Endosperm-based post-zygotic hybridization barriers in *Arabidopsis*

Ida Marie Johannessen

Dissertation presented for the degree of
*Philosophiae Doctor* (PhD)
2019

Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo
Acknowledgements

I would like to start by thanking all my colleagues for being who they are, joining in on the scientific discussions and the often not so serious lunches and coffee breaks. These events made the everyday work life over the past four and a half years quite nice. Most importantly, many thanks to my supervisors prof. Paul Grini, prof. Anne Brysting and dr. Katrine Bjørkan for all your feedback and time, and putting up with me all these years. I thank you for letting me develop into the person I am today (science-wise). It has been four and a half intense, and extremely interesting, years. Thank you for giving me this opportunity.

My work would not have been possible in the way that it has been, without the important and great work of our technical staff. Thank you so much Solveig, Roy, Ingrid, Marit, Bjørn and Chiara. You made my work so much easier and our plant facilities would not be up and running if it were not for you.

Friends and family have been a constant source of support, and comfort when needed. I would like to thank my mother and father for being there for me in every way they could think of. A special thanks to my fiancé whom have been a vital supporter both in work related matters and everything else. Without you, it would have been very much harder to get to the end of this quite long journey. Thanks to everybody that has been cheering me on and believing in me.

Ida Marie Johannessen

Oslo, January 2019
# Table of Contents

Acknowledgements .................................................................................................................................. I

List of Papers........................................................................................................................................ V

Summary .............................................................................................................................................. VII

Introduction ............................................................................................................................................. 1

Seed development................................................................................................................................ 1

The *Arabidopsis* genus ....................................................................................................................... 2

Presence and bypass of *Arabidopsis* post-zygotic hybridization barriers ........................................ 4

Triploid block and bridge ..................................................................................................................... 5

Parental specific effects......................................................................................................................... 5

Silencing mechanisms and parent of origin gene expression ......................................................... 6

MADS box type I transcription factors .............................................................................................. 7

Endosperm Balance Number (EBN) theory .......................................................................................... 8

Aim of Study ......................................................................................................................................... 9

Results and Discussion .......................................................................................................................... 11

* A. arenosa* and *A. lyrata* compatibility investigated by different cross combinations and populations ................................................................. 11

The diploid barrier can be bypassed by genome increase in a species dependent manner ............ 13

The hybridization barrier is stronger between non-introgressed tetraploids .............................. 14

The post-zygotic hybridization barrier between *A. lyrata* and *A. arenosa* is endosperm-based .... 15

The barrier between *A. thaliana* and *A. arenosa* is population and temperature dependent ...... 19

MADS box type I genes and their role in the endosperm-based hybridization barrier ................... 20

Regulation of MADS box type I genes ............................................................................................... 20

MADS box type I genes are regulated in incompatible Arabidopsis hybrids ................................ 22

Possible interaction between MADS box type I TFs in *A. arenosa* and *A. lyrata* ....................... 25

Not just a triploid block, but also a triploid bridge .......................................................................... 26

The Endosperm Balance Number (EBN) theory and the observed role of introgression is conflicting .......................................................................................................................... 27

Possible regulatory mechanism for deleterious gene expression in incompatible hybrids .......... 29

Conclusion ........................................................................................................................................... 31

Future perspectives ............................................................................................................................. 33

List of References ............................................................................................................................... 34

Paper I

Paper II

Paper III
List of Papers

Paper I  Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe

Paper II  Function and conservation of imprinted MADS-box transcription factors in the genus *Arabidopsis*

Paper III  Genetic and molecular analysis of the *Arabidopsis lyrata* and *Arabidopsis arenosa* hybridization barrier

*Joint first authors*
Summary

Investigations of post-zygotic hybridization barriers in *Arabidopsis* have led us to a more detailed understanding of the endosperm-based barrier that prevents different *Arabidopsis* species from producing viable offspring. In Paper I, we describe the barrier between diploid *A. lyrata* and diploid *A. arenosa* as an endosperm-based barrier, where the timing of the endosperm cellularization is affected. When diploid *A. arenosa* acts as seed donor, the endosperm of the hybrid seeds cellularizes early resulting in small and unviable seeds. When diploid *A. arenosa* act as pollen donor, the endosperm cellularizes late or not at all, resulting in collapsed seeds. In both crossing directions, very few seeds germinate, showing a near complete post-zygotic hybridization barrier. We found a bypass of the hybridization barrier by increasing the genome dosage of *A. lyrata*, from diploid to tetraploid, resulting in similar seed germination frequencies as in intraspecies crosses of the parents. The seeds of these rescued crosses also show similar timing of endosperm cellularization as the parents, i.e. after 9 days after pollination (DAP), when the embryo is at the globular stage, and before 15 DAP, when the embryo is around late heart stage. Interestingly, the hybridization barrier is not bypassed when increasing the genome dosage of *A. arenosa* from diploid to tetraploid. In this case, the resulting seed phenotype is very small seeds when tetraploid *A. arenosa* acts as seed donor and collapsed seeds when tetraploid *A. arenosa* acts as pollen donor. Germination frequencies at 0% and 1%, respectively, show a complete hybridization barrier. In conclusion, although the success of the cross between diploid *A. arenosa* and tetraploid *A. lyrata* is not due to increased parental genome dosage in general, it indicates that tetraploidization of *A. lyrata* is sufficient for the reciprocal bypass of the hybridization barrier. It has been hypothesized that the success of interspecies crosses depends on whether the parental species have contributions to the endosperm in the correct maternal to paternal ratio or not (Endosperm Balance Number (EBN)). In line with this, we suggest that the incompatibility of diploid *A. lyrata* and tetraploid *A. arenosa* is due to a difference in EBN, seen as a non-reciprocal effect on endosperm development, underpinning that the hybridization barrier is linked to parental contribution to endosperm development.

