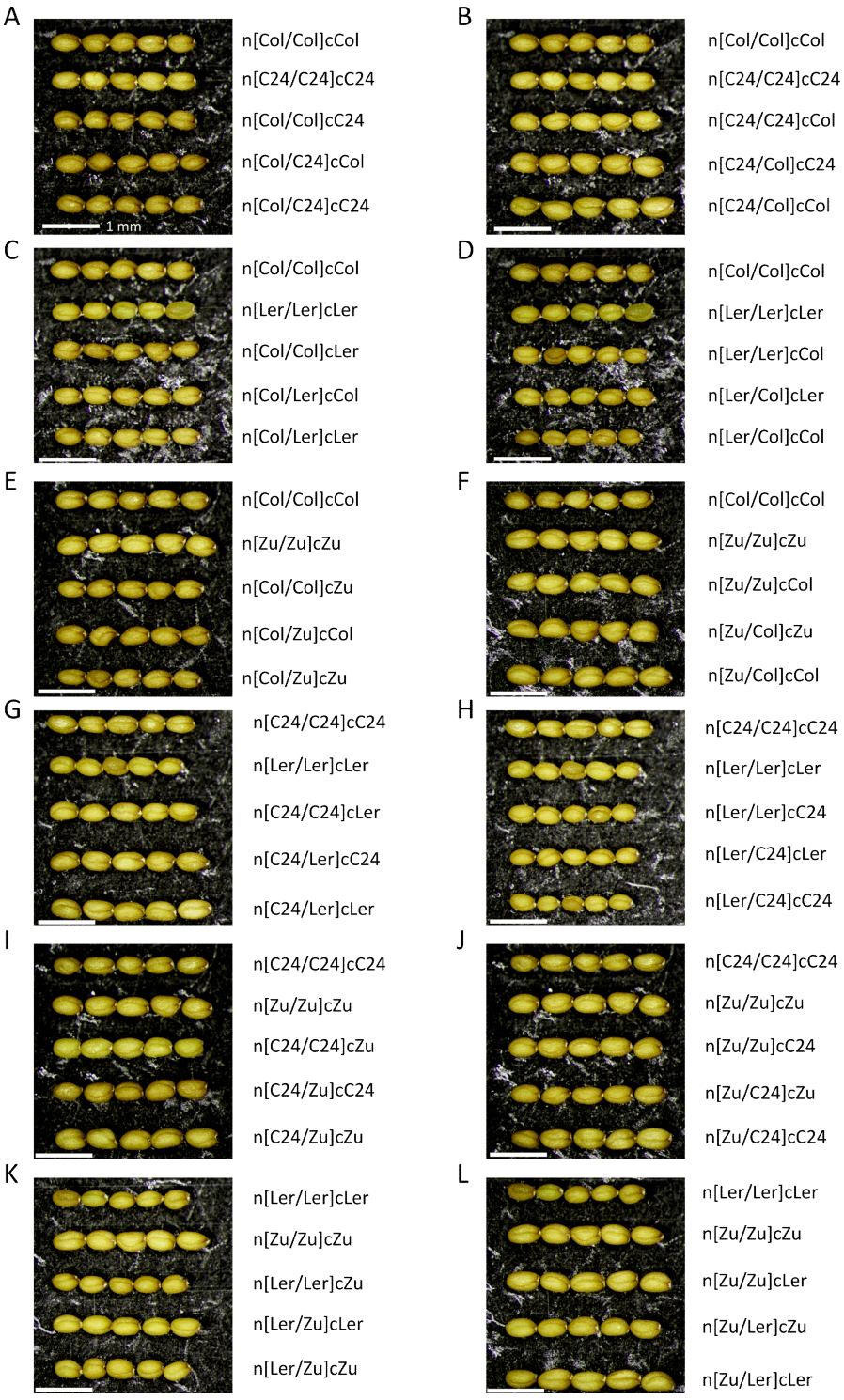
Relative expression of gene targets from the Col-0 group (**A**) and from the Zu-0 group (**B**). Relative gene expression of each gene is normalized with housekeeping gene *EF1α* expression and compared to a control sample using the Livak method. Stars indicate sample conditions for with gene expression fold change is statistically significant according to Student test (p-value < 0.05).

