Evolution of genomic imprinting in the *Capsella* genus

Marcelinus Rocky Hatorangan
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2017

Acta Universitatis agriculturae Sueciae
2017:85

Cover: Silhouette of the *Capsella* genus
(Image: MRH. Marcelinus Rocky Hatorangan)

ISSN 1652-6880
ISBN (print version) 978-91-7760-060-2
ISBN (electronic version) 978-91-7760-061-9

Print: SLU Repro, Uppsala 2017

Evolution of genomic imprinting in the *Capsella* genus

Abstract

According to the biological species concept, species are defined as groups of interbreeding populations reproductively isolated from other such groups. Hybrid seed lethality is a common reproductive barrier between plant species. Most frequently its developmental cause lies in the endosperm, but its molecular basis is not well understood. Hybrid seed defects differ depending on the cross direction, suggesting parent-of-origin-specific effects at work. Therefore, genomic imprinting, leading to the parent-of-origin-specific expression of genes in the endosperm, has been proposed to underlie hybrid seed lethality. Nevertheless, this hypothesis remains to be thoroughly tested.

Therefore, the main goal of this thesis was to explore the link between genomic imprinting and hybrid seed lethality in *Capsella*. The first part of this work aimed at characterizing the imprintome (the set of imprinted genes) of *Capsella rubella* and to compare it with the imprintome of the closely related species *Arabidopsis thaliana* (10-14 million years apart). This revealed that the imprintomes of both species are poorly conserved. Nevertheless, the pathways regulating genomic imprinting target transposable elements (TEs) in both species. Furthermore, studying the imprintomes of three *Capsella* species supported the notion of poor imprinting conservation between related species. This work also revealed that imprintome divergence between *Capsella* species is based on the divergence of TE insertions and consequent silencing mechanisms. Furthermore, this work discovered that hybrid seed lethality is widespread between each of the *Capsella* species. This phenomenon originates in the endosperm and exhibits a parent-of-origin pattern. Importantly, this work revealed that endosperm-based hybridization barriers in *Capsella* correlate with the number and expression of paternally-expressed imprinted genes (PEGs). In addition, the mating system strongly impacts on the number of PEGs, which suggests that transitions of mating systems fuel the establishment of postzygotic hybridization barriers.

Altogether, this thesis proposes a molecular and evolutionary explanation for the arising of endosperm-based hybridization barriers, in connection with genomic imprinting, TE dynamics and mating system. These data are expected to have a strong impact on plant breeding strategies and to promote further studies in this direction of research.

Keywords: hybridization barriers, genomic imprinting, transposable elements, endosperm, *Capsella*, *Arabidopsis*

Author's address: Marcelinus Rocky Hatorangan, SLU, Department of Plant Biology, P.O. Box 7080, SE-750 07 Uppsala, Sweden

Evolusi rekaman genomik pada genus *Capsella*

Abstrak

Menurut konsep spesiasi biologi, spesies didefinisikan sebagai kelompok populasi antarbiak yang saling terisolasi secara reproduktif. Letalitas benih hibrida adalah sawar reproduksi umum antara spesies tanaman. Seringkali penyebabnya terpaut dengan endosperma tapi dasar molekularnya belum dipahami secara utuh. Kecacatan berbeda benih hibrida, menunjukkan efek khas induk asal memainkan peranan penting dalam membentuk sawar hibridisasi pascazigotik. Oleh sebab itu, rekaman genomik yang mengarahkan ekspresi gen khas induk asal di endosperma diusulkan sebagai penyebab letalitas benih hibrida. Walaupun demikian, hipotesis ini belum sepenuhnya teruji.

Oleh sebab itu, tujuan utama disertasi ini adalah untuk menjelajahi hubungan antara rekaman genomik dan letalitas benih hibrida pada *Capsella*. Bagian pertama karya ini ditujukan pada karakterisasi imprintom (himpunan gen-gen terekam) dari *Capsella rubella* dan membandingkannya dengan imprintom spesies berkekerabatan dekat, *Arabidopsis thaliana* (10-14 juta tahun terpisah). Hasilnya menunjukkan imprintom kedua spesies tidak dilestarikan dengan baik. Meskipun demikian, lintasan yang mengatur rekaman genomik menyasar transposon pada kedua spesies. Selanjutnya, imprintom dari tiga spesies *Capsella*, dibangun dan dibandingkan. Hasilnya semakin mendukung buruknya pelestarian rekaman antar spesies berkekerabatan dekat. Perbandingan wilayah pengapit gen-gen terekam menunjukkan bahwa perbedaan imprintom antar spesies *Capsella* terletak pada divergensi penyisipan transposon dan kelangsungan mekanisme penyenyapan. Selanjutnya, karya ini menemukan adanya persebaran letalitas benih hibrida antar spesies *Capsella*. Fenomena ini konsisten muncul di dalam endosperma dan menampilkan pola induk asal. Puncaknya, karya ini menunjukkan bahwa sawar hibridisasi berbasis endosperma pada *Capsella* berkorelasi dengan jumlah dan ekspresi gen terekam sebapak (GTB). Akhirnya, sistem penyilangan sangat berpengaruh terhadap jumlah GTB dan transisi sistem penyilangan memacu pembentukan sawar hibridisasi pascazigotik.

Secara keseluruhan, disertasi ini mengusulkan penjelasan molekular dan evolusioner bagi kemunculan sawar hibridisasi berbasis endosperma dalam hubungannya dengan rekaman genomik, dinamika transposon, dan sistem penyilangan. Data ini diharapkan memiliki dampak yang kuat terhadap strategi pembiakan tanaman dan upaya mempromosikan studi lebih lanjut searah dengan tujuan riset ini.

Kata kunci: sawar hibridisasi, rekaman genomik, transposon, endosperma, *Capsella*, *Arabidopsis*

Alamat penulis: Marcelinus Rocky Hatorangan, SLU, Departemen Biologi Tanaman, Kotak Pos 7080, SE-750 07 Uppsala, Swedia

Preface

When I started my project on genomic imprinting in *Capsella*, the first thing that came into my mind is: “Will my study give important contributions to the scientific community and the people around the world?" I guess, this question has motivated not only me but also many scientists. I think my supervisor, Claudia, has done a really good work to keep that fire alive until I’m able to finish this thesis. I might finally have something that is worth it!

Anyway, back to the important content in this thesis, I want to note that I’m using the term shepherd’s purse as a generic name for *Capsella* genus. Before, this name has been used specifically for *C. bursa-pastoris*, where the last binomial name for this species actually means the same in Latin. The reason is quite simple. Untrained eyes won’t be able to see the differences between these species. In addition to that, at least two of the *Capsella* species, *C. bursa-pastoris* and *C. rubella* are known to be edible. I recommend you to try them! These plants hold a cultural importance especially for people in China and mentioned in some classic literatures. “As sweet as shepherd’s purse” is one of Chinese sayings to describe an “adequate” experience. For me it sounds similar to Swedish word, *lagom* (free to correct me if I’m wrong). This word describes my feeling a lot when I finished this thesis! *Capsella* might be not as important as other crops in human civilization but I think, it is enough to keep me to study for more than 4 years.

Now we have come to another quite important part. If you are not familiar with molecular biology (or whatever science nerdy stuff), my way to describe genomic imprinting or imprinted genes in this study is the way Mama Plants and Papa Plants communicate before deciding to start a family. In *Capsella*-ways, a too strong (or stubborn) plant needs another strong partner as an equal parent for having healthy children. Although it might not make any sense in some “Western” culture, I found this concept might work in my culture where actually two people need to be as stubborn as they can before they started a family. A stubborn person with a submissive partner won’t work because there’s no way to convey the feeling of one partner to the other. So in my philosophical way, you can say that this study is all about how plants communicate before they start a family, how to measure their stubbornness and how they can succeed as parents. If this concept intrigues you, please keep reading and enjoy!

Dedication

To my wife and my family for their hope, love, and understanding.

Sān chūn jì cài ráo yǒu wèi，jiǔ shú yīng táo zuì yǒu míng 。Qīng xīng bù gū zhū jiǔ bàn，líng rén wàng què yì xiāng qíng 。Gust the shepherd's purse at the third spring; best the ripe of cherry on the moon of ninth. Regret them not, comrades of wine; forget the lonely self in a foreign line.

Zheng Xie (1693–1765)

Contents

List of publications **11**

List of figures **13**

Abbreviations **15**

1 Introduction **17**
1.1 Seed formation and endosperm development 17
1.2 Molecular mechanisms underlying genomic imprinting 20
1.2.1 DNA Methylation 21
1.2.2 Histone Modification 23
1.2.3 Transposable Elements 24
1.2.4 Conservation of imprinted genes among angiosperms 24
1.3 Evolutionary forces driving genomic imprinting 26
1.3.1 The Kinship Theory 26
1.3.2 Mating system-dependent TE dynamics and genomic … 27
1.4 Interspecies hybridization barriers in the endosperm 27
1.5 Shepherd's Purse (*Capsella*) 30

2 Aims of the study **33**

3 Results and discussion **35**
3.1 Rapid evolution of genomic imprinting in two species of the … 35
3.2 Non-reciprocal interspecies hybridization barriers in the *Capsella* … 38
3.3 Paternally expressed genes correlate with hybridization success … 40

4 Conclusions **43**

5 Future perspectives **44**

References **46**

Popular science summary **57**

Populärvetenskaplig sammanfattning **59**

Rangkuman sains populer **61**

Acknowledgements **63**

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Hatorangan, M.R., Laenen, B., Steige, K., Slotte, T., Köhler, C*. (2016). Rapid evolution of genomic imprinting in two species of the Brassicaceae. *Plant Cell* tpc.00304.2016. doi:10.1105/tpc.16.00304

II Rebernig, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T., Köhler, C*. (2015). Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. *PLoS Genet* 11, e1005295. doi:10.1371/journal.pgen.1005295

III Lafon-Placette, C.¤, Hatorangan, M.R.¤, Steige, K.A., Cornille, A., Lascoux M., Slotte, T., Köhler, C*. Paternally expressed imprinted genes correlate with hybridization success and likely underpin the Endosperm Balance Number in the *Capsella* genus. (submitted)

* Corresponding author. ¤ Equal contribution

The contribution of Marcelinus Rocky Hatorangan to the papers included in this thesis was as follows:

I Designed, performed, analysed the data, and wrote the paper

II Performed and analysed the data

III Designed, performed, analysed the data, and wrote the paper

List of figures

Figure 1. Schematic double fertilization events in angiosperm with nuclear type endosperm. 18
Figure 2. Endosperm cellularization types. In cellular endosperm, each nuclear division is followed by cellularization … 19
Figure 3. Low conservation of imprinted genes among species. Venn diagram was generated by identifying *A. thaliana* … 25
Figure 4. The phylogenetic tree of *Capsella*. Distances are not scaled. Credit goes to Clément Lafon-Placette for the concept and idea. 30
Figure 5. Different epigenetic regulation of MEGs and PEGs in *C. rubella*. (A) Biallelically expressed genes … 37
Figure 6. Interspecies hybrid seeds in *Capsella* mimicking interploidy hybridization phenotypes in *Arabidopsis thaliana* (*At*) ... 39
Figure 7. Hypothetical changes in number of imprinted genes and EBN related to species transition time from outcrossing to selfing … 42

Abbreviations

DFE	Distribution of Fitness Effects
EBN	Endosperm Balance Number
kya	Thousand (kilo) years ago
MEG	Maternally Expressed Gene
Mya	Million (Mega) years ago
PEG	Paternally Expressed Gene
TE	Transposable Element

1 Introduction

1.1 Seed formation and endosperm development

The nutrition of both, humans and animals, including livestock, heavily depends on seeds and products derived thereof. This means that seed crops, and cereals in particular, are considered staple crops, whose productivity is of unparalleled relevance from economical, societal and ecological standpoints. It is interesting to note that grass seeds have been proposed to be a part of the hominin diet, in some parts of the world, for as long as 100,000 years (Mercader, 2009) and some evidence has been found supporting the claim that the domestication of cereals may have started as early as 23,000 years ago (Snir *et al.*, 2015). Since then, seeds of grass species have had a prominent role during the establishment of human civilization. Although the type of seeds consumed varied between different cultures and civilizations around the world, the domestication of cereals as a food source has been proposed to target the selection of plants with seeds containing a large endosperm, the tissue responsible for the storage of nutrients (Fuller, 2007; Preece *et al.*, 2017). Interestingly, at least in the case of wheat, increase in endosperm size was connected with a relative decrease in embryo size (Golan *et al.*, 2015).

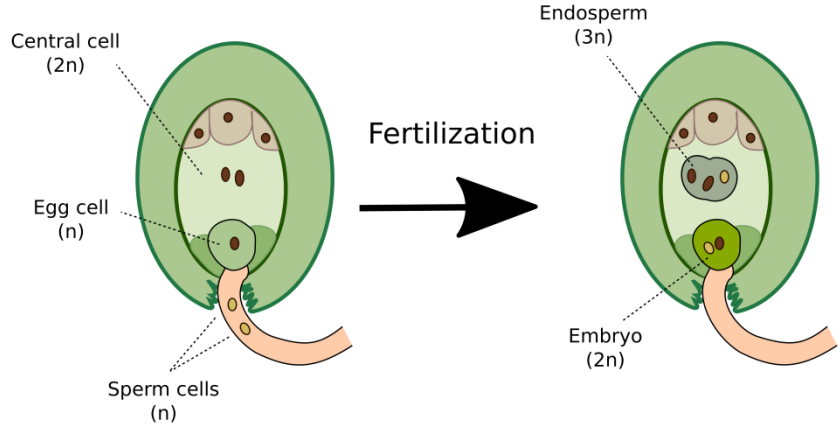

Figure 1. Schematic double fertilization events in angiosperm with nuclear type endosperm.