In Paper II, we describe a detailed expression pattern of MADS box type I transcription factors (TFs), and the effects of regulation mechanisms, in *A. thaliana* during seed development. The TFs show an expression peak before the onset of, or during, cellularization. In addition, we show conserved imprinting status of *AGL36* in *A. arenosa*, while the imprinting status varies in contrasting ways in different *Arabidopsis* hybrids. Surprisingly, we find that the hybridization barrier between *A. thaliana* and *A. arenosa* is temperature dependent, with less seeds germinating after developing under 22 °C than under 18 °C.

Finally, in Paper III we investigate the MADS box type I TFs at 9 and 15 DAP in developing seeds in *A. arenosa*, *A. lyrata* and their hybrids, revealing that the incompatible hybrids exhibit a non-

VII
reciprocal, deviating expression pattern. Seeds from the incompatible hybrids with maternal excess linked phenotypes (early cellularization) show a low gene expression of these genes, indicating that the expression decreased early. In contrast, the incompatible hybrids with paternal excess linked phenotypes (cellularization late, or not at all) show a continued expression, similar to what is observed before normal cellularization. In addition, there is an absence of expression associated with ongoing and completed endosperm cellularization. Our results suggests that MADS box type TFs are linked to endosperm development and timing of cellularization, and that the expression of some of the genes are strongly affected in incompatible hybrids. The compatible hybrids show a MADS box type I gene expression comparable with the parents. In order to investigate alternative explanations for the bypass of the barrier, we show that all interploidy F1 plants are triploid, ruling out the involvement of unreduced gametes in their formation. The mature F1 hybrids from reciprocal crosses between diploid *A. arenosa* and tetraploid *A. lyrata* were backcrossed to the parents for two generations. The ploidy of the backcrossed offspring shows that the mature F1 hybrid plants are semi-fertile and produce a range of aneuploid as well as some haploid and diploid gametes, indicating that chromosome number irregularities are tolerated in *Arabidopsis*. Further investigations of the bypass of the barrier conducted using colchicine-induced tetraploids (*Paper I*) and plants from both introgressed and non-introgressed natural populations (*Paper I and III*) confirm that the hybridization barrier between diploid *A. arenosa* and tetraploid *A. lyrata* is not bypassed due to introgressed genetic elements in the tetraploid. Finally, we find that the hybridization barrier between tetraploid *A. lyrata* and tetraploid *A. arenosa* is stronger when using non-introgressed compared to introgressed plants (*Paper III*), and comparable to what is seen at the diploid level (*Paper I*).
Introduction

Species often originate by a bifurcating process leading to sister linages that diverge over time by natural selection and genetic drift (Grant 1981). The biological species concept as defined by Mayr (1942) states that species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups. Mayr (1970, 1996) revised his definition adding that isolating mechanisms are “biological properties of individuals which prevent the interbreeding [fusion] of populations”. This species concept has often been challenged in the world of plants due to high degree of hybridization, which can lead to new species or merge existing ones (Abbott et al. 2013). Today we know about approximately 369 000 species of flowering plants, which makes up more than 90 percent of all land plants (Lughadha et al. 2016). The dominance of this plant group, combined with high frequencies of hybridization, makes the flowering plants a manifold, colorful and complicated part of life.

Despite the high degree of hybridization, plant species are typically isolated by multiple barriers, preventing hybridization. These act either before (pre-zygotic barriers) or after fertilization (post-zygotic barriers) (Widmer et al. 2009). Pre-zygotic barriers include pre-pollination barriers that limit the transfer of pollen from individuals of one species to stigmas of other species as for instance ecological or temporal barriers, and post-pollination barriers that act after pollination but before fertilization, such as the failure of gametes from non-conspecific pollen in fertilizing eggs. Post-zygotic barriers act after fertilization and can take the form of seed development arrest and seeds that do not germinate, and the production of inviable or infertile offspring (Coyne and Orr 2004, Rieseberg and Blackman 2010). Learning more about how new species form, and how they are prevented from crossing with other species, can aid our understanding of current biodiversity as well as how new combinations of plant species, e.g. through artificial crossing, can tackle ongoing and future food and cultivating challenges. Analyzing the mechanism of post-zygotic reproductive isolation is a step in the direction of these learning outcomes.

Seed development

After the event of normal fertilization in a cross, the zygote and the endosperm are formed, the seed develops and germinates when mature and triggered by the right conditions, and the seedling grows into a viable, fertile plant. Flowering plants have in common the process of double fertilization where two sperm cells are delivered by pollen to the female gametophyte. One sperm cell fertilizes the egg cell, giving rise to the zygote that develops into the embryo, and the second sperm cell fertilizes the central cell, leading to the nutritious tissue, the endosperm (Figure 1). In Arabidopsis (and many other plants) the central cell is a fusion of two female gametes generated by mitosis (homodiploid), making the fertilized endosperm triploid with two genome copies from the mother and one genome copy from the
father. The endosperm development starts with a nuclear phase, where the endosperm nuclei divides without the successive formation of cell walls (Olsen 2004). The dominating and most stable form of endosperm development in flowering plants is the nuclear mode (Geeta 2003).