Name	Marker type	Sense	Foward primer	Reference
UPSC2-12717	Nuclear	Forward	5'-CAAAGGAAACACCTGCATCA-3'	Păcurar et al. 2012
		Reverse	5'-AAAATGGGGCCTAATACGTTG-3'	
L5P	Nuclear	Forward	5'-ATCCTCTCGAGGTAAGCGGT-3'	This chapter
		Reverse	5'-TCTTCTCAGGTCGGTGTGGA-3'	
CAPS Cyto I	Cytoplasmic	Forward	5'-GTCATTTACCCTGTTAGTCCG-3'	Ravi et al. 2014
		Reverse	5'-GAAATACAAGACAGCCAATCC-3'	
CAPS Cyto II	Cytoplasmic	Forward	5'-TGAACTAGCTGCTGCTTGTG-3'	Azhagiri et al. 2007
		Reverse	5'-GCTTTACTTAGCTCACCTCTG-3'	
MitoSeq	Cytoplasmic	Forward	5'-CCGCTCCGACTGCATACATA-3'	This chapter
		Reverse	5'-TCATTCTGTGCGCTAGTGGT-3'	

Supplementary table S4.1: List of primers used for the genotyping of *de novo* generated Cybrid lines.

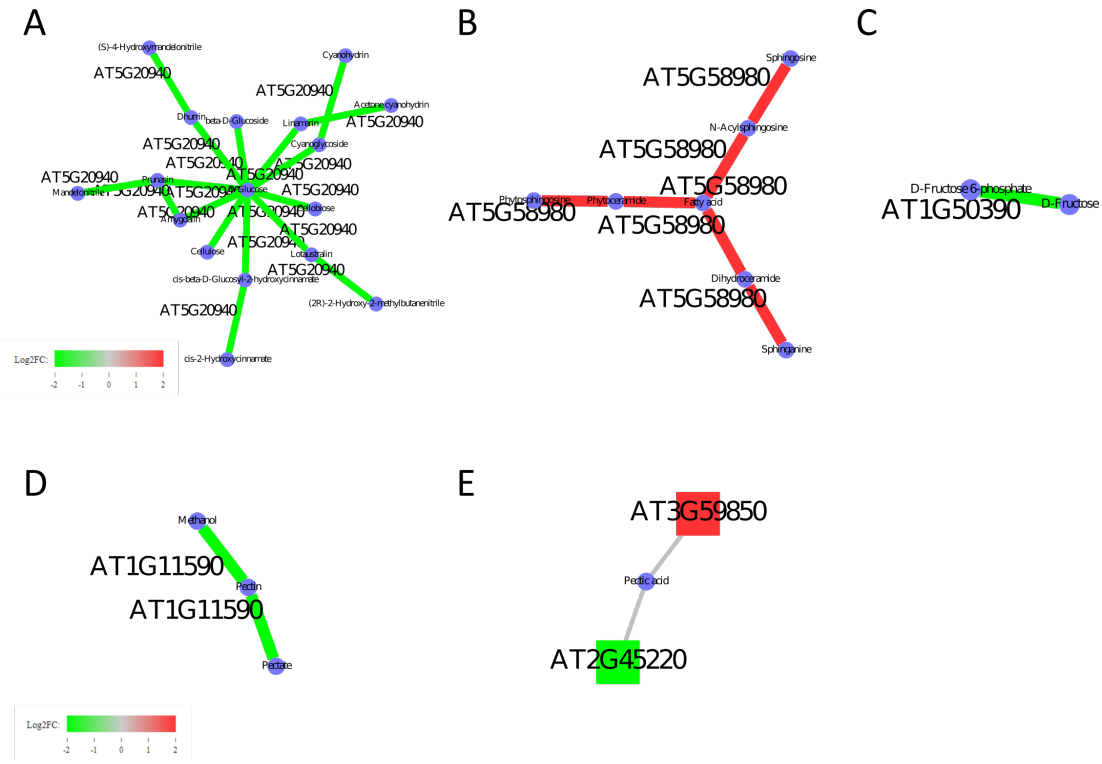
Gene	Sense	Sequence
At5g50390	Forward	5’-ACCTCTTTGGCCATCACCTG-3’
	reverse	5’-TCCACAGCGTCAACATGGAA-3’
At5g58980	Forward	5’-TCAATGCTCTTGTGGACGGT -3’
	Reverse	5’- CGTTCATGTGCCGGATTGTT-3’
At5g20940	Forward	5’-CACTCGCAAACGCCAAGTAT-3’
	Reverse	5’-GCTAAACACACTCCCGACGA-3’
At1g11590	Forward	5’-AGGGCAAGAATAGCTTTGGCT-3’
	Reverse	5’-CTTCGCGACCACAACATCAG-3’
TCH4	Forward	5’-GGAACAGTCACAACACTTTACTTGA-3’
	Reverse	5’-GCTGTTGGATCAAACCAGAGT-3’
ACS6	Forward	5’-TGTCTAAAATCGCCTCCGGT-3’
	Reverse	5’-GCATCAAATCTCCACAAAGCTGA-3’
EF-1α	Forward	5’-TCACCCTTGGTGTCAAGCAGAT-3’
	Reverse	5’-CAGGGTTGTATCCGACCTTCTT-3’