The development of a seed starts after the fertilization of the maternal gametes, egg cell and central cell, by the paternal gametes, the sperm cells, which are carried into the maternal ovule by the pollen tube (reviewed in Drews and Yadegari, 2002). Thus, seed development is preceded by a double fertilization event (Figure 1) leading to two fertilization products: the embryo and the endosperm. The embryo constitutes the next generation and its development is supported by the endosperm, which accumulates nutrients originating from the maternal tissues and conveys them to the developing embryo and/or germinating seedling (Lafon-Placette and Köhler, 2014). Surrounding the embryo and endosperm is the seed coat, which physically protects the fertilization products, and also participates in establishing seed dormancy, in facilitating water uptake during germination, and in seed dispersal, among other functions (Debeaujon *et al.*, 2000; Figueiredo and Köhler, 2014; Radchuk and Borisjuk, 2014). While the seed coat is a purely maternal tissue, the endosperm and embryo are directly derived from double fertilization. In order for the seed to develop successfully, these three structures have to communicate with one another and coordinate their growth and development (Figueiredo and Köhler, 2016; Ingram and Gutierrez-Marcos, 2015; Ingram, 2010; Lafon-Placette and Köhler, 2014). Embryo development, in particular, is highly dependent on the endosperm. Defects in endosperm formation significantly affect seed viability. Depending on the plant species, the development of the endosperm can be classified in three distinct categories: nuclear, cellular and helobial endosperms. Nuclear endosperm is the most prevalent type in angiosperms and is characterized by the absence of a cell plate between the mitotically-dividing daughter nuclei (Olsen, 2004; Sabelli

and Larkins, 2009). Thus, this type of endosperm initially develops as a syncytium, which later in development undergoes cellularization (Figure 2). The second type of endosperm, *ab initio* cellular, occurs in several plant species, including several members of the Solanaceae family (Briggs, 1993), and is characterized by the absence of a syncytium, meaning that cell walls are formed after each mitotic nuclear division. Finally, the helobial type of endosperm development, which is specific to several monocot species, is characterized by the formation of a cell wall following the first endosperm division, after which two chambers are formed; one that develops as a syncytium and one that undergoes cellular endosperm development (Swamy and Parameswaran, 1963).

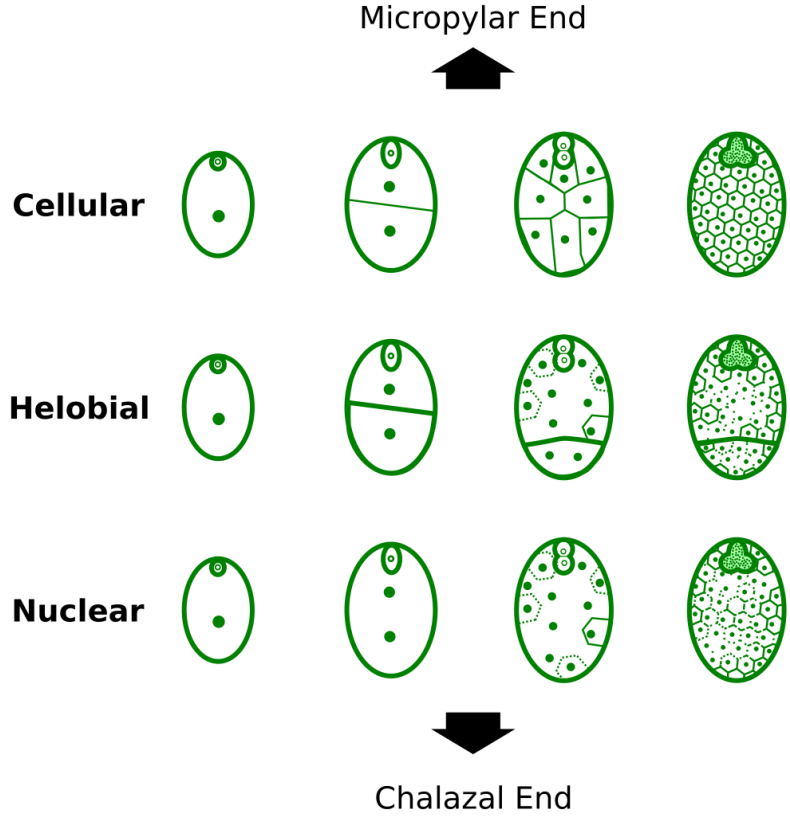

Figure 2. Endosperm cellularization types. In cellular endosperm, each nuclear division is followed by cellularization. In nuclear endosperm, the endosperm initially develops as syncytium and cellularization starts after a defined round of nuclear divisions. Helobial endosperm is characterized by the formation of two chambers. One follows the nuclear type of endosperm development and one the cellular type (Tobe and Kadokawa, 2010).

The nuclear type of endosperm development is the most common type of endosperm development and characteristic for cereal crops, as well as for the model plant species *Arabidopsis*, among many others. As previously mentioned, this type of endosperm is characterized by its initial development as a syncytium, which later cellularizes. Importantly, the timing of endosperm cellularization is crucial for the development of the embryo and for the viability of the seed (Hehenberger *et al.*, 2012; Pignocchi *et al.*, 2009). In instances where the endosperm fails to cellularize, or where cellularization is delayed, this can lead to embryo arrest and abortion of the seed (Hehenberger *et al.*, 2012; Pignocchi *et al.*, 2009; Scott *et al.*, 1998). Conversely, if the endosperm undergoes cellularization earlier than expected, this can lead to smaller seeds and, in extreme cases, may result in abortion of the progeny (Kang *et al.*, 2008; Scott *et al.*, 1998; Wang *et al.*, 2010).

It is interesting to note that the three developing structures within a seed are genetically distinct: in diploid species, the seed coat, given that it is not derived from a fertilization event, is diploid and contains only maternal genetic material; the embryo is also diploid, but contains one copy of both the maternal and paternal genomes; and the endosperm, which is derived from the fertilization of the homodiploid central cell by one of the paternal sperm cells, is triploid and contains two maternal and one paternal genome copies (Drews and Yadegari, 2002). Importantly, this ratio of maternal and paternal genomic copies is vital for the correct development of a seed. Deviations from this ratio, such as in crosses between parents of different ploidy levels leads to abnormal seeds which, in many cases, are not viable (Johnston *et al.*, 1980; Scott *et al.*, 1998; Josefsson *et al.*, 2006). These observations suggest that the parental genomes are not functionally equivalent during endosperm development and that there are likely parental-specific molecular factors whose activity needs to be tightly controlled in order to achieve successful seed development.

1.2 Molecular mechanisms underlying genomic imprinting

The non-equivalence of parental genomes suggests on the molecular level that certain genes are expressed in a parent-of-origin-specific manner (Arnaud and Feil, 2006; Jiang and Köhler, 2012). Genomic imprinting is an epigenetic phenomenon which causes certain genes to be specifically active only when either maternally or paternally inherited. Genes where the paternal alleles are silenced and the maternal alleles are predominantly active are called maternally expressed genes (MEGs). Conversely, genes where the paternal alleles are predominantly active are called paternally expressed genes (PEGs). In plants,

genomic imprinting is mostly confined to the endosperm; imprinting in the embryo is rare and transient (Hsieh *et al.*, 2011; Jahnke and Scholten, 2009; Nodine and Bartel, 2012; Pignatta *et al.*, 2014; Raissig *et al.*, 2013). This differs from the occurrence of imprinting in mammals, where imprinted genes are found in most tissue types (Pires and Grossniklaus, 2014).

There are different epigenetic mechanisms acting during male and female gametogenesis, causing both parental genomes to be epigenetically distinct (Gehring and Satyaki, 2017; Rodrigues and Zilberman, 2015). The epigenetic imprints applied during gametogenesis are mainly differences in histone modifications and DNA methylation. Most known imprinting mechanisms in plants depend on at least one of them if not both and will be discussed below.

1.2.1 DNA Methylation

DNA methylation, the addition of a methyl group to DNA, is a covalent modification that mainly affects cytosines in eukaryotes (Gehring and Henikoff, 2008). In plants, DNA methylation exists in three different sequence contexts, CG, CHG, and CHH methylation, where H can be any base but G (Law and Jacobsen, 2010).

Methylation in each context is catalysed by different enzymatic pathways and has different roles. In *Arabidopsis*, CG methylation is mostly maintained by the DNA methyltransferase MET1 and is present in high density in transposable elements (TEs) and gene bodies. While in TEs, this modification is associated with TE silencing (Law and Jacobsen, 2010), the role of CG methylation in gene regulation remains elusive (Bewick and Schmitz, 2017; Niederhuth and Schmitz, 2017). As for CG methylation, non-CG methylation (CHG and CHH) is also associated with TE silencing and heterochromatin formation. Non-CG methylation in CHG context is established and maintained by the CHROMOMETHYLTRANSFERASE 3 (CMT3) that is recruited by binding to dimethylated histone H3 on lysine 9 (H3K9me2) (Law and Jacobsen, 2010). CHH methylation is established by two different pathways; in the RNA-dependent DNA methylation (RdDM) pathway the DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) enzyme is recruited by ARGONAUTE4 (AGO4) to loci forming a chromatin-associated non-coding transcript by RNA Polymerase V (PolV). Small RNAs (sRNAs) bound by AGO4 that have sequence complementarity to PolV transcripts will effectively guide AGO4-DRM2 to target loci and establish CHH methylation (Law and Jacobsen, 2010; Matzke and Mosher, 2014; Stroud *et al.*, 2014, 2013). An alternative pathway leading to CHH methylation involves CMT2 and is facilitated by the chromatin remodeler DDM1 that possibly removes or shifts

the linker histone H1 to allow methyltransferases to access DNA (Zemach *et al.*, 2013).

Among all plant tissues, the epigenetic landscape in the endosperm is unique and determined before fertilization in the gametes. During female gametogenesis, the DNA glycosylase DEMETER (DME) is specifically expressed in the central cell, the precursor of the endosperm and removes methylated cytosines via the base-excision repair mechanism (Choi *et al.*, 2002; Park *et al.*, 2016). As a consequence, the genome of the central cell becomes globally demethylated in all sequence contexts, with small euchromatic TEs being preferentially targeted by DME (Hsieh *et al.*, 2009; Ibarra *et al.*, 2012; Park *et al.*, 2016). In the vegetative cell of pollen the activity of DME together with the silencing of *DDM1* leads to low DNA methylation, TE reactivation and sRNAs production (Calarco *et al.*, 2012; Schoft *et al.*, 2011). In particular, a class of 21/22-nt sRNAs, called epigenetically-activated small interfering RNAs (easiRNAs) is specifically enriched in pollen (Calarco *et al.*, 2012; Slotkin *et al.*, 2009). EasiRNAs can travel from the vegetative nucleus to the sperm cells, where they accumulate and potentially reinforce TE silencing (Martínez *et al.*, 2016). Indeed, 21/22-nt sRNAs have been shown to initiate DNA methylation via a modified RdDM pathway involving *RDR6* (Nuthikattu *et al.*, 2013). However, the level of CHH methylation remains low in sperm cells (Calarco *et al.*, 2012), pointing to a potential post-fertilization role of easiRNAs.

Thus, before fertilization, an epigenetic asymmetry exists between parental genomes: the maternal genome is demethylated while the paternal genome remains highly methylated at least in CG and CHG contexts. After fertilization and during the first stages of endosperm development, DNA methylation remains low and parentally asymmetric (Moreno-Romero *et al.*, 2016). This is consistent with members of the RdDM pathway being silenced during the nuclear phase of endosperm development (Belmonte *et al.*, 2013), thus unable to remethylate the maternal genome. This epigenetic asymmetry between parental genomes provides a potential molecular basis for parent-of-origin-specific gene expression. Consistently, genes that are normally maternally expressed (MEGs) become biallelically expressed (Hsieh *et al.*, 2011) upon met1 pollination, suggesting that the paternal allele of these genes is normally silenced due to DNA methylation while the maternal allele is demethylated and expressed. Nevertheless, this type of regulation does not affect all MEGs (Hsieh *et al.*, 2011), suggesting alternative pathways for MEG regulation (Gehring and Satyaki, 2017). However, it is also possible that many genes have been erroneously assigned as MEGs but are in fact seed coat expressed (Schon and Nodine, 2017), explaining their lack of response to *met1* pollination. For

PEGs, the use of *met1* as pollen donors leads to a reduction of paternal allele expression (Hsieh *et al.*, 2011), suggesting that DNA methylation in the paternal genome prevents the action of repressing factors, namely Polycomb group proteins.