![Diagram of Seed Development in Arabidopsis](image)

**Figure 1 Seed development in Arabidopsis from mature female gametophyte to mature seed.** (A) The female gametophyte with the central cell with two polar nuclei, which fuse together before fertilization. a, antipodes; ec, egg cell; pn, polar nuclei; s, synergids. (B) The fertilized central cell has led to the early developing endosperm. The triploid endosperm nucleus (en) is dividing and nuclei are migrating from the micropylar region (mp) toward the chalazal end (cz). Making up most of space is a large central vacuole (cv), which is surrounded by a thin line of cytoplasm (cy). (C) and (D) Three distinct regions have formed: the micropylar endosperm (MCE), which surrounds the embryo (e), the central or peripheral endosperm (PEN), and the region of the chalazal endosperm (CZE), which contains the chalazal cyst (cz) that remains uncellularized throughout. (E) Globular embryo stage. (F) The endosperm has cellularized in the area around the embryo (MCE), radial microtubule system (rms) and initial cell walls (alveoli – alv) are forming in PEN, and multinucleate endosperm nodules (no) form in CZE. (G) Completely cellular endosperm (ce) and late heart embryo stage (e). The endosperm is now a source of energy for the developing embryo. (H) Surrounding the mature embryo (me) is the peripheral aleurone-like cell (alc) layer. A, adopted from Laux et al. (2004). B-H, adopted from Olsen (2004).

The endosperm remains in the syncytial phase until the embryo is at the early heart stage and cellularization begins (Mansfield and Briarty 1990). Cell walls are formed around the endosperm nuclei, starting with a radial microtubule system (rms) coming out from the surface of the endosperm nuclei followed by the alveolation process where the initial cell walls form (Olsen 2004). Endosperm cellularization marks the transition from resource sink to source for the embryo (Hehenberger et al. 2012). Early on it was shown that abnormal endosperm development leads to inviable hybrid seeds, suggesting that endosperm has a central role in hybridization barriers (Brink and Cooper 1947).

**The Arabidopsis genus**

*Arabidopsis thaliana* is a well-known and the preferred model organism in plant biology. In addition to access to a well-annotated genome, studies using *A. thaliana* have the additional advantage that a
vast number of mutants exist, which can be used to identify candidate genes involved in various processes. Taking advantage of the close relationship to *A. thaliana*, other species of the genus (e.g. *A. lyrata* and *A. arenosa*) have been established as model systems for investigating various evolutionary and ecological questions, as for instance hybridization barriers (Comai et al. 2000, Nasrallah et al. 2000, Beaulieu et al. 2009, Muir et al. 2015).

The split between *A. thaliana* and its sister lineage is suggested to have happened 6 million years ago (Hohmann et al. 2015). Within its sister lineage, the split between the *A. lyrata* and *A. arenosa* lineages is estimated to have happened 0.6 million years ago (Novikova et al. 2016) (Figure 2). The *A. lyrata* lineage is circumpolar and widely distributed with the two main subspecies growing in North America (*A. lyrata* ssp. *lyrata*) and Eurasia (*A. lyrata* ssp. *petrea*), respectively (Schmickl et al. 2010). Eastern parts of Central Europe are considered centres of diversity for the *A. arenosa* lineage, which also extends its distribution into northern and south-eastern Europe (Schmickl et al. 2012, Arnold et al. 2015). As opposed to the selfing annual *A. thaliana*, *A. lyrata* and *A. arenosa* are self-incompatible perennial or biennial plants.

![Figure 2 Relatedness between Arabidopsis lineages and suggested polyploid origins (colored lines). The dashed square represents the last 500 000 years within which period the polyploids have been suggested to have originated. Figure adopted from Novikova et al. (2018).](image)

Both diploid and tetraploid populations of *A. lyrata* and *A. arenosa* exist. The tetraploid *A. arenosa* is predicted to be an autotetraploid assumed to have originated from a single population about 15 000 to 19 000 generations ago (Arnold et al. 2015). The terms auto- and allotetraploid are used in the sense that polyploidization has occurred without and with the influence of a different species, respectively (Ramsey and Schemske 1998). The origin of tetraploid *A. lyrata* has not fully been established (Novikova et al. 2018). The genome size of tetraploid *A. lyrata* is not twice that of the diploid, which could indicate either an allotetraploid origin or that it has experienced genome reduction after the
polyploidization event (Dart et al. 2004). However, the tetraploid does not show disomic inheritance (Schneck and Koch 2011), which normally is seen in allopolyploids (Soltis et al. 1993). Tetraploid *A. lyrata* is dated to have originated around 160 000 years ago (Hohmann and Koch 2017). Tetraploid *A. arenosa* and *A. lyrata* are, thus, suggested to have originated in the last and second to last glaciation, respectively (Novikova et al. 2018), in line with other studies supporting high prevalence of polyploidy formation during the Pleistocene glaciations (Brochmann et al. 2004), e.g. *Draba* (Brochmann et al. 1992), *Parnassia* (Borgen and Hultgård 2003), *Cerastium* (Brysting et al. 2007), *Primula* (Casazza et al. 2012).