Supplementary table S4.2: List of primers used for RT-qPCR in chapter 4.



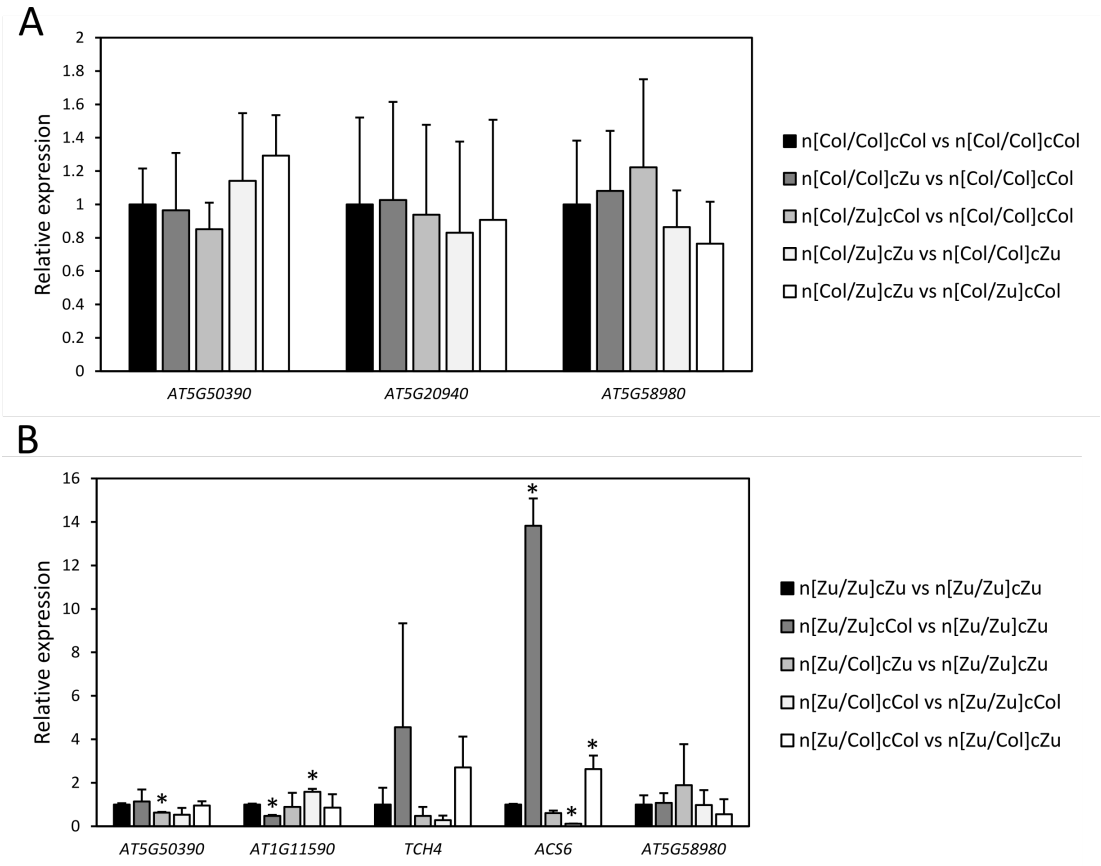
Supplementary Figure S4.1: Cybrids, hybrids and hybrid-cybrid seed size heterosis.