1.2.2 Histone Modification

Polycomb group proteins are chromatin associated factors involved in the transcriptional regulation of target loci (Mozgova *et al.*, 2015). They act in complexes, called Polycomb Repressive Complex (PRC) 1 and PRC2. PRC2 catalyses the trimethylation of lysine 27 of histone H3 (H3K27me3), a repressive histone mark recognized by LIKE HETEROCHROMATIN PROTEIN1 (LHP1) that facilitates recruitment of PRC2, forming a feed-forward loop of H3K27me3 deposition (Derkacheva *et al.*, 2013), there are different PRC2 complexes acting during defined stages of plant development (Mozgova *et al.*, 2015). Two members of PRC2 are specifically expressed during plant reproduction: *FERTILIZATION-INDEPENDENT2* (*FIS2*) and *MEDEA* (*MEA*) (Luo *et al.*, 2000; Vielle-Calzada *et al.*, 1999), with *FIS2*, the name given to subunit of the FIS-PRC2, being specifically expressed in the endosperm. Both, *FIS2* and *MEA* are MEGs and activated in the central cell by the demethylation of their promoter via DME action (Jullien *et al.*, 2006). *MEA* and *FIS2* are essential for seed viability, as shown by the complete seed lethality in *mea* and *fis2* (Chaudhury *et al.*, 1997; Grossniklaus *et al.*, 1998). Interestingly, the endosperm defects of *fis2* and *mea* mutants can be rescued by either increasing the number of maternal genome copies, or by removing the paternal contribution from the endosperm (Kradolfer *et al.*, 2013a; Nowack *et al.*, 2007). This suggests that the FIS-PRC2 is involved in the regulation of the relative dosage of maternal and paternal genomes. Consistently, the silenced maternal alleles of most PEGs are marked by H3K27me3, while the active paternal allele is devoid of H3K27me3 (Moreno-Romero *et al.*, 2016). The cause for this parental asymmetry is partly understood. As mentioned above, due to the extensive DNA demethylation in the central cell (Ibarra *et al.*, 2012), the paternal genome has increased methylation levels compared to the maternal one. PRC2 is thought to be excluded from densely DNA methylated regions (Deleris *et al.*, 2012; Weinhofer *et al.*, 2010), providing an explanation for the specific presence of H3K27me3 on the maternal allele, leading to its silencing and establishing the imprinting pattern (PEG). The majority of PEGs in *Arabidopsis* exhibit this regulatory pattern (Moreno-Romero *et al.*, 2016), as do PEGs in maize and rice (Du *et al.*, 2014; Zhang *et al.*, 2014), suggesting a broad conservation of mechanisms regulating genomic imprinting in plants.

1.2.3 Transposable Elements

Transposable elements (TEs) are fragments of DNA that can move from one location in the genome to another (Fedoroff, 2012). TEs can be grouped into two classes: the first class, called retrotransposons, move via a "copy-and-paste" mechanism, while the second class, DNA transposons, move via a "cut-and-paste" fashion (Kim, 2017). TEs ability to multiply and insert in any part of the genome creates the risk of recombination errors, gene disruption, and genome instability (Chen, 2007; Fultz *et al.*, 2015). Eukaryotic organisms therefore evolved defense mechanisms that lead to the epigenetic silencing of TE activity. In plants, this involves the action of the RdDM and CMT2-based pathways, establishing DNA methylation in all sequence contexts (Matzke and Mosher, 2014; Stroud *et al.*, 2014; Zemach *et al.*, 2013). The TE content of a genome over generations is highly dynamic and depends on TE activity, host epigenetic silencing, TE elimination by mutations and recombination, and last but not least natural selection (Underwood *et al.*, 2017). Indeed, TEs can be an important source of genetic variation on which natural selection can act (Kim, 2017; Underwood *et al.*, 2017), especially when TEs affect gene expression. TEs can evolve as *cis* regulatory sequences recognized by the host transcriptional machinery (Chuong *et al.*, 2017). In addition, the epigenetic silencing of TEs can affect genes, either in *cis* as the chromatin repression of TEs will affect nearby genes, or in *trans* with TE-derived sRNAs targeting similar sequences that flank distant genes (Chuong *et al.*, 2017; Kim, 2017; Underwood *et al.*, 2017). In terms of genomic imprinting, TE insertions are thus a potential trigger for the establishment of the epigenetic asymmetry between parents. Consistently, there is a strong association between the presence of TEs in *cis* and the imprinting status of flanking genes in *Arabidopsis* (Pignatta *et al.*, 2014; Wolff *et al.*, 2011). More precisely, a specific family of TEs, helitrons, has been associated with PEGs, even though the reason for this association remains unclear (Pignatta *et al.*, 2014; Wolff *et al.*, 2011). In addition, sRNAs targeting MEGs accumulate to high levels in sperm cells and the seed, suggesting that TEs play a yet unclear role in MEG regulation via sRNAs, either in *cis* or in *trans* (Calarco *et al.*, 2012).

1.2.4 Conservation of imprinted genes among angiosperms

With the rise of next generation sequencing techniques, the imprintomes (the total of imprinted genes) of several plant species have been elucidated (Klosinska *et al.*, 2016; Luo *et al.*, 2011; Pignatta *et al.*, 2014; Waters *et al.*, 2013; Xu *et al.*, 2014; Zhang *et al.*, 2016). Their comparison revealed that the number of commonly imprinted genes is rather low (Figure 3). This weak

conservation seems independent of the genetic distance between species; thus, there is low conservation between monocots and dicots that diverged more than 140 million years ago (Mya), but also between sister species like *Arabidopsis thaliana* and *A. lyrata* that diverged about 5 Mya (Klosinska *et al.*, 2016; Luo *et al.*, 2011; Waters *et al.*, 2013). In addition, even high intraspecific variation of imprinted genes has been observed in *A. thaliana* (Pignatta *et al.*, 2014). Nevertheless, while there are species-specific imprinted genes, their biological function is largely conserved: they are enriched for transcriptional and chromatin regulators, suggesting that they play a role in the control of endosperm development (Luo *et al.*, 2011; Wolff *et al.*, 2011; Zhang *et al.*, 2016). In addition, conserved PEGs include auxin biosynthesis genes (Luo *et al.*, 2011), an important phytohormone essential for the initiation of endosperm and seed coat development (Figueiredo *et al.*, 2016, 2015).

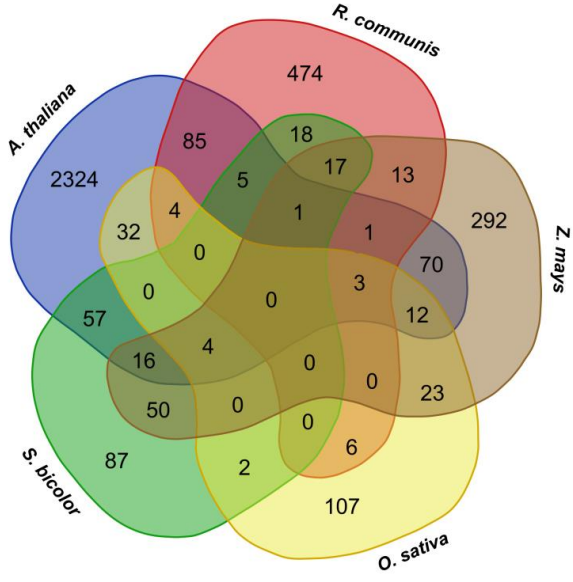

Figure 3. Low conservation of imprinted genes among species. Venn diagram was generated by identifying *A. thaliana* homologs of imprinted genes reported in each species using a simple BLAST search (e-value < 0.000001).

Accession-specific imprinting in *A. thaliana* has been associated with DNA methylation variation at TEs neighbouring imprinted genes (Pignatta *et al.*, 2014). This suggests that variation in TEs and their epigenetic silencing could be the molecular explanation for imprinting variation between species, a hypothesis that remains to be thoroughly tested.

1.3 Evolutionary forces driving genomic imprinting

While in the previous chapter 1.2 eluded on the molecular mechanisms causing genes to adopt an imprinted expression patters, in the following I will discuss the evolutionary forces that maintain imprinted genes in a population.

1.3.1 The Kinship Theory

The parental conflict or kinship theory developed by David Haig and Mark Westoby is undoubtly the most widely known theory explaining the evolution of genomic imprinting (Haig, 2000; Haig and Westoby, 1989). The theory posits that in species where the maternal parent mates with several males and invests resources in the offspring, a conflict arises over the allocation of resources to the developing progeny (Patten *et al.*, 2013). While it will be advantageous for the maternal parent to distribute resources equally to all of its offspring, paternal parents will favor if their offspring acquires more resources, on the expense of their non-related siblings. This scenario implies an intragenomic parental conflict among the offspring. It also implies that selection acts antagonistically on males and females concerning the evolution of traits regulating nutrient transfer to the progeny. While mono-allelic expression is generally a disadvantage, this conflict could be the reason that imprinted expression of certain genes is evolutionary conserved (Haig and Westoby, 1989). Consistent with the prediction of the kinship theory, paternally expressed genes in mammals favor prenatal growth, while maternally expressed genes act as inhibitors (Haig, 2004). Similarly, in *Arabidopsis*, the paternal genome has a positive influence on endosperm growth, while the maternal genome has the opposite effect (Scott *et al.*, 1998). In addition, parental conflict is expected to be stronger in outbreeding species compared to selfing species (Brandvain and Haig, 2005). Consistently, imprinted genes are expressed at a higher level in the outbreeding *A. lyrata* compared to its selfing congener *A. thaliana* (Klosinska *et al.*, 2016) and a high number of imprinted genes has been found in maize (Waters *et al.*, 2013).

If this theory is true, one would expect not to encounter imprinted genes in selfing plants. However, to date, the majority of studies about genomic imprinting in plants has been led on selfing species, and discovered a significant number of imprinted genes (Gehring *et al.*, 2011; Luo *et al.*, 2011; Wolff *et al.*, 2011; Xu *et al.*, 2014). One likely explanation is that selfing species retained genomic imprinting from the ancestral outbreeding mating type (Igic *et al.*, 2008). Alternatively, it is possible that imprinted genes acquired a function that is crucial for seed development, independently of any parental conflict. Consistent with assumption, some imprinted genes that are

conserved across selfers share similar functions (Figueiredo *et al.*, 2015; Luo *et al.*, 2011). To understand how mating system transitions impact on the fate of imprinted genes, it is required to survey genomic imprinting in additional outbreeding species and their related selfers.

1.3.2 Mating system-dependent TE dynamics and genomic imprinting

A major argument against the kinship theory is the extreme imprintome variability across angiosperms and the absence of phenotypic defects in mutants of imprinted genes in *Arabidopsis* (Luo *et al.*, 2011; Waters *et al.*, 2013; Wolff *et al.*, 2015, 2011; Xu *et al.*, 2014; Zhang *et al.*, 2016). Instead of being functionally relevant, genomic imprinting could be seen as a mere by-product of the dynamic epigenetic TE silencing occurring around fertilization (Barlow, 1993; Calarco *et al.*, 2012; Ibarra *et al.*, 2012). Nevertheless, both seemingly opposing scenarios can be reconciled when considering the silencing of TEs as the molecular cause for imprinted expression that is of functional relevance for only a few genes. The imprinted expression of those genes will be maintained over long evolutionary timescales, explaining conservation of some imprinted genes in a wide range of species (Klosinska *et al.*, 2016).

In this perspective, the evolution of genomic imprinting would follow the evolution of TEs. Hybridization, the rate of outcrossing, or the effective population size are proposed factors influencing the evolution of TE transposition rate in a given population (Ågren and Wright, 2011). Especially, it is expected that in selfing species, TEs get eliminated from the host genome over generations (Ågren and Wright, 2011; Boutin *et al.*, 2012). Thus, if one considers genomic imprinting solely as a by-product of TE silencing, less imprinted genes should be found in selfing species compared to outcrossing ones. This expectation meets the prediction made by the kinship theory (Brandvain and Haig, 2005; Haig and Westoby, 1989). To further explore this hypothesis, a thorough study comparing TE and genomic imprinting evolution in selfing and related outbreeding species is needed.

1.4 Interspecies hybridization barriers in the endosperm

Hybridization between related species has fuelled angiosperm evolution and is widely employed for breeding purposes. Many important crops for human civilization such as rapeseed (Arias *et al.*, 2014) and oranges (Wu *et al.*, 2014) are hybrids between two different species. Hybridization events allow gene flows that can lead to the acquisition of new adaptive traits (Goulet *et al.*,

2017). Nevertheless, there are multiple barriers that generally prevent hybridizations under natural conditions that can act either before (prezygotic) or after fertilization (postzygotic) (Seehausen *et al.*, 2014). This is reflected in the "biological species concept" that defines species as groups of interbreeding natural populations that are reproductively isolated from other populations (Mayr, 1996; Seehausen *et al.*, 2014).

In plants, postzygotic reproductive barriers can act directly after fertilization, impairing growth of the embryo and/or the endosperm. Alternatively, embryogenesis can be completed, but the resulting hybrid is not viable. Hybrid seed inviability is a major problem for plant breeding, and has thus been extensively studied during the last century (Cooper and Brink, 1945, 1942; Florez-Rueda *et al.*, 2016; Gill and Waines, 1978; Johnston and Hanneman, 1982; Lafon-Placette *et al.*, 2017; Lee and Cooper, 1958; Sansome *et al.*, 1942; von Wangenheim, 1957). Failure in endosperm development has been suspected as the main cause of hybrid seed lethality, since the hybrid embryo frequently survives and grows to maturity if dissected and provided with suitable nutrients (Chen *et al.*, 2016; Lafon-Placette *et al.*, 2017; Sharma *et al.*, 1996). Thus, it appears that negative epistasis between parental alleles in the hybrid as predicted by the Dobzhansky-Muller model of speciation (Dobzhansky and Dobzhansky, 1937; Muller, 1942, 1940) is a frequent phenomenon in the endosperm.