Unlike diploid *A. lyrata* and *A. arenosa*, where no introgression is seen (Hohmann and Koch 2017), gene flow due to introgression has been highly detected between tetraploid *A. lyrata* and *A. arenosa* (Jørgensen et al. 2011, Schneckl and Koch 2011, Hohmann and Koch 2017). Hybridization occurs naturally in hybrid zones also between diploid and tetraploid *A. arenosa* (Schneckl et al. 2012, Arnold et al. 2015, Kolář et al. 2015, Kolář et al. 2016).

**Presence and bypass of *Arabidopsis* post-zygotic hybridization barriers**

Hybridization and post-zygotic hybridization barriers have been investigated for many years using *Arabidopsis* species, both in interspecies and interploidy crosses. When crossing diploid and tetraploid *A. thaliana* ecotype Landsberg erecta (Ler) to tetraploid *A. arenosa*, fertilization occurred, however, mainly inviable seeds were produced, indicating a strong post-zygotic barrier (Comai et al. 2000). On the other hand, crossing tetraploid *A. thaliana* of the ecotype accession C24 to tetraploid *A. arenosa* produced as much as 80 % germinating seeds, suggesting that seed parent accession affects hybrid viability (Bushell et al. 2003). It has further been shown that the post-zygotic barrier, otherwise present in a cross between diploid *A. thaliana* and diploid *A. arenosa*, is lifted when using a tetraploid *A. thaliana* to cross with a diploid *A. arenosa*, suggesting that hybrid seed viability is also affected by parental genome dosage (Josefsson et al. 2006). In this system, fertilization only occurs if *A. thaliana* is the maternal cross partner, thus reciprocal crosses cannot be performed (Comai et al. 2000). This led to the hypothesis that by increasing the maternal genome dosage, the post-zygotic hybridization barrier will be lifted (Josefsson et al. 2006).

It was established that *A. thaliana* pollen was incompatible with *A. lyrata*, but using diploid *A. lyrata* ssp. *lyrata* as pollen donor in crosses with *A. thaliana* ecotype Col resulted in viable hybrids (Nasrallah et al. 2000). Similar (but with other accessions) crosses between *A. thaliana* ecotype Col-4 mother and *A. lyrata* ssp. *petraea* father, surprisingly produced hybrids where chromosome doubling was observed in as much as 25 % of the F1 offspring, restoring the fertility of the otherwise sterile hybrid (Beaulieu et al. 2009).
Recently, diploid *A. arenosa* and *A. lyrata* crosses were used to investigate hybridization in *Arabidopsis*, revealing that the fitness of the hybrid offspring depends on the direction of the cross, with 15% and 40% germinating seeds with *A. arenosa* and *A. lyrata* as the seed donor, respectively (Muir et al. 2015). Here, the availability of performing reciprocal crosses, allowing for all the controls needed to unravel effects of barrier components, makes *A. arenosa* and *A. lyrata* ideal for investigating the post-zygotic hybridization barriers.

**Triploid block and bridge**

Numerous studies have investigated the outcome of diploid to tetraploid crosses in both directions, and documented that triploid progeny are largely absent (Woodell and Valentine 1961, Ramsey and Schemske 1998). The rarity or absence of viable triploid progeny in diploid to tetraploid reciprocal crosses is known as the triploid block (Marks 1966). This triploid block, observed in interspecies interploidy crosses producing inviable seeds, has been shown to be linked to failure of endosperm development (Schatlowski and Köhler 2012). Marks (1966) suggested that in the case of a triploid block, unreduced gametes in the diploid are favored leading to more tetraploid offspring. In addition, Marks (1966) pointed out that given this scenario, introgression can occur between species without the link of triploid individuals. If first achieved, tetraploidization is proposed to be enough to bypass interploidy crossing barriers (Hanneman and Peloquin 1968, Johnston and Hanneman 1980). If, on the other hand, triploid offspring are formed, backcrossing of these to diploids or tetraploids can form euploid offspring when occasionally haploid or diploid gametes are formed (Ramsey and Schemske 1998). Ramsey and Schemske (1998) used existing data to create a model to estimate the formation of auto- and allotetraploids, either through a triploid bridge or by unreduced gametes. From the model, they showed that the triploid bridge has a great role in the formation of tetraploids, except in the case of allotetraploids with the ability to self.