A and B: Col-0/C24 combination, separated as Col-0 subset **(A)** and C24 subset **(B)**. **C and D**: Col-0/Ler-0 combination, separated as Col-0 subset **(C)** and Ler-0 subset **(D)**. **E and F**: Col-0/Zu-0 combination, separated as Col-0 subset **(E)** and Zu-0 subset **(F)**. **G and H**: C24/Ler-0 combination, separated as C24 subset **(G)** and Ler-0 subset **(H)**. **I and J**: C24/Zu-0 combination, separated as C24 subset **(I)** and Zu-0 subset **(J)**. **K and L**: Ler-0/Zu-0 combination, separated as Ler-0 subset **(K)** and Zu-0 subset **(L)**.



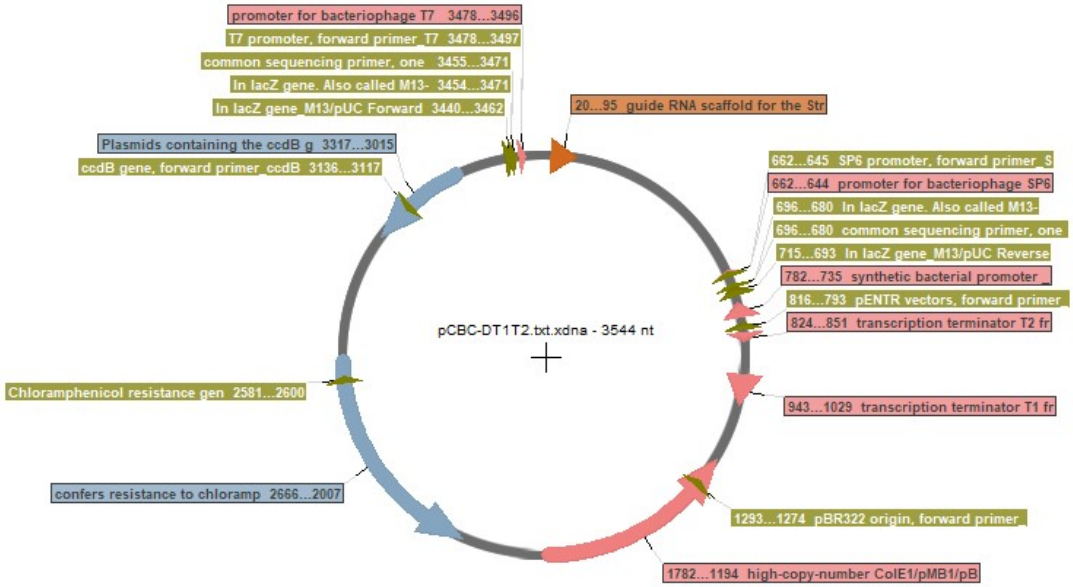
Supplementary Figure S4.2: Dysregulated pathways.

Significantly dysregulated pathways between reciprocal cybrid, hybrid and hybrid cybrid lines from Col-0 subset (**A, B and C**) and Zu-0 Subset (**D**) and between isogenic lines (**E**). **A**: Glucosyl hydrolase pathwath is significantly downregulated in cybrid n[Col/Col]cZu in comparison to WT n[Col/Col]cCol. **B**: Alkalyne ceramidase pathway is upregulated in hybrid n[Col/Zu]cCol and hybrid cybrid n[Col/Zu]cZu in comparison to their reciprocal isogenic lines as well as between isogenic n[Zu/Zu]cZu and n[Col/Col]cCol lines **C**: carbohydrate metabolism pathway is downregulated in hybrid cybrid n[Col/Zu]cZu in comparison to normal hybrid. **D**: Pectin methyl-esterase inhibition pathway is downregulated in hybrid n[Zu/Col]cZu and hybrid cybrid n[Zu/Col]cCol in comparison to their reciprocal isogenic lines. E: Cell wall loosening pathway is upregulated in isogenic n[Zu/Zu]cZu line in comparison to n[Col/Col]cCol line. Pathway images were generated using ShinyGam.