Interestingly, the endosperm defects are repeatedly observed after interspecific hybridization and very similar to problems of hybrid seeds originating from parents of the same species but of different ploidies (Florez-Rueda *et al.*, 2016; Ishikawa *et al.*, 2011; Lafon-Placette *et al.*, 2017; Lafon-Placette and Köhler, 2016; Lee and Cooper, 1958; Scott *et al.*, 1998; Sekine *et al.*, 2013; von Wangenheim, 1957). Namely, in species with nuclear endosperm, pollinations of maternal plants with pollen derived from a higher ploidy parent lead to increased seed growth, failure of endosperm cellularization, and prolonged endosperm proliferation (Scott *et al.*, 1998; Sekine *et al.*, 2013). The reciprocal hybridization leads to smaller seeds with precociously cellularized endosperm and reduced endosperm proliferation. The similarity of phenotypes in hybrid seeds between different species of the same ploidy suggests that rather than an intrinsic qualitative incompatibility, the cause of endosperm defects has a quantitative nature. Consistent with this idea, raising the ploidy level of one of the parents frequently increases the viability of hybrid seeds and restores endosperm development (Johnston and Hanneman, 1982; Josefsson *et al.*, 2006; Lafon-Placette *et al.*, 2017). These results suggest that species have different "effective" ploidies, which, upon hybridization lead to interploidy-like endosperm problems that can be restored

by manipulating their actual ploidy (Johnston and Hanneman, 1982). This has been conceptualized as Endosperm Balance Number (EBN, Johnston *et al.*, 1980). The EBN is a relative number given to a species according to its ability to form viable hybrid seeds with another species of defined EBN (Johnston and Hanneman, 1982). For example, a species does not form viable seeds when crossed to a species whose EBN is set to 2. If doubling the ploidy of the species with unknown EBN leads to viable seed formation, its EBN will be set to 1.

With the EBN concept, the problems affecting the endosperm from interploidy/interspecies hybridization come down to one same cause: a disturbed parental ratio of two maternal to one paternal (2m:1p) genome copies that is crucial for endosperm development (Johnston *et al.*, 1980). This can be due to an actual imbalance in parental copy numbers (interploidy hybrids), or an imbalance of "effective" ploidies while the actual genome copy numbers are still in a 2m:1p ratio (interspecies hybrids). Thus, identifying the genetic elements establishing hybridization barriers is likely to reveal the genetic basis of the EBN. Early on, it was proposed that the EBN could be under the control of few genetic loci (Ehlenfeldt and Hanneman, 1988; Johnston *et al.*, 1980). Nevertheless, recent QTL studies found multiple loci to be involved in hybrid seed lethality that have an additive effect (Burkart-Waco *et al.*, 2012; Garner *et al.*, 2016). This is consistent with the common idea that multiple small loci slowly diverge over time to establish postzygotic barriers, also named "snow-ball effect" (Coyne and Orr, 2004; Seehausen *et al.*, 2014). Identifying such loci is inherently difficult, explaining why no causal genetic element has yet been found to be responsible for interspecies hybrid seed lethality.

Nevertheless, during the last decades major progress has been made in identifying the molecular mechanisms underlying hybrid seed lethality. Especially, the parent-of-origin defects observed in interploidy and interspecies hybrid seeds suggest that imprinted genes play a causative role in establishing hybridization barriers in the endosperm (Gutierrez-Marcos *et al.*, 2003). Consistently, disturbed imprinting has been reported in hybrids of several species, ranging from the *Arabidopsis* to *Solanum* genera (Burkart-Waco *et al.*, 2015; Florez-Rueda *et al.*, 2016; Kirkbride *et al.*, 2015). The most compelling evidence comes from genetic studies showing that mutating PEGs can rescue seeds derived from interploidy hybridization that are otherwise inviable (Kradolfer *et al.*, 2013b; Wolff *et al.*, 2015). Interestingly, mutating these genes does not cause any phenotypic effect in diploid seeds (Wolff *et al.*, 2015), revealing that the functional role of imprinted genes is masked in a balanced genomic background. Nevertheless, whether the same genes underpin interspecies hybrid seed lethality remains to be tested.

What could be the reason that related species have different EBNs? If imprinted genes underpin the EBN, the same evolutionary forces driving genomic imprinting, namely TEs, are driving differences in EBN. Following the logic of the kinship theory predicting differences in imprinted genes between selfing and outbreeding species, the Weak Inbreeder/Strong Outbreeder (WISO) theory predicts outbreeding species to have a higher EBN compared to selfing species (Brandvain and Haig, 2005). This theory is supported by the strong correlation between EBN and mating systems in the *Solanum* genus (Brandvain and Haig, 2005; Lafon-Placette and Köhler, 2016). Nevertheless, this evidence remains correlative and more evidence is needed to conclude.

1.5 Shepherd's Purse (*Capsella*)

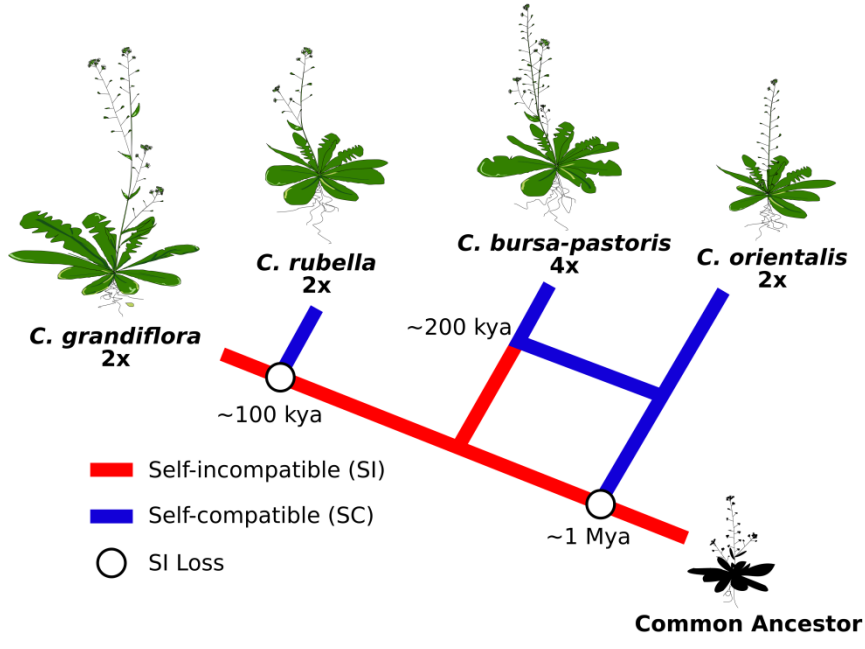

Figure 4. The phylogenetic tree of *Capsella*. Distances are not scaled. Credit goes to Clément Lafon-Placette for the concept and idea.

Capsella is a plant genus in the *Brassicaceae* family, close to the *Arabidopsis* genus (~10-14 Mya apart; Koch and Kiefer, 2005; Mitchell-Olds, 2001). It

consists of four species: the diploid *C. orientalis*, *C. rubella* and *C. grandiflora* and the allotetraploid *C. bursa-pastoris* (Figure 4).

C. grandiflora is the only outbreeding species of the genus due to its gametophytic self-incompatibility (SI) mating system (Paetsch *et al.*, 2006). In the *Brassicaceae*, it is assumed that gametophytic SI is the ancestral mating state (Igic *et al.*, 2008). Therefore, the self-compatible (SC) *Capsella* species (*C. orientalis*, *C. rubella*, and *C. bursa-pastoris*) most probably evolved from the ancestral lineage by the loss of SI (Douglas *et al.*, 2015). *C. rubella* arose around 100,000 years ago, as a likely consequence of the loss of SI followed by a strong bottleneck (Douglas *et al.*, 2015; Guo *et al.*, 2009; Slotte *et al.*, 2013). Meanwhile, *C. orientalis* is predicted to have diverged around 1 million years ago from the *C. rubella* / *C. grandiflora* lineage (Douglas *et al.*, 2015). While current evidence does not support a definitive conclusion, the high homozygosity observed in *C. orientalis* and the strong "selfing syndrome" (small flower size, high ovule/pollen ratio, absence of scent) argues for a long time of inbreeding, substantially exceeding the time of inbreeding of *C. rubella* (Douglas *et al.*, 2015; Hurka *et al.*, 2012). The emergence of *C. orientalis* could be concomitant with its loss of SI, but data are missing to conclude on this point. Finally, *C. bursa-pastoris* is an allotetraploid that originated from the hybridization between the *C. orientalis* and *C. rubella*/*grandiflora* lineages that occurred around 200,000 years ago (Douglas *et al.*, 2015).

TE evolution has been studied in all *Capsella* species (Ågren *et al.*, 2016, 2014). Among the diploid species, these studies revealed that *C. orientalis* has the lowest TE content among all species. In addition, *C. grandiflora* has the highest occurrence of gene-neighboring TEs, followed by *C. rubella* and *C. orientalis* (Ågren *et al.*, 2014). These findings are consistent with the theoretical prediction that in selfing species, TEs get eliminated over generations (Boutin *et al.*, 2012). Also, the lower TE incidence in *C. orientalis* compared to *C. rubella* argues in favor of a longer selfing time of the former species.

Regarding all the elements developed in the present introduction, several questions arise. Firstly, bearing in mind the general low conservation of genomic imprinting between species, how comparable is the imprintome of the closely related species *C. rubella* and *A. thaliana?* How distinctive are the imprintomes of all *Capsella* species? Are the imprintomes of *C. rubella*, *C. grandiflora* and *C. orientalis* associated with the TE composition of their genomes? Then, current predictions imply that outbreeders have a higher EBN compared to selfers. Are the selfer *C. rubella* and the outbreeder *C. grandiflora* distant enough to have evolved a different EBN? Does this trend stand with *C.*

orientalis? And finally, how is the imprintome of each species related to its EBN?

2 Aims of the study

The aims of this study are to:

1 Survey the imprintome of three related species, *Capsella rubella*, *C. grandiflora,* and *C. orientalis*;
2 Characterize the hybrid seed lethality occurring between the three species;
3 Establish a link between genomic imprinting and hybridization success in *Capsella*.

3 Results and discussion

3.1 Rapid evolution of genomic imprinting in two species of the Brassicaceae

The comparison of imprintomes across species is an important approach to better understand the mechanisms establishing genomic imprinting. Most studies reached the conclusion that genomic imprinting is poorly conserved among species (Luo *et al.*, 2011; Pignatta *et al.*, 2014; Waters *et al.*, 2013; Xu *et al.*, 2014; Zhang *et al.*, 2016). Nevertheless, those studies compared either species that have diverged long time ago (e.g. monocots vs dicots) or species that differ for traits that might affect genomic imprinting (e.g. outbreeding vs inbreeding). *A. thaliana* and *C. rubella* are closely related species that are predicted to have diverged around 10-14 million years ago (Koch and Kiefer, 2005; Mitchell-Olds, 2001). Both species are similar in terms of life cycle, mating system or genome size, making them well suitable to study conservation of genomic imprinting. The present work compared the imprintomes of *A. thaliana* and *C. rubella* and their epigenetic regulation.

Interestingly, the number of genes commonly imprinted in both, *C. rubella* (*Cr*) and *A. thaliana* (*At*), is rather low. The conservation of imprinting of PEGs is higher compared to MEGs, which is consistent to what has been observed in other species (Klosinska *et al.*, 2016; Luo *et al.*, 2011; Pignatta *et al.*, 2014; Waters *et al.*, 2013; Xu *et al.*, 2014; Zhang *et al.*, 2016). When comparing all possible homologs of *Cr* imprinted genes in *At*, around 29% of *Cr* PEGs were found to have imprinted homologs in *At*, meanwhile only 19% of MEGs in *Cr* were found to have their homologs imprinted in *At*. Genes that shared a common imprinting status in *Cr* and *At* were found to have similar functional roles. Based on GO enrichment analysis, conserved MEGs are enriched for genes involved in transcriptional regulation and metabolism, while

conserved PEGs are only enriched for genes involved in transcriptional regulation. This suggests that conserved imprinted genes might play an important role in regulating endosperm development. In contrast, non-conserved imprinted genes are not enriched for specific functional roles.

Given the low conservation of imprinted genes, their functional relevance could be questioned. To test this, signs of purifying selection and adaptive evolution of imprinted genes were investigated by measuring the distribution of fitness effects (DFE) and the rate of adaptive substitutions relative to the rate of neutral evolution (ωa). The DFE analysis revealed that PEGs have a higher proportion of nearly neutral nonsynonymous variants than average, suggesting relaxed purifying selection. Nevertheless, imprinted genes have a significantly higher ωa compared to a control gene set, suggesting imprinted genes undergo adaptive evolution. This argues for an important functional role of imprinted genes.