**Parental specific effects**

Interploidy crosses often result in deviating seed phenotypes, depending on the direction of the cross. The observed phenotypes can be explained by the parental conflict theory, which states that the maternal genome produces factors limiting resources in order to counteract the paternal genome factors aiming to increase resources to single seeds (Haig and Westoby 1989). Hence, in parental excess crosses, there can be imbalance of maternal and paternal genomes, which is then observed as inhibited growth and smaller seeds in maternal excess crosses, and promoted growth and large seeds in paternal excess crosses (von Wangenheim and Peterson 2004). A possible mechanistic explanation for the concept of parental conflict is imprinted genes that are deregulated in the endosperm (Haig and Westoby 1991). Imprinted genes are suggested to be dosage sensitive as the gene copies of one of the parents are silenced, and the introduction of extra gene copies in parental excess crosses will disrupt
the normal gene dosage (Dilkes and Comai 2004). In support of this theory, the paternally expressed imprinted gene *ADMETOS* was found to have a role in the triploid block between diploid and tetraploid *A. thaliana*, with the effect of producing viable seeds in the *admetos* mutant (Kradolfer et al. 2013). Other paternally expressed genes have been shown to have the same effect (Wolff et al. 2015). Outside the *Arabidopsis* system, imprinting and endosperm development have also been indicated as important in interspecies crosses of *Mimulus* (Garner et al. 2016), *Capsella* (Rebernig et al. 2015), and tomato (Florez-Rueda et al. 2016). In addition, earlier studies have shown parental specific effects on interspecies and interploidy crosses in e.g. wheat (Gill and Waines 1978) and maize (Lin 1984).

The paternally expressed imprinted gene *PHERES1* (*PHE1/AGL37*) (Köhler et al. 2003), along with the maternally expressed imprinted gene *AGAMOUS-LIKE* (*AGL*) 36 (Shirzadi et al. 2011), have been shown to be affected in interspecies crosses of *A. thaliana* and *A. arenosa* (Josefsson et al. 2006, Walia et al. 2009). In particular, the MADS box type I gamma genes have many genes expected to have parental-biased gene expression. This was discovered in a screening using the *cdka;1* mutant (Shirzadi et al. 2011), which produces pollen with only one functioning sperm cell, and where the endosperm can develop without fertilization initiated by the fertilization of the egg cell (Nowack et al. 2006).

**Silencing mechanisms and parent of origin gene expression**

In order to have a normally developing endosperm, it is postulated that correct regulation of imprinted genes is important (Gill and Waines 1978, Haig and Westoby 1991), which is highly relevant in interspecies crosses since imprinted genes have been shown to differ between species, even between related species (Klosinska et al. 2016). The imprinting state can come about through different mechanisms leading to silencing of one of the parental gene copies or by removing the repression marks on one but not the other. The silencing mechanism through DNA METHYLTRANSFERASE1 (MET1) (Finnegan and Dennis 1993) is a maintenance process of CG methylation involved in imprinting of genes. Both parental copies of many imprinted genes are initially methylated by MET1 (Jullien et al. 2006, Satyaki and Gehring 2017). In the central cell in the female gametophyte, the activity of a DNA glycosylase DEMETER removes methylated cytosins on the maternal copies, activating the gene (Choi et al. 2002). In pollen, DEMETER is only found to be active in the vegetative cell and does not lift methylation marks in the male germline (Schoft et al. 2011). The result is that the maternal copy of the gene is active while the paternal copy delivered by the sperm cells remains silenced. The FERTILIZATION INDEPENDENT SEED (FIS) Polycomb Repressive Complex 2 (PRC2) gene, *MEDEA* (*MEA*) is also a subject to MET1 methylation which is removed by DEMETER, and thus activating it (Kinoshita et al. 1999, Choi et al. 2002, Gehring et al. 2006). MEA is a SET domain polycomb protein shown to function in the suppression of endosperm development through gene silencing (Kiyosue et al. 1999). MEA activity suppresses through methylating histone 3 lysine 27 (H3K27) (Köhler and Hennig 2010), and is important in regulating
imprinted genes (Luo et al. 2000). Shortly after fertilization MEA takes part in regulates MADS box type I genes, downregulating and later turning off these genes (Shirzadi et al. 2011, Zhang et al. 2018).

A different silencing mechanism is through Polymerase (PolIV)-mediated short interfering (si) RNAs that can result in de novo DNA methylation through the canonical RNA directed DNA methylation (RdDM) pathway (Wierzbicki et al. 2012). The pathway results in methylation of CHH sequences where H represents the nucleotide A, T or C. Double stranded RNA molecules based on PolIV transcripts are diced into 24 nucleotide (nt) fragments, which are loaded onto an ARGONAUTE (AGO) RNA-Induced Silencing Complex (RISC) protein complex. The siRNA-AGO complex is directed to PolV transcripts where DOMAIN REARRENGED METHYLASE 2 (DRM2) de novo methylates nearby CHH sequences, resulting in silencing through heterochromatin formation (Wierzbicki et al. 2012). The pathway has been shown to have a role in imprinting (Vu et al. 2013), and hence is of interest when investigating imprinted genes. MADS box type I genes have been shown to be deregulated in interploidy A. thaliana crosses and in A. thaliana mutants of proteins important for the PolIV RdDM (Lu et al. 2012).