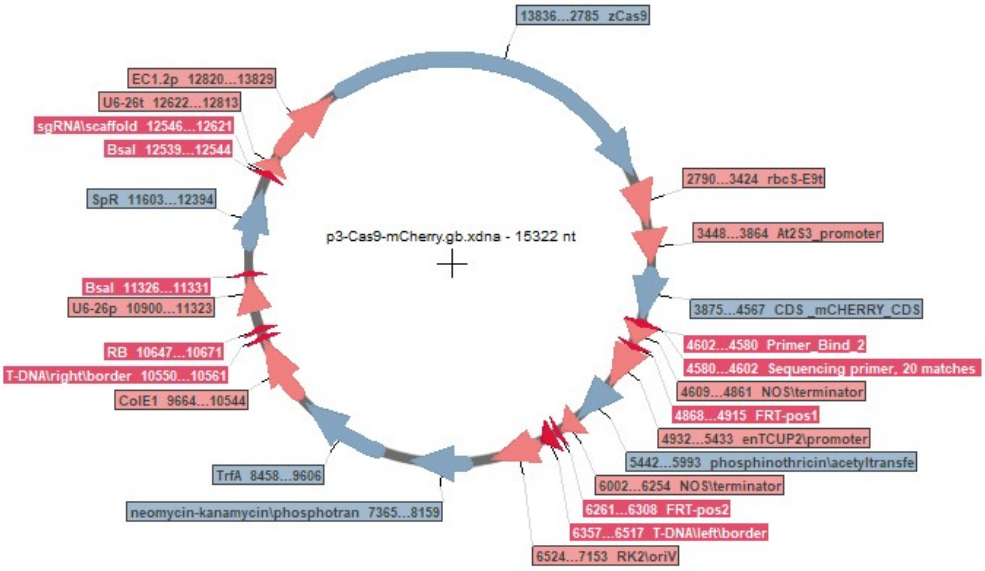


Supplementary Figure S4.3: RT-qPCR validation of RNAseq experiment

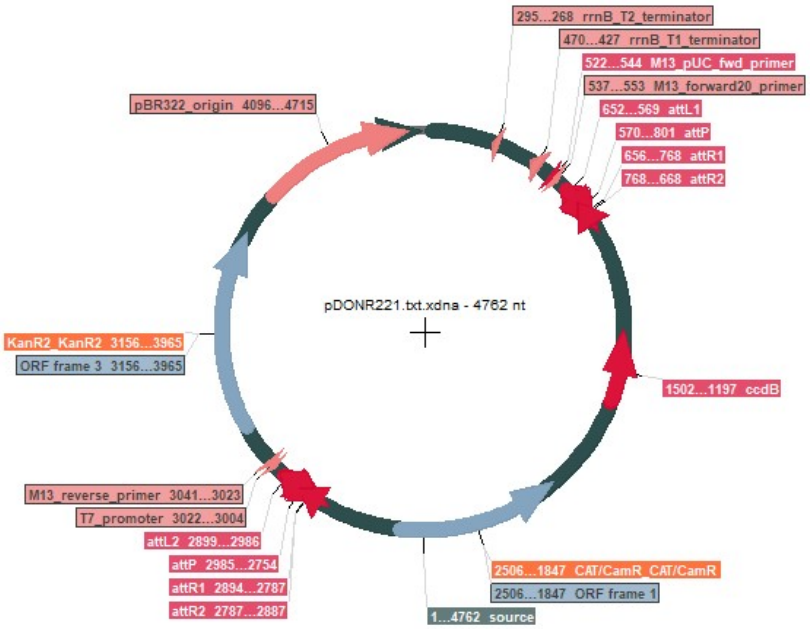
Relative expression of gene targets from the Col-0 group (**A**) and from the Zu-0 group (**B**). Relative gene expression of each gene is normalized with housekeeping gene *EF-1α* expression and compared to a control sample using the Livak method. Stars indicate sample conditions for with gene expression fold change is statistically significant according to Student test (p-value < 0.05).