The molecular regulation of imprinted genes in *Cr* was then investigated and compared to *At*. In *At*, transposable elements (TEs), especially helitrons and MuDR transposons, are associated with imprinted genes (Pignatta *et al.*, 2014; Wolff *et al.*, 2011). Enrichment of TEs was found in the proximity of PEGs in *Cr*, confirming the trend observed in *At* (Pignatta *et al.*, 2014; Wolff *et al.*, 2011). In particular, helitrons and MuDR transposons were strongly associated with upstream and downstream regions of PEGs. No clear association between MEGs and TEs could be found. Consistent with the presence of TEs, non-CG methylation in PEGs was found to be significantly higher than average. Interestingly, while there was no association between MEGs and TEs, non-CG methylation was also significantly higher in MEGs. DNA methylation in MEGs is likely to be driven by small RNAs (sRNAs), as shown by the strong accumulation of sRNAs around these genes. This suggests a *trans* regulation of MEGs controlled by distant TEs. In addition, the level of non-CG methylation and sRNA accumulation was different between MEGs and PEGs, suggesting different epigenetic regulation between MEGs and PEGs in *Cr*, as suggested for *At* (Figure 5; Moreno-Romero *et al.*, 2016; Pignatta *et al.*, 2014; Wolff *et al.*, 2011).

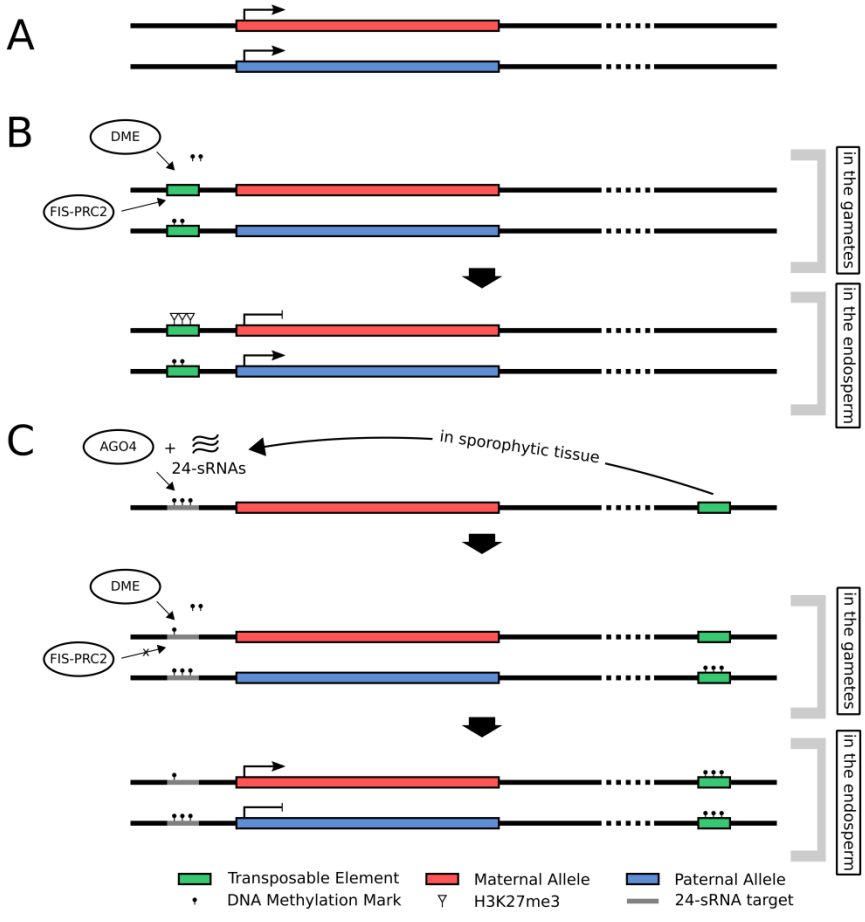

Figure 5. Different epigenetic regulation of MEGs and PEGs in *C. rubella*. (A) Biallelically expressed genes. (B) Demethylated TEs associated with the maternal allele of PEGs allow access of the FIS-PRC2, establishing H3K27me3 on the maternal allele. DNA methylation on the paternal allele prevents access of the FIS-PRC2, allowing paternal allele expression. (C) In sporophytic tissues, maternal and paternal alleles of MEGs will be targeted by *trans*-acting 24-nt small RNAs (sRNAs) generated by TEs distantly located from MEGs. This will cause repeat elements located in the vicinity of MEGs to be methylated via the RNA-dependent DNA methylation (RdDM) pathway. Activity of DEMETER (DME) in the central cell of the female gametophyte will remove DNA methylation on maternal alleles. Since DME is not active in sperm, paternal alleles will remain methylated. The RdDM pathway is not active in the early endosperm, maintaining DNA methylation differences on maternal and paternal alleles after fertilization.

3.2 Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm

The selfer *C. rubella* (*Cr*) and the outbreeder *C. grandiflora* (*Cg*) are very closely related (~100 kya; Douglas *et al.*, 2015). The stochastic accumulation of Dobzhansky-Muller incompatibilities is a function of time (Coyne and Orr, 2004; Seehausen *et al.*, 2014), we therefore wondered whether postzygotic hybridization barriers could evolve in this short timeframe. Since endosperm-based hybridization barriers have been proposed to quickly arise between outbreeding species and their inbreeder congeners (Brandvain and Haig, 2005), we specifically tested for the presence of endosperm-based postzygotic hybridization barriers between both species.

Following crosses between *Cr* and *Cg*, a significant rate of seed lethality was found. This rate was cross direction-dependent: while *Cr* × *Cg* seeds were shrivelled and mostly inviable, 60% of *Cg* × *Cr* seeds survived. In addition, seed size varied depending on the cross direction: *Cr* × *Cg* seeds grew larger than non-hybrid seeds, while *Cg* × *Cr* were substantially smaller.

The developmental cause for this hybrid seed inviability was investigated. Interestingly, the embryo could develop normally if dissected out of the seed and grown on MS medium, strongly suggesting that endosperm defects are the cause for hybrid seed lethality and not the embryo. For this reason, endosperm development was further investigated. Endosperm cellularization has been shown to be a crucial developmental transition for embryo survival (Hehenberger *et al.*, 2012; Pignocchi *et al.*, 2009). Consistently, *Cr* × *Cg* seeds showed delayed or even failed endosperm cellularization while in contrast, *Cg* × *Cr* endosperm cellularization occurred prematurely compared to the parental species. Such non-reciprocal endosperm defects have been described in hybrid seeds originating from parents of different ploidies (Scott *et al.*, 1998; Sekine *et al.*, 2013). This suggests that *Cr* × *Cg* non-reciprocal endosperm defects are a consequence of dosage imbalance between *Cr* and *Cg* genomes, even though *Cr* and *Cg* are both diploid. To further test whether interploidy and interspecies endosperm defects share the same mechanism, the transcriptome of reciprocal *Cr* × *Cg* hybrid seeds was compared to the transcriptome of *Arabidopsis* interploidy hybrid seeds (produced using the *osd1* mutant forming unreduced gametes). Interestingly, a significant number of genes that were up and downregulated in *Cr* × *Cg* compared to the parental species were misregulated the same way in 2x × 4x interploidy hybrid seeds. In particular, among upregulated genes we found *AGAMOUS-LIKE* (*AGL*) MADS box transcription factor genes that are putatively involved in inhibiting endosperm cellularization (Kang *et al.*, 2008). Conversely, a significant number of misregulated genes in *Cg* × *Cr* were also misregulated in 4x × 2x *Arabidopsis* seeds. Altogether,

these results show that the endosperm defects in *Cr* × *Cg* reciprocal hybrid seeds are similar to interploidy seed defects both at the developmental and molecular level. This suggests a dosage imbalance as the cause for *Cr* × *Cg* hybrid seed lethality, with *Cg* behaving as a species of higher ploidy level compared to *Cr*, even though both species are diploid (Figure 6).

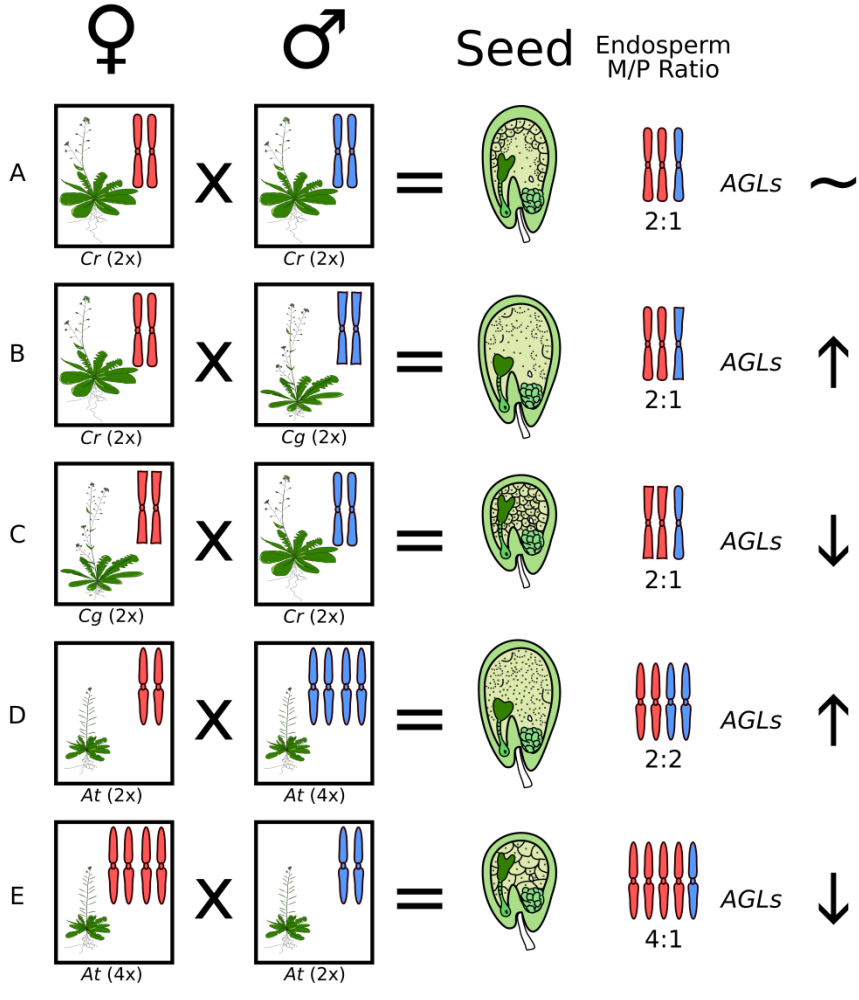

Figure 6. Interspecies hybrid seeds in *Capsella* mimicking interploidy hybridization phenotypes in *Arabidopsis thaliana* (*At*). In B and D, large seeds with delayed cellularization were observed when *Capsella grandiflora* (*Cg*, strong parent) or 4x *Arabidopsis thaliana* (*At*) served as pollen donor. While in C and E, small seeds with precocious cellularization were observed when *Capsella rubella* (*Cr*, weaker parent) or 2x *At* served as pollen donor. *AGAMOUS-LIKE* (AGL) genes are following similar expression patterns in seeds that share similar phenotype. No changes in endosperm proliferation rates were observed in B and C type of crosses compared to D and E.

Nevertheless, there are differences between seeds derived from interploidy and interspecies hybridization. While endosperm proliferation is altered in interploidy hybrid seeds (Scott *et al.*, 1998), aberrant nuclei proliferation was not observed in *Capsella* hybrid seeds. This suggests that endosperm cellularization and endosperm proliferation are two unconnected processes.

Finally, to determine the paternal genetic loci involved in *Cr* × *Cg* hybrid seed lethality, a QTL analysis was performed using *CgCr* Recombinant Inbred Lines (RILs; Sicard *et al.*, 2011). This work revealed that at least two or more loci in the *Cg* genome are responsible for the aborted hybrid seed phenotype. This is strengthened by subsequent finding of *Cg* PEGs (Paper III) that are located within the mentioned QTL, namely *Carubv10020920m.g* (homolog of translation factor, *AT1G66070*), *Carubv10021385m.g* (homolog of F-box protein, *AT1G66300*) and *Carubv10026304m.g* (homolog of ripening-responsive protein, *AT5G65380*).

3.3 Paternally expressed genes correlate with hybridization success and likely underpin the Endosperm Balance Number in the *Capsella* genus

Most of described cases of hybrid seed failure are related to a dosage imbalance between parental genomes in the endosperm (Ishikawa *et al.*, 2011; Johnston and Hanneman, 1982; Lafon-Placette *et al.*, 2017; Parrott and Smith, 1986; Valentine and Woodell, 1963). Even parents with the same ploidy level can have different genome dosages, or "Endosperm Balance Numbers" (EBNs; Johnston *et al.*, 1980; Johnston and Hanneman, 1982). The reasons for different EBNs between species, and thus hybrid seed lethality, are still unclear. The parent-of-origin defects frequently observed in interploidy and interspecies hybrid seeds suggest that parent-of-origin molecular mechanisms - namely genomic imprinting - are the processes underlying hybrid seed lethality (Gutierrez-Marcos *et al.*, 2003; Haig and Westoby, 1989; Lafon-Placette and Köhler, 2016). Consistently, a causal link between imprinted genes and interploidy seed lethality has been established (Kradolfer *et al.*, 2013b; Wolff *et al.*, 2015). Nevertheless, such evidence is missing for interspecific hybridization failures. In parallel, a strong correlation between mating system and EBN exists: selfing is usually associated with low EBN while outcrossing correlates with high EBN (Brandvain and Haig, 2005; Lafon-Placette and Köhler, 2016). The reason for this correlation remains elusive. In this work, a link between EBN, genomic imprinting, mating system, and hybridization success in *Capsella* is proposed.