MADS box type I transcription factors

The MADS acronym comes from the founding proteins found in four different species: MINICHROMOSOME MAINTENANCE1 (Saccharomyces cervisiae), AGAMOUS (Arabidopsis thaliana), DEFICIENS (Antirrhinum majus), and SERUM RESPONSE FACTOR (Homo sapiens). The gene family is defined by the MADS-box domain, coding for an approximately ~58 amino acids long sequence having DNA and protein interaction properties. The MADS box genes are divided into two main groups, namely Type I or M-type, which has the subgroups alpha, beta and gamma, and MIKC also known as Type II. In addition it is a group called MIKC* or delta, which has protein domains in common with MIKC (Alvarez-Buylla et al. 2000, De Bodt et al. 2003, Parenicova et al. 2003). Parenicova et al. (2003) reported 107 MADS box genes in A. thaliana, which have been used as the starting point for many following studies. As noted by Bemer et al. (2010), Parenicova et al. (2003) did not include AGL101 for unknown reasons. The complete list of all A. thaliana MADS box genes, which including AGL101 makes 108 in total, was used in a protein-protein interaction study performed by de Folter et al. (2005). They showed that many of the proteins have several interaction partners within the MADS box gene group, indicating that these proteins can function as dimers and tetramers. For instance, AGL62 is shown to interact with AGL28 and AGL36. The MADS box type I genes have largely not been assigned a function but are shown to be expressed during gametophyte and seed development (Masiero et al. 2011). One of the described genes is AGL62, which has a role in endosperm cellularization, shown to lead to early cellularization if mutated (Kang et al. 2008). In the incompatible hybrid cross between diploid A. thaliana and diploid A. arenosa, the gene expression of the MADS box type I genes AGL35, AGL36, AGL40, AGL62, AGL90, PHERES1 and PHERES2 are
shown to be upregulated (Walia et al. 2009). In addition, Walia et al. (2009) showed the necessity of maintaining expression of PRC2 genes in order to produce viable hybrid seeds.

**Endosperm Balance Number (EBN) theory**

It is suggested that a reason for hybrid failure in interploidy crosses is due to the imbalance of maternal and paternal contribution in the endosperm, formulated as the Endosperm Balance Number (EBN) theory (Johnston et al. 1980). The EBN theory was followed up with supporting results in Johnston and Hanneman (1980). The theory states that if viable seeds are produced, the two crossing partners have the same EBN, regardless of actual ploidy and species. The 2:1 ratio of maternal to paternal contribution in the endosperm is the foundation of the theory, ignoring the maternal seed coat tissue and its ploidy. Johnston et al. (1980) proposed that certain endosperm factors decide the EBN and the effective ploidy of a species, and that these factors must be in a 2:1 ratio. Given that EBN is caused by a few specific genes, an evolutionary implication could be that a mutation in one of them could result in rapid formation of reproductive isolation and lead to the evolution of a new sympatric species. EBN can potentially be used to predict the outcome of interspecies crosses from the known crossing success of related species (Johnston et al. 1980).
Aim of Study

As a step towards understanding how new species form and how existing barriers prevent different species from producing viable offspring when crossed, this study aims to investigate the post-zygotic hybridization barrier between *Arabidopsis* species, with a specific focus on the role of MADS box type I transcription factors.

We use different ecotype accessions of diploid *A. thaliana*, diploid and tetraploid *A. arenosa*, and diploid and tetraploid *A. lyrata* to perform crosses, in controlled growth conditions. Combining these species in interspecies and interploidy crosses, we aim to answer questions related to the phenotype and genetic basis of the barrier. How is the seed and endosperm phenotype related to the post-zygotic hybridization barrier? Does an increase of the maternal genome dosage bypass the post-zygotic hybridization barrier? What is the underlying genetic mechanism and which MADS box type I genes are involved in the barrier?

In more detail, we aim to investigate imprinted genes in connection with the hybridization barrier. The MADS box type I gene family is expected to have many imprinted genes and identifying these genes and the imprinting regulatory mechanism will aid in understanding the role of imprinting and imprinted genes.

Choosing the *Arabidopsis* species as our plant system enables us to make use of the vast amount of *A. thaliana* resources and data. The great advantage of including *A. lyrata* and *A. arenosa* is that they can be reciprocally crossed, unlike *A. thaliana* that can only be used as mother in crosses with its sister species. Using perennial and heterozygous plants from natural populations of *A. lyrata* and *A. arenosa* adds extra challenges, but by doing so we hope to aid in bridging knowledge from experimental *A. thaliana* studies to studies on natural populations.
Results and Discussion

The barrier between diploid *A. lyrata* and diploid *A. arenosa* is nearly complete, with only some germinating seeds resulting from crosses between the two (Muir et al. 2015). Previously it has been shown that the species barrier between *A. thaliana* and *A. arenosa* could be bypassed in maternal excess crosses, suggesting that increasing the maternal genome dosage has a role in bypassing the barrier (Josefsson et al. 2006). This would be in line with the majority of studies (ten out of eleven) surveyed by Ramsey and Schemske (1998), which show that viable triploid offspring result from maternal excess crosses. However, based on a new survey of recent studies, Vallejo-Marín et al. (2016) showed that paternal excess crosses also have a significant role in producing interploid interspecies viable offspring. Moreover, the role of the endosperm in post-zygotic hybridization barriers has been highlighted in a growing number of recent studies in several species, e.g. *Mimulus* (Oneal et al. 2016, Kinser et al. 2018), *Capsella* (Rebernig et al. 2015), and tomato (Roth et al. 2018a).

With the aim of describing the post-zygotic hybridization barrier by identifying the phenotype on a morphological level, and finding genes involved in the *Arabidopsis* barrier, we crossed diploids and tetraploids to each other, including interploid intraspecies crosses. Many MADS box type I transcription factors are expressed during early seed development (Bemer et al. 2010), and are shown to be affected in *Arabidopsis* hybrid crosses (Walia et al. 2009), making this gene family our genetic focus.