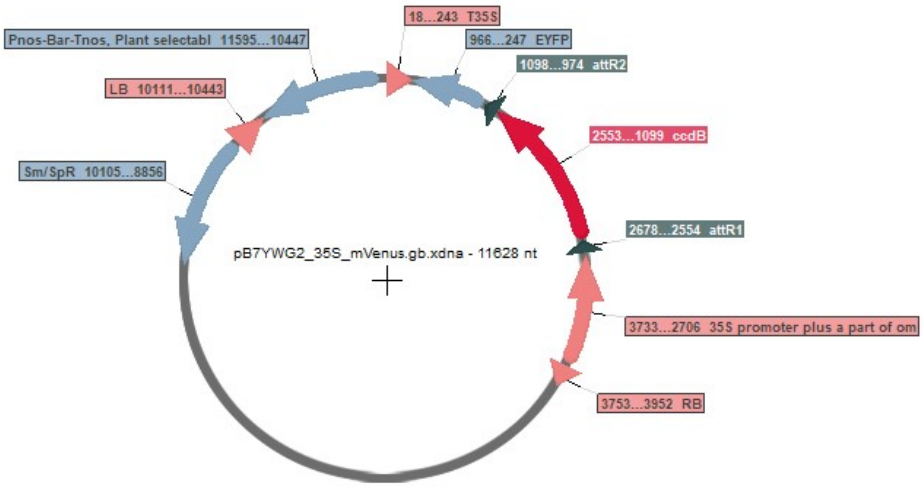
Supplementary figure S5.1: Map of the Cas9 cassette plasmid pCBC-DT1T2.



Supplementary figure S5.2: Map of the p3-mcherry used for Cas9-directed mutagenesis on Col-0 to generate the *prpl5-2* mutant line.

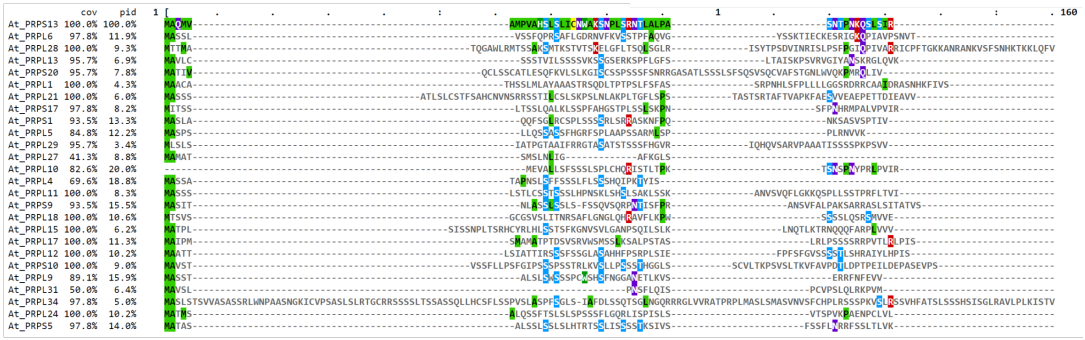


Supplementary figure S5.3: Map of Gateway pDONR221 vector.

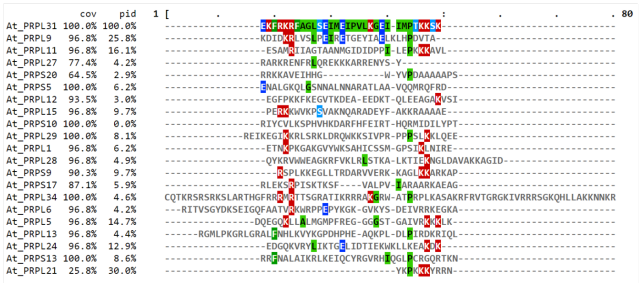


Supplementary figure S5.4: Map of pB7YWG2 used for as destination vector for subcellular localisation of PRPL5:EYFP.

A



B



Supplementary figure S5.5: Alignement of PRPs predicted signalling peptides.

Alignment of predicted cTP (**A**) and NLS (**B**) sequences in nuclear encoded PRPs.