This work focused on the three diploid *Capsella* species: the outbreeder *C. grandiflora* (*Cg*) and the early and late selfers *C. orientalis* (*Co*, estimated divergence ~1 Mya) and *C. rubella* (*Cr*, divergence ~100 kya; Douglas *et al.*, 2015). All species were reciprocally crossed with each other. Overall, hybrid seeds from all cross combinations were highly inviable, showing a strong level of reproductive isolation between the three species. Parent-of-origin hybrid seed defects were observed: both, *Co* × *Cr* and *Co* × *Cg* were dark and shrivelled while the reciprocal seeds were extremely small, resembling *Cr* × *Cg* reciprocal hybrid seeds (Rebernig *et al.*, 2015). Endosperm defects are the cause for *Co* × *Cr* and *Co* × *Cg* seed failure like previously described for *Cr* × *Cg* reciprocal hybrid seeds (Rebernig *et al.*, 2015), while endosperm cellularization failed to occur in seeds with *Co* as maternal plant, it occurred precociously in *Cr* × *Co*. The endosperm in *Cg* × *Co* even arrested before this developmental transition. In addition, hybrid embryos between the three species could all be grown to viable plants if dissected out from the seed and incubated on MS medium, confirming that the endosperm is the cause for hybrid seed inviability (Rebernig *et al.*, 2015; Sicard *et al.*, 2015). The phenotype of the endosperm mimics the effects of interploidy hybridization (Scott *et al.*, 1998). Importantly, viable *Cr* × *Cg* hybrid seed could be produced when using a tetraploid *Cr* as maternal parent, strongly suggesting that *Cg* has the highest "effective ploidy" or EBN, followed by *Cr* and *Co*.

The imprintome of all three species was established and compared. The comparison of the number of imprinted genes between species showed that *Cg* has the highest number of PEGs, followed by *Cr* and *Co*, correlating with the EBN ranking of the species. Such correlation did not hold for MEGs, *Cr* has a much higher number of MEGs compared to the two other species. In addition, PEGs were expressed to a higher level in *Cg*, the species with the highest EBN, compared to *Cr*. No significant difference was found for MEGs. Finally, in *Cr* × *Cg* hybrid seeds, PEGs from both, *Cr* and *Cg*, were specifically upregulated. These results suggest that the number and/or the expression level of PEGs are involved in the variation of EBN between *Capsella* species, the primary cause of the defects in the hybrid endosperm.

The molecular cause for the variation in PEG number between species was investigated. Genomic imprinting has been associated with TEs and DNA methylation (Du *et al.*, 2014; Pignatta *et al.*, 2014; Wolff *et al.*, 2011). By comparing imprinted genes and their non-imprinted homologs in other *Capsella* species, it appeared that species-specific PEGs were associated with species-specific CHH methylation in the regions neighboring PEGs. CHH methylation, as mentioned above, is a hallmark of TE silencing. This suggests that differential TE insertions between species drive the emergence of species-

specific PEGs. Consistently, the highest number of PEGs found in *Cg*, followed by *Cr* and *Co*, matches with the highest occurrence of TEs nearby genes in *Cg*, followed by *Cr* and *Co* (Ågren *et al.*, 2014). The different TE abundance between *Cg*, *Cr*, and *Co* is consistent with theoretical predictions proposing that TEs are more efficiently eliminated in selfing species compared to outbreeding species (Boutin *et al.*, 2012). As a consequence, this suggests that the mating system, influencing the rate of TE insertions in the genome, is a strong evolutionary driver of genomic imprinting and the EBN. This could explain the difference of EBNs often found between selfing and outbreeding species (Figure 7; Brandvain and Haig, 2005; Lafon-Placette and Köhler, 2016).

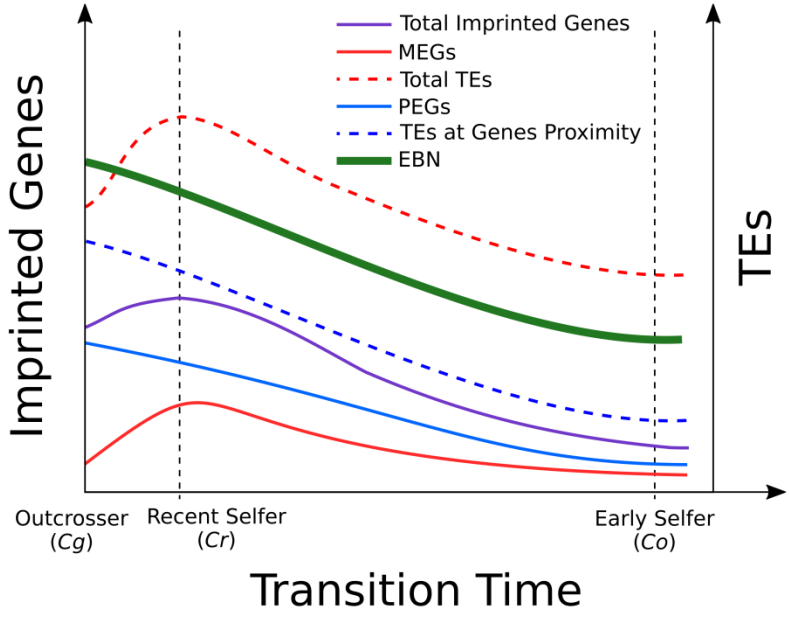

Figure 7. Hypothetical changes in number of imprinted genes and EBN related to species transition time from outcrossing to selfing. Number of imprinted genes correlates with the number of TEs in respective species and decreases as the TEs being purged in species with a longer selfing history. Different to MEGs, PEG numbers directly correlate with the EBN and the number of TEs occupying the proximity of genes.

4 Conclusions

By investigating the evolution of genomic imprinting in the *Capsella* genus, this work strongly supports the hypothesis that genomic imprinting is a rapidly evolving phenomenon in plants. There is low conservation of imprinted genes among the three investigated species, but also the number of imprinted genes and their expression level substantially differ. Genomic imprinting is driven by TEs and host-mediated silencing mechanisms of TEs that impact on the regulation of neighbouring genes. Selfing and outbreeding species differ in number of TEs, providing a compelling explanation for the correlation of imprinted genes with the plant mating type. This work furthermore shows that *Capsella* species are reproductively isolated by an endosperm-based postzygotic hybridization barrier. Each *Capsella* species has a specific EBN that correlates with the number of PEGs, providing a link between EBN, PEGs and mating type. Not only the number of PEGs but also their expression level is substantially higher in outbreeding compared to selfing species, indicating that higher selection pressure in outbreeding species enforces PEG expression. In conclusion, variation in EBN between selfing and outbreeding species is likely a result of different TE dynamics establishing genomic imprinting, or a result of different levels of selection pressure (parental conflict), or both.

5 Future perspectives

Based on the conclusions of this study, genomic imprinting, TEs and EBN are closely connected. Nevertheless, the causal link between TE insertion and establishment of imprinting remains to be demonstrated. A way to test the causality of this relationship could be to create an imprinted gene, i.e. by introducing a TE into the promoter sequence of a gene and testing whether this renders the expression status of the gene.

Another interesting aspect worth of further investigation is genomic imprinting in *C. bursa-pastoris*, an allopolyploid species derived after hybridization of *C. orientalis* and *C. rubella/grandiflora* lineages (Douglas *et al.*, 2015). Its genome therefore consists of two subgenomes, and it would be interesting to know if the two subgenomes retained the genomic imprinting pattern from the parental species, or whether they evolved differently. Notably, TEs in *C. bursa-pastoris* are more abundant in gene-rich regions compared to the parental species (Ågren *et al.*, 2016), which is expected to lead to higher numbers of imprinted genes. Unpublished work by the Köhler research group revealed that *C. rubella* and *C. bursa-pastoris* have similar EBNs, thus, hybridizations between both species lead to viable hybrid seeds. Whether this correlates with similar number and expression of imprinted genes remains an exciting question to be addressed.

Furthermore, substantial evidence points that imprinted genes and/or the regulatory machinery responsible for genomic imprinting are establishing interspecies hybridization barriers; nevertheless, a causal link remains to be identified. Current work in the lab aims at identifying the causal loci underlying the identified QTLs. The identification of those loci will be a major advance for our understanding of the evolution of postzygotic hybridization barriers between related species.

Finally, the knowledge generated in this study has the potential to be utilized in plant breeding, which is worth to be further explored. Many crop species have a different EBN compared to their ancestral varieties, causing

problems when introgressing traits from ancestral varieties into modern crops. Knowing the molecular mechanisms underlying the EBN will allow to bypass this barrier, by either generating mutants in the identified genes or manipulating their expression. The Köhler research group has shown that manipulating DNA methylation levels allows bypassing the interploidy hybridization barrier (Schatlowski *et al.*, 2014). Similar approaches could be used to test whether they suffice to bypass interspecies hybridization barriers established by differences in EBN. Overcoming these barriers is furthermore of high relevance to engineer apomixes, a trait allowing clonal seed production. Inducing clonal embryo formation by genetic manipulation is possible (Marimuthu *et al.*, 2011); however, engineering a functional endosperm is hindered by the parental-dosage sensitivity of the endosperm. In fact, most apomictic species are pseudoapomicts, meaning that the endosperm is developing after fertilization. Dosage sensitivity in the endosperm is built by imprinted genes; therefore, knowing the genes responsible for building dosage sensitivity in the endosperm will have a strong impact to engineer apomictic seed production. Finally, changes in parental genome dosage cause strong effects on seed size, a trait of high relevance in breeding programs. Identifying the causal genes will be a worthwhile effort of future investigations.

References

Ågren, J.A., Huang, H.-R., Wright, S.I., 2016. Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. *Am. J. Bot.* 103, 1197–1202. doi:10.3732/ajb.1600103

Ågren, J.A., Wang, W., Koenig, D., Neuffer, B., Weigel, D., Wright, S.I., 2014. Mating system shifts and transposable element evolution in the plant genus *Capsella*. *BMC Genomics* 15. doi:10.1186/1471-2164-15-602

Ågren, J.A., Wright, S.I., 2011. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosome Res.* 19, 777. doi:10.1007/s10577-011-9229-0

Arias, T., Beilstein, M.A., Tang, M., McKain, M.R., Pires, J.C., 2014. Diversification times among Brassica (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. *Am. J. Bot.* 101, 86–91. doi:10.3732/ajb.1300312

Arnaud, P., Feil, R., 2006. *MEDEA* takes control of its own imprinting. *Cell* 124, 468–470. doi:10.1016/j.cell.2006.01.020

Barlow, D.P., 1993. Methylation and imprinting: from host defense to gene regulation? *Science* 260, 309–310. doi:10.1126/science.8469984

Belmonte, M.F., Kirkbride, R.C., Stone, S.L., Pelletier, J.M., Bui, A.Q., Yeung, E.C., Hashimoto, M., Fei, J., Harada, C.M., Munoz, M.D., Le, B.H., Drews, G.N., Brady, S.M., Goldberg, R.B., Harada, J.J., 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc. Natl. Acad. Sci.* 110, E435–E444. doi:10.1073/pnas.1222061110

Bewick, A.J., Schmitz, R.J., 2017. Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.*, 36 Genome studies and molecular genetics 36, 103–110. doi:10.1016/j.pbi.2016.12.007

Boutin, T.S., Le Rouzic, A., Capy, P., 2012. How does selfing affect the dynamics of selfish transposable elements? *Mob. DNA* 3, 5. doi:10.1186/1759-8753-3-5

Brandvain, Y., Haig, D., 2005. Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. *Am. Nat.* 166, 330–338. doi:10.1086/an.2005.166.issue-3

Briggs, C.L., 1993. Endosperm development in *Solanum nigrum* L. Formation of the zone of separation and secretion. *Ann. Bot.* 72, 303–313. doi:10.1006/anbo.1993.1112

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

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

Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., Martienssen, R.A., 2012. Reprogramming of dna methylation in pollen guides epigenetic inheritance via small RNA. *Cell* 151, 194–205. doi:10.1016/j.cell.2012.09.001

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., Peacock, W.J., 1997. Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 94, 4223–4228.

Chen, C., E, Z., Lin, H.-X., 2016. Evolution and molecular control of hybrid incompatibility in plants. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01208

Chen, X., 2007. A marked end. *Nat. Struct. Mol. Biol.* 14, 259–260. doi:10.1038/nsmb0407-259

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., Fischer, R.L., 2002. DEMETER, a DNA Glycosylase Domain Protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* 110, 33–42. doi:10.1016/S0092-8674(02)00807-3

Chuong, E.B., Elde, N.C., Feschotte, C., 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86. doi:10.1038/nrg.2016.139

Cooper, D.C., Brink, R.A., 1945. Seed collapse following matings between diploid and tetraploid races of *Lycopersicon pimpinellifolium*. *Genetics* 30, 376–401.

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

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Incorporated Publishers.

Debeaujon, I., Léon-Kloosterziel, K.M., Koornneef, M., 2000. Influence of the Testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol.* 122, 403–414.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., Jacobsen, S.E., 2012. Loss of the DNA Methyltransferase MET1 induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. *PLOS Genet.* 8, e1003062. doi:10.1371/journal.pgen.1003062

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., Bischof, S., Gruissem, W., Hennig, L., 2013. *Arabidopsis* MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* 32, 2073–2085. doi:10.1038/emboj.2013.145

Dobzhansky, T., Dobzhansky, T.G., 1937. Genetics and the origin of species. Columbia university press.