To summarize the main findings in Paper I, II and III, firstly I will give an overview of the *A. arenosa* to *A. lyrata* crossing success, studied as seed phenotype and seed germination frequency, and how introgression influence the results and affects the strength of the barrier. Secondly, I will summarize our investigations on seed development in *A. thaliana*, *A. lyrata*, *A. arenosa* and their hybrids, showing the importance of endosperm cellularization as an explanation for the barrier. Thirdly, I will elaborated on the gene expression of MADS box type I genes in *A. thaliana*, *A. lyrata*, *A. arenosa* and a selection of their hybrids, as these genes are linked to endosperm development. Finally, I will share some thoughts on how different populations of *A. lyrata* and *A. arenosa* can be related to the Endosperm Balance Number (EBN) theory, and how introgression complicates this relationship.

*A. arenosa* and *A. lyrata* compatibility investigated by different cross combinations and populations

In order to describe the post-zygotic hybridization barrier between *A. lyrata* and *A. arenosa*, we reciprocally crossed diploids and tetraploids of both species in all combinations, as denoted below:

\[
4x A. arenosa \leftrightarrow 2x A. lyrata \leftrightarrow 2x A. arenosa \leftrightarrow 4x A. lyrata \\
2x A. lyrata \leftrightarrow 4x A. lyrata \leftrightarrow 4x A. arenosa \leftrightarrow 2x A. arenosa
\]
The plants used for our crossings were grown from seeds collected from the different populations localized in Austria and Slovakia (Jørgensen et al. 2011) (Roswitha Schmickl, personal communication), as shown in Figure 3.

![Figure 3 Localities of A. arenosa and A. lyrata populations used in Paper I and III. A. arenosa and A. lyrata populations in red and blue, respectively. Introgressed tetraploid populations (4LWAC and 4AKER) are written in bold. 4A and 2A: tetraploid and diploid A. arenosa, respectively. 4L and 2L: tetraploid and diploid A. lyrata, respectively. Population abbreviations used in the papers indicated by roman numerals: KRO: Kronsegg (III), KER: Kernhof (I, III), NT: Nizke Tatry (2A I and III, 4A III), WAC: Wachau (I, III), PER: Pernitz (I, III), LIC: Wienerwald Castle Lichtenstein (III).](image)

Inspection of seed phenotypes (Figure 4) and germination assays (Paper I Figure 1 and 3, Paper III Figure 2) revealed that the species barrier was intact in some crossing combinations and bypassed in others. We found that very small seeds and dark, collapsed seeds can be associated with germination failure (Paper I Figure 1 and 3). The diploid interspecies crosses were produced with two sets of diploid populations, which both produced hybrid seeds with low germination frequency (Paper I Figure 1, Paper III Figure 2), showing a strong species barrier. In Paper I and Paper III, the diploid populations used in crosses with tetraploid plants were A. arenosa from Nizke Tatry (2A-NT) and A. lyrata from Pernitz (2L-PER) (Figure 3). In Paper I, we focus on the diploid and tetraploid interspecies crosses in addition to the interploidy interspecies crosses (Figure 4 A-C) in order to investigate the effect of increasing parental genome dosage on the hybridization barrier. In Paper III we complete the crossing scheme with interploidy intraspecies crosses (Figure 4 D). In order to investigate the effect of introgression on hybridization barriers, additional tetraploid populations were included (Paper III Figure 2).
The diploid barrier can be bypassed by genome increase in a species dependent manner

The strong species barrier in diploid interspecies crosses resulted in collapsed or small seeds (Figure 4 A) and approximately only 10% seeds germinating (Paper I Figure 1, Paper III Figure 2), which is less than what has been previously published for similar crosses (Muir et al. 2015). A possible explanation for variation in germination frequency could be due to natural variation between populations (Bushell et al. 2003, Burkart-Waco et al. 2012).

![Figure 4 Seed phenotypes.](image)

When increasing the genome dosage of *A. lyrata* using a tetraploid, the maternal excess interspecies cross resulted in a complete bypass of the barrier. Moreover, so did also the reciprocal paternal excess cross (Figure 4 B), indicating that tetraploidization of *A. lyrata* contributes to bypass the barrier in a manner not dependent on the parental contribution. However, the bypass effect was only seen in the combination of tetraploid *A. lyrata* and diploid *A. arenosa*. The cross between diploid *A. lyrata* and tetraploid *A. arenosa* appeared to be incompatible. The observed seed phenotypes were very small seeds in the maternal excess cross and larger, collapsed seeds in the paternal excess cross, suggesting that the parental excess crosses have an imbalance of growth affecting factors. This is in line with the
parental-conflict theory (Haig and Westoby 1989), which states that the maternal genome produces growth limiting factors to ensure allocation of resources to all offspring, whereas the paternal genome produces growth stimulating factors to better secure the fate of single offspring. Our results falls into line with the published outcomes of several other studies showing similar seed phenotypes from hybrid crosses, as reviewed by von Wangenheim and Peterson (2004).

In addition, we observed that in the *A. arenosa* interploidy crosses, 10-50% of the seeds in the paternal excess cross germinated. In contrast, the *A. lyrata* interploidy crosses produced only a single viable hybrid seed, and a clear non-reciprocal effect is seen as expected in interploidy crosses (Figure 4 D).