Douglas, G.M., Gos, G., Steige, K.A., Salcedo, A., Holm, K., Josephs, E.B., Arunkumar, R., Ågren, J.A., Hazzouri, K.M., Wang, W., Platts, A.E., Williamson, R.J., Neuffer, B., Lascoux, M., Slotte, T., Wright, S.I., 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc. Natl. Acad. Sci.* 112, 2806–2811. doi:10.1073/pnas.1412277112

Drews, G.N., Yadegari, R., 2002. development and function of the angiosperm female gametophyte. *Annu. Rev. Genet.* 36, 99–124. doi:10.1146/annurev.genet.36.040102.131941

Du, M., Luo, M., Zhang, R., Finnegan, E.J., Koltunow, A.M.G., 2014. Imprinting in rice: the role of DNA and histone methylation in modulating parent-of-origin specific expression and determining transcript start sites. *Plant J.* 79, 232–242. doi:10.1111/tpj.12553

Ehlenfeldt, M.K., Hanneman, R.E., 1988. Genetic control of Endosperm Balance Number (EBN): three additive loci in a threshold-like system. *Theor. Appl. Genet.* 75, 825–832. doi:10.1007/BF00258041

Fedoroff, N.V., 2012. Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–767. doi:10.1126/science.338.6108.758

Figueiredo, D.D., Batista, R.A., Roszak, P.J., 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, D.D., Batista, R.A., Roszak, P.J., Köhler, C., 2015. Auxin production couples endosperm development to fertilization. *Nat. Plants* 1, nplants2015184. doi:10.1038/nplants.2015.184

Figueiredo, D.D., Köhler, C., 2016. Bridging the generation gap: communication between maternal sporophyte, female gametophyte and fertilization products. *Curr. Opin. Plant Biol.* 29, 16–20. doi:10.1016/j.pbi.2015.10.008

Figueiredo, D.D., Köhler, C., 2014. Signalling events regulating seed coat development. *Biochem. Soc. Trans.* 42, 358–363. doi:10.1042/BST20130221

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

Fuller, D.Q., 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the old world. *Ann. Bot.* 100, 903–924. doi:10.1093/aob/mcm048

Fultz, D., Choudury, S.G., Slotkin, R.K., 2015. Silencing of active transposable elements in plants. *Curr. Opin. Plant Biol.* 27, 67–76. doi:10.1016/j.pbi.2015.05.027

Garner, A.G., Kenney, A.M., Fishman, L., Sweigart, A.L., 2016. Genetic loci with parent-of-origin effects cause hybrid seed lethality in crosses between *Mimulus* species. *New Phytol.* 211, 319–331. doi:10.1111/nph.13897

Gehring, M., Henikoff, S., 2008. DNA methylation and demethylation in *Arabidopsis*. *Arab. Book Am. Soc. Plant Biol.* 6. doi:10.1199/tab.0102

Gehring, M., Missirian, V., Henikoff, S., 2011. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* 6, e23687. doi:10.1371/journal.pone.0023687

Gehring, M., Satyaki, P.R., 2017. Endosperm and imprinting, inextricably linked. *Plant Physiol.* 173, 143–154. doi:10.1104/pp.16.01353

Gill, D.B.S., Waines, J.G., 1978. Paternal regulation of seed development in wheat hybrids. *Theor. Appl. Genet.* 51, 265–270. doi:10.1007/BF00274813

Golan, G., Oksenberg, A., Peleg, Z., 2015. Genetic evidence for differential selection of grain and embryo weight during wheat evolution under domestication. *J. Exp. Bot.* 66, 5703–5711. doi:10.1093/jxb/erv249

Goulet, B.E., Roda, F., Hopkins, R., 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiol.* 173, 65–78. doi:10.1104/pp.16.01340

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M.A., Gagliano, W.B., 1998. Maternal control of embryogenesis by *MEDEA*, a Polycomb Group Gene in *Arabidopsis*. *Science* 280, 446–450. doi:10.1126/science.280.5362.446

Guo, Y.-L., Bechsgaard, J.S., Slotte, T., Neuffer, B., Lascoux, M., Weigel, D., Schierup, M.H., 2009. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc. Natl. Acad. Sci.* 106, 5246–5251. doi:10.1073/pnas.0808012106

Gutierrez-Marcos, J.F., Pennington, P.D., Costa, L.M., Dickinson, H.G., 2003. Imprinting in the endosperm: a possible role in preventing wide hybridization. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 358, 1105–1111. doi:10.1098/rstb.2003.1292

Haig, D., 2004. Genomic imprinting and kinship: how good is the evidence? *Annu. Rev. Genet.* 38, 553–585. doi:10.1146/annurev.genet.37.110801.142741

Haig, D., 2000. The kinship theory of genomic imprinting. *Annu. Rev. Ecol. Syst.* 31, 9–32. doi:10.1146/annurev.ecolsys.31.1.9

Haig, D., Westoby, M., 1989. Parent-specific gene expression and the triploid endosperm. *Am. Nat.* 134, 147–155.

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

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., Zilberman, D., 2009. Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324, 1451–1454. doi:10.1126/science.1172417

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., Fischer, R.L., 2011. Inaugural article: regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl. Acad. Sci.* 108, 1755–1762. doi:10.1073/pnas.1019273108

Hurka, H., Friesen, N., German, D.A., Franzke, A., Neuffer, B., 2012. 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). *Mol. Ecol.* 21, 1223–1238. doi:10.1111/j.1365-294X.2012.05460.x

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R.L., Tamaru, H., Zilberman, D., 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* 337, 1360–1364. doi:10.1126/science.1224839

Igic, B., Lande, R., Kohn, J.R., 2008. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* 169, 93–104. doi:10.1086/523362

Ingram, G., Gutierrez-Marcos, J., 2015. Peptide signalling during angiosperm seed development. *J. Exp. Bot.* 66, 5151–5159. doi:10.1093/jxb/erv336

Ingram, G.C., 2010. Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* 247, 195–214. doi:10.1007/s00709-010-0184-y

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., Kinoshita, T., 2011. Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm

without change to the rate of syncytial nuclear division. *Plant J.* 65, 798–806. doi:10.1111/j.1365-313X.2010.04466.x

Jahnke, S., Scholten, S., 2009. Epigenetic resetting of a gene imprinted in plant embryos. *Curr. Biol.* CB 19, 1677–1681. doi:10.1016/j.cub.2009.08.053

Jiang, H., Köhler, C., 2012. Evolution, function, and regulation of genomic imprinting in plant seed development. *J. Exp. Bot.* 63, 4713–4722. doi:10.1093/jxb/ers145

Johnston, S.A., den Nijs, T.P., Peloquin, S.J., Hanneman, R.E., 1980. The significance of genic balance to endosperm development in interspecific crosses. *Theor. Appl. Genet.* 57, 5–9. doi:10.1007/BF00276002

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

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

Jullien, P.E., Kinoshita, T., Ohad, N., Berger, F., 2006. Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. *Plant Cell* 18, 1360–1372. doi:10.1105/tpc.106.041178.

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

Kim, N.-S., 2017. The genomes and transposable elements in plants: are they friends or foes? *Genes Genomics* 39, 359–370. doi:10.1007/s13258-017-0522-y

Kirkbride, R.C., Yu, H.H., Nah, G., Zhang, C., Shi, X., Chen, Z.J., 2015. An epigenetic role for disrupted paternal gene expression in postzygotic seed abortion in *Arabidopsis* interspecific hybrids. *Mol. Plant* 8, 1766–1775. doi:10.1016/j.molp.2015.09.009

Klosinska, M., Picard, C.L., Gehring, M., 2016. Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat. Plants* 2, 16145. doi:10.1038/nplants.2016.145

Koch, M.A., Kiefer, M., 2005. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species - *Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. *Am. J. Bot.* 92, 761–767. doi:10.3732/ajb.92.4.761

Kradolfer, D., Hennig, L., Köhler, C., 2013a. Increased maternal genome dosage bypasses the requirement of the FIS Polycomb Repressive Complex 2 in *Arabidopsis* seed development. *PLoS Genet* 9, e1003163. doi:10.1371/journal.pgen.1003163

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

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

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

Lafon-Placette, C., Köhler, C., 2014. Embryo and endosperm, partners in seed development. *Curr. Opin. Plant Biol.* 17, 64–69. doi:10.1016/j.pbi.2013.11.008

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi:10.1038/nrg2719

Lee, J.H., Cooper, D.C., 1958. Seed development following hybridization between diploid *Solanum* species from Mexico, Central and South America. *Am. J. Bot.* 45, 104–110. doi:10.2307/2439357

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., Chaudhury, A., 2000. Expression and parent-of-origin effects for *FIS2*, *MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc. Natl. Acad. Sci.* 97, 10637–10642. doi:10.1073/pnas.170292997

Luo, M., Taylor, J.M., 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 Genet* 7, e1002125. doi:10.1371/journal.pgen.1002125

Marimuthu, M.P.A., Jolivet, S., Ravi, M., Pereira, L., Davda, J.N., Cromer, L., Wang, L., Nogué, F., Chan, S.W.L., Siddiqi, I., Mercier, R., 2011. Synthetic clonal reproduction through seeds. *Science* 331, 876. doi:10.1126/science.1199682

Martínez, G., Panda, K., Köhler, C., Slotkin, R.K., 2016. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nat. Plants* 2, nplants201630. doi:10.1038/nplants.2016.30

Matzke, M.A., Mosher, R.A., 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408. doi:10.1038/nrg3683

Mayr, E., 1996. What is a species, and what is not? *Philos. Sci.* 63, 262–277.

Mercader, J., 2009. Mozambican Grass Seed Consumption During the Middle Stone Age. Science 326, 1680–1683. doi:10.1126/science.1173966

Mitchell-Olds, T., 2001. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol. Evol.* 16, 693–700. doi:10.1016/S0169-5347(01)02291-1

Moreno-Romero, J., Jiang, H., Santos-González, J., Köhler, C., 2016. Parental epigenetic asymmetry of PRC2-mediated histone modifications in the *Arabidopsis* endosperm. *EMBO J.* 35, 1298–1311. doi:10.15252/embj.201593534

Mozgova, I., Köhler, C., Hennig, L., 2015. Keeping the gate closed: functions of the polycomb repressive complex PRC2 in development. *Plant J.* 83, 121–132. doi:10.1111/tpj.12828

Muller, H.J., 1942. Isolating Mechanisms, Evolution, and Temperature. Biological Symposia. *Isol. Mech. Evol. Temp. Biol. Symp.* 71–125.

Muller, H.J., 1940. Bearings of the "*Drosophila*" work on systematics. *New Syst.* 185–268.

Niederhuth, C.E., Schmitz, R.J., 2017. Putting DNA methylation in context: from genomes to gene expression in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 149–156. doi:10.1016/j.bbagrm.2016.08.009

Nodine, M.D., Bartel, D.P., 2012. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* 482, 94–97. doi:10.1038/nature10756

Nowack, M.K., Shirzadi, R., Dissmeyer, N., Dolf, A., Endl, E., Grini, P.E., Schnittger, A., 2007. Bypassing genomic imprinting allows seed development. *Nature* 447, 312–315. doi:10.1038/nature05770

Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N., Slotkin, R.K., 2013. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol.* 162, 116–131. doi:10.1104/pp.113.216481

Olsen, O.-A., 2004. Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16, S214–S227. doi:10.1105/tpc.017111

Paetsch, M., Mayland-Quellhorst, S., Neuffer, B., 2006. Evolution of the self-incompatibility system in the Brassicaceae: identification of S-locus receptor kinase (SRK) in self-incompatible *Capsella grandiflora*. *Heredity* 97, 283–290. doi:10.1038/sj.hdy.6800854

Park, K., Kim, M.Y., Vickers, M., Park, J.-S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., Scholten, S., 2016. DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. *Proc. Natl. Acad. Sci.* 113, 15138–15143. doi:10.1073/pnas.1619047114

Parrott, W.A., Smith, R.R., 1986. Evidence for the existence of endosperm balance number in the true clovers (*Trifolium* spp.). *Can. J. Genet. Cytol.* 28, 581–586. doi:10.1139/g86-085

Patten, M.M., Úbeda, F., Haig, D., 2013. Sexual and parental antagonism shape genomic architecture. *Proc. R. Soc. Lond. B Biol. Sci.* 280, 20131795. doi:10.1098/rspb.2013.1795

Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., Gehring, M., 2014. Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife* e03198. doi:10.7554/eLife.03198

Pignocchi, C., Minns, G.E., Nesi, N., Koumproglou, R., Kitsios, G., Benning, C., Lloyd, C.W., Doonan, J.H., Hills, M.J., 2009. ENDOSPERM DEFECTIVE1 is a novel microtubule-associated protein essential for seed development in *Arabidopsis*. *Plant Cell* 21, 90–105. doi:10.1105/tpc.108.061812

Pires, N.D., Grossniklaus, U., 2014. Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Rep.* 6. doi:10.12703/P6-63

Preece, C., Livarda, A., Christin, P.-A., Wallace, M., Martin, G., Charles, M., Jones, G., Rees, M., Osborne, C.P., 2017. How did the domestication of Fertile Crescent grain crops increase their yields? *Funct. Ecol.* 31, 387–397. doi:10.1111/1365-2435.12760

Radchuk, V., Borisjuk, L., 2014. Physical, metabolic and developmental functions of the seed coat. *Front. Plant Sci.* 5. doi:10.3389/fpls.2014.00510

Raissig, M.T., Bemer, M., Baroux, C., Grossniklaus, U., 2013. Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genet.* 9, e1003862. doi:10.1371/journal.pgen.1003862

Rebernig, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T., Köhler, C., 2015. Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. *PLoS Genet.* 11, e1005295. doi:10.1371/journal.pgen.1005295

Rodrigues, J.A., Zilberman, D., 2015. Evolution and function of genomic imprinting in plants. *Genes Dev.* 29, 2517–2531. doi:10.1101/gad.269902.115

Sabelli, P.A., Larkins, B.A., 2009. The development of endosperm in grasses. *Plant Physiol.* 149, 14–26. doi:10.1104/pp.108.129437

Sansome, E.R., Satina, S., Blakeslee, A.F., 1942. Disintegration of ovules in tetraploid-diploid and in incompatible species crosses in *Datura*. *Bull. Torrey Bot. Club* 69, 405–420. doi:10.2307/2481421

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*. *Plant Cell* tpc.114.130120. doi:10.1105/tpc.114.130120

Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.-S., Park, G.T., Fischer, R.L., Tamaru, H., 2011. Function of the DEMETER DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. *Proc. Natl. Acad. Sci.* 108, 8042–8047. doi:10.1073/pnas.1105117108

Schon, M.A., Nodine, M., 2017. Widespread contamination of *Arabidopsis* embryo and endosperm transcriptome datasets. *Plant Cell* tpc.00845.2016. doi:10.1105/tpc.16.00845

Scott, R.J., Spielman, M., Bailey, J., Dickinson, H.G., 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Dev. Camb. Engl.* 125, 3329–3341.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C.S., Eroukhmanoff, F., Feder, J.L., Fischer, M.C., Foote, A.D., Franchini, P., Jiggins, C.D., Jones, F.C., Lindholm, A.K., Lucek, K., Maan, M.E., Marques, D.A., Martin, S.H., Matthews, B., Meier, J.I., Möst, M., Nachman, M.W., Nonaka, E., Rennison, D.J., Schwarzer, J., Watson, E.T., Westram, A.M., Widmer, A., 2014. Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176–192. doi:10.1038/nrg3644

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

Sharma, D.R., Kaur, R., Kumar, K., 1996. Embryo rescue in plants - a review. *Euphytica* 89, 325–337. doi:10.1007/BF00022289

Sicard, A., Kappel, C., Josephs, E.B., Lee, Y.W., Marona, C., Stinchcombe, J.R., Wright, S.I., Lenhard, M., 2015. Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in *Capsella*. *Nat. Commun.* 6, 7960. doi:10.1038/ncomms8960

Sicard, A., Stacey, N., Hermann, K., Dessoly, J., Neuffer, B., Bäurle, I., Lenhard, M., 2011. Genetics, evolution, and adaptive significance of the selfing syndrome in the genus *Capsella*. *Plant Cell* 23, 3156–3171. doi:10.1105/tpc.111.088237

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A., Martienssen, R.A., 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136, 461–472. doi:10.1016/j.cell.2008.12.038

Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.-L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K., Wang, W., Mandáková, T., Vello, E., Smith, L.M., Henz, S.R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T.T., Blanchette, M., Clark, R.M., Quesneville, H., Nordborg, M., Gaut, B.S., Lysak, M.A., Jenkins, J., Grimwood, J., Chapman, J., Prochnik, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D.,

Wright, S.I., 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831–835. doi:10.1038/ng.2669

Snir, A., Nadel, D., Groman-Yaroslavski, I., Melamed, Y., Sternberg, M., Bar-Yosef, O., Weiss, E., 2015. The origin of cultivation and proto-weeds, long before neolithic farming. *PLoS ONE* 10, e0131422. doi:10.1371/journal.pone.0131422

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., Jacobsen, S.E., 2014. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* 21, 64–72. doi:10.1038/nsmb.2735

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., Jacobsen, S.E., 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* 152, 352–364. doi:10.1016/j.cell.2012.10.054

Swamy, B.G.L., Parameswaran, N., 1963. The Helobial Endosperm. *Biol. Rev.* 38, 1–50. doi:10.1111/j.1469-185X.1963.tb00653.x

Underwood, C.J., Henderson, I.R., Martienssen, R.A., 2017. Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr. Opin. Plant Biol.* 36, 135–141. doi:10.1016/j.pbi.2017.03.002

Valentine, D.H., Woodell, S.R.J., 1963. Studies in British primulas. X. Seed incompatibility in intraspecific and interspecific crosses at diploid and tetraploid levels. *New Phytol.* 62, 125–143.

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

von Wangenheim, K.-H.F., 1957. Untersuchungen über den zusammenhang zwischen chromosomenzahl und kreuzbarkeit bei *Solanum*-arten. *Z. Für Indukt. Abstamm.-Vererbungslehre* 88, 21–37. doi:10.1007/BF00593652

Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A., Berger, F., Peacock, W.J., Dennis, E.S., Luo, M., 2010. The VQ motif protein IKU1 regulates endosperm growth and seed size in *Arabidopsis*. *Plant J.* 63, 670–679. doi:10.1111/j.1365-313X.2010.04271.x

Waters, A.J., Bilinski, P., Eichten, S.R., Vaughn, M.W., Ross-Ibarra, J., Gehring, M., Springer, N.M., 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl. Acad. Sci.* 201309182. doi:10.1073/pnas.1309182110

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., Köhler, C., 2010. H3K27me3 profiling of the endosperm implies exclusion of Polycomb Group Protein targeting by DNA methylation. *PLoS Genet.* 6, e1001152. doi:10.1371/journal.pgen.1001152

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

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M., Köhler, C., 2011. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet.* 7, e1002126. doi:10.1371/journal.pgen.1002126

Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., Takita, M.A., Labadie, K., Poulain, J., Couloux, A., Jabbari, K., Cattonaro, F., Del Fabbro, C., Pinosio, S., Zuccolo, A., Chapman, J., Grimwood, J., Tadeo, F.R., Estornell, L.H., Muñoz-Sanz, J.V., Ibanez, V., Herrero-Ortega, A., Aleza, P., Pérez-Pérez, J., Ramón, D., Brunel, D., Luro, F., Chen, C., Farmerie, W.G., Desany, B., Kodira, C., Mohiuddin, M., Harkins, T., Fredrikson, K., Burns, P., Lomsadze, A., Borodovsky, M., Reforgiato, G., Freitas-Astúa, J., Quetier, F., Navarro, L., Roose, M., Wincker, P., Schmutz, J., Morgante, M., Machado, M.A., Talon, M., Jaillon, O., Ollitrault, P., Gmitter, F., Rokhsar, D., 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* 32, 656–662. doi:10.1038/nbt.2906

Xu, W., Dai, M., Li, F., Liu, A., 2014. Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. *Nucleic Acids Res.* gku375. doi:10.1093/nar/gku375

Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA Methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193–205. doi:10.1016/j.cell.2013.02.033

Zhang, M., Li, N., He, W., Zhang, H., Yang, W., Liu, B., 2016. Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J.* 85, 424–436. doi:10.1111/tpj.13116

Zhang, M., Xie, S., Dong, X., Zhao, X., Zeng, B., Chen, J., Li, H., Yang, W., Zhao, H., Wang, G., Chen, Z., Sun, S., Hauck, A., Jin, W., Lai, J., 2014. Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res.* 24, 167–176. doi:10.1101/gr.155879.113

Popular science summary

Many crops important for human civilization are hybrids, derived after hybridization of different species, varieties, or cultivars. Hybrids often outperform their parents by having increased growth rate, yield, and endurance to stress. However, the production of hybrids is laborious and often fails due to the lethality of seeds produced from crosses between related species.

This study aimed at understanding the molecular mechanisms underlying hybridization success or failure. Especially, it focused on the endosperm, the nourishing part of the seed, which is known to develop abnormally in hybrid seeds. The study models were three species of the Shepherd's purse (*Capsella*) genus. The data presented here suggest that epigenetic mechanisms regulating mobile DNA elements are causally involved in hybrid seed lethality. These regulatory pathways cause genes to be active only when inherited from one of the parents, either the mother or the father. The epigenetic modification, also termed imprint, is applied during gamete formation and maintained after fertilization. This study revealed that closely related species differ in number and activity of imprinted genes, explaining that hybrid seed defects differ depending on which species served as male or female parent. Thus, the results of this work suggest that imprinted genes can serve as a predictor for the hybridization success between species.

Populärvetenskaplig sammanfattning

Många grödor som är viktiga för mänskligheten är hybrider som kan härledas från korsningar mellan olika arter och sorter. Hybrider överträffar ofta sina föräldrar genom att ha snabbare tillväxt, högre avkastning och stresstålighet. Framställning av hybrider är dock mödosam och misslyckas ofta på grund av att frön från korsningen aborteras.

Målet med den här studien var att förstå de molekylära mekanismer som avgör om en hybridisering lyckas eller inte. Studien fokuserade särskilt på endospermet, den näringsgivande delen av fröet, vilket har visat sig utvecklas onormalt i hybridfrön. Modellväxterna bestod av tre olika *Capsella*arter (lommeört). De data som presenteras här tyder på att epigenetiska mekanismer som reglerar mobila DNA-element är involverade i frödödligheten hos hybrider. Dessa regleringsvägar gör gener aktiva beroende på om de ärvs från moder- eller faderplantan. Den epigenetiska modifieringen, även kallad ”imprint”eller prägling, sker under gametformationen och upprätthålls efter befruktningen. Denna studie visade att närbesläktade arter skiljer sig åt beträffande antal imprintade gener och aktiviteten hos dessa och att hybridiseringsdefekter skiljer sig åt beroende på vilken art som fungerar som moder- respektive faderplanta. Således antyder resultatet av studien att imprintade gener kan studeras för att förutsäga hybridiseringsframgången mellan olika arter.

Rangkuman sains populer

Banyak tanaman penting bagi peradaban manusia merupakan hibrid dari dua spesies, varietas, atau kultivar. Tanaman hibrida menampilkan karakter yang cenderung melampaui induknya, seperti meningkatnya laju pertumbuhan, hasil, kekebalan terhadap penyakit, dan ketahanan terhadap stress. Namun, produksi tanaman hibrida sering kali gagal karena benih-benih letal yang dihasilkan dari penyilangan dua tanaman yang berkekerabatan dekat.

Studi ini ditujukan untuk memahami apa yang menyebabkan kecocokan dua tanaman dalam menghasilkan benih-benih hibrida hidup pada tingkat molekular. Studi ini difokuskan pada endosperma, sumber gizi yang menghidupi benih, yang juga diketahui menyebabkan abnormalitas pada benih hibrida. Disertasi ini menggunakan model studi dari tiga spesies tanaman dompet gembala (*Capsella*). Data yang disuguhkan disini menunjukkan bahwa mekanisme epigenetik pengatur elemen DNA bergerak terlibat sebagai penyebab dalam letalitas benih hibrida. Lintasan pengatur ini menyebabkan gen menjadi aktif hanya ketika diwariskan oleh salah satu induk, oleh ibu atau bapak. Modifikasi epigenetik ini, yang juga disebut rekaman, terjadi ketika pembentukan gamet dan dipelihara setelah pembuahan. Studi ini mengungkapkan bahwa spesies berkekerabatan dekat berbeda dalam hal jumlah dan aktivitas gen-gen terekam; menjelaskan bahwa kecacatan berbeda pada benih hibrida tergantung pada spesies mana yang dijadikan induk jantan atau betina. Pada akhirnya, karya ini mengusulkan bahwa gen-gen terekam dapat dijadikan sebagai penaksir kesuksesan hibridisasi antarspesies.

Acknowledgements

After a long and tedious period of study, I'm finally able to finish this project in three diploid *Capsella* species. I would like to deliver my sincere gratitude for everyone who has been involved in my study.

Thanks to my wife, Vina, and my family who with all their help, love, and understanding, made me able to overcome my problems. Especially for my wife, your cooking keeps me alive.

Thanks to my supervisor group. Firstly to Claudia, for her patience and care in guiding me every time I stumbled on some problem. This project won't be successful without your keen supervision and ideas. Thanks to Tanja and his wonderful research team, Kim and Ben, for helping me a lot especially on the evolutionary part. Thanks to Sofia for guiding me especially on my second manuscript.

Big thanks to the angelic trio, Clément, Duarte, and Rita. All of you have become a very important aspect for keeping me up every time a huge obstacle seems appear at every step of my research. This dissertation won't be possible without your valuable input, mentoring, guidelines, and company. Thanks also to my lunch mate Lejon, Lauriane, Eva, Jeni, Miyuki, Iva, and Catarina for all the motivating and complex discussion.

Thanks to all people in Claudia's and Lars' group. Thanks Juan, Lei, and Minerva for providing support on the algorithm and coding part. Thanks to Matt for your assistance with the Bioanalyzer. Thanks Cecilia for providing all the things that I need in the lab. Thanks to the oriental group, Hua, Jun, Wenjia, and Zhenxing for making me feel almost at home. Thanks German for your input in the epigenetics aspects. Special thanks to another trio Jordi, Nici, and Caroline for the tutorial on how to do things with *Capsella*, its endosperm, and its library.

Thanks also to the student teaching group, Jim, Girma, Pascal, Sultana, Daniel, Prashanth, Krist, and Anna. All of you have made my life easier each time I need to handle the students. And finally, thanks for all members of the

Bio Center and SLU, Marie, Daniel, Peter, Eva, Qing, Lotta, Urban, Per, Björn, Mona, Anneli, and others that have supported me by giving a very friendly environment to study.

And I almost forget, thanks to the “Big X Factor”. I’m so grateful that nothing more wrong than this can be happened in this path of my life.