From theory (e.g. the Endosperm Balance Number theory), we would expect to find the same barrier between the two species at tetraploid level as exists at diploid level, but to our surprise, the tetraploid interspecies crosses mainly produced viable seeds with no sign of a species barrier (Figure 4 C). In our initial interspecies interploidy crosses, we used plants from a tetraploid *A. arenosa* population from Kernhof (4A-KER) and a tetraploid *A. lyrata* population from Wachau (4L-WAC) (Figure 4 B-D) (Figure 3), which to some degree are introgressed by the other species (Schmickl and Koch 2011).

This opened for the possibility that the bypass of the interspecies interploidy barrier could be due to introgressed genetic material. To test if genome doubling in itself has an effect on the bypass of the hybridization barrier, a colchicine-induced tetraploid was made from diploid *A. lyrata* and crossed to diploid *A. arenosa*. The resulting hybrid seeds germinated well with a frequency of 50-70% (*Paper I Figure S5*), suggesting that the doubling of the genome circumvents the post-zygotic barrier.

In order to give further support to the hypothesis that the observed bypass of the barrier is not due to introgression, we used tetraploid *A. lyrata* plants from an estimated non-introgressed population collected in Wienerwald (Castle Lichtenstein) in Austria (4L-LIC, Roswitha Schmickl, personal communication) (Figure 3). The result of the crosses between diploid *A. arenosa* and tetraploid *A. lyrata* 4L-LIC produced mainly viable seeds (*Paper III Figure 2*), in line with the results from the crosses with the introgressed 4L-WAC (*Paper I, Figure 3*), and supporting the results from crosses using the synthetic tetraploid. The paternal excess crosses produced on average 70% germinating seeds, and the maternal excess crosses produced on average 50% germinating seeds, which is a significant reduction compared to the parents (*Paper III Figure 2*). Based on these results, introgression does not appear to have a role in the bypass of the interspecies interploidy hybridization barrier, suggesting that mere tetraploidization of *A. lyrata* is the cause for production of viable hybrid offspring. Whether the rescue of the cross is due to the doubling of the whole genome or just the adjusted gene expression of a few genes, or other genetic elements, remains to be investigated.

**The hybridization barrier is stronger between non-introgressed tetraploids**

Based on the observed success of the tetraploid interspecies crosses (*Paper I*) and the reported ongoing introgression between the tetraploid species in hybrid zones, we hypothesized that the barrier
would be stronger between plants of less introgressed tetraploid *A. lyrata* and *A. arenosa* populations. To investigate this, we performed reciprocal tetraploid interspecies crosses using estimated non-introgressed populations (4A-KRO, 4A-NT and 4L-LIC; the introgression level for 4A-KRO is estimated by Schmickl and Koch (2011) and for 4A-NT by Monnahan et al. (2019)) (Figure 3). In contrast to the seed phenotypes of the crosses between 4A-KER and 4L-WAC (Figure 4 C), which show only slight seed size differences (*Paper I Figure 3*), a non-reciprocal effect is seen in crosses including the non-introgressed 4L-LIC and accompanied by a significant reduction in germination frequency (*Paper III Figure 2*). The low germination frequencies when 4L-LIC acts as mother are linked to collapsed seeds (*Paper III Figure 2*). The reciprocal crosses when 4L-LIC acts as father result in seeds varying from small seeds to more viable-looking ones. Fewest germinating seeds, and an almost complete barrier in both directions, are obtained from crosses when 4L-LIC are crossed to 4A-NT (both non-introgressed). This cross produces only 2% and 10% germinating seeds when 4L-LIC is mother and father, respectively. A stronger asymmetrical barrier is found in crosses with the introgressed 4A-KER, from which 18% and 50% seeds germinate with 4L-LIC as mother and father, respectively, but with a large variation between the biological replicates. Unexpectedly, an asymmetrical barrier is also seen in crosses with 4A-KRO (non-introgressed), with 2% and 46% (significantly different *p*<0.05) germinating seeds when 4L-LIC is mother and father, respectively. This suggests that not only introgression, but also natural variation between populations have an effect on the barrier, as shown previously in interspecies crosses using different *A. thaliana* ecotypes (Burkart-Waco et al. 2012).

**The post-zygotic hybridization barrier between *A. lyrata* and *A. arenosa* is endosperm-based**

In order to investigate the phenotype of the post-zygotic hybridization barrier in more detail, we analysed seed development (*Paper I and II*). Embryo analysis showed that diploid *A. lyrata* develops faster compared to diploid *A. lyrata*, and that the hybrid crosses between them follow the maternal crossing partner (*Paper I Figure S2*). Likewise, the embryo develops faster in tetraploid *A. arenosa* than in tetraploid *A. lyrata*, and the hybrid crosses between them follow the maternal crossing partner (*Paper I Figure S4*). Diploid hybrid seeds are mostly inviable and the tetraploid hybrid seeds are mainly viable and lead to mature plants. The interspecies interploidy crosses between diploid *A. lyrata* and tetraploid *A. arenosa* display slow embryo development compared to all the other crosses, while the crosses between diploid *A. arenosa* and tetraploid *A. lyrata* have embryos at developmental stages similar to the other compatible crosses (*Paper I Figure S4*).

Endosperm has for a long time been acknowledged to have an important role in the success of interspecies crosses and hybrid seed failure (Brink and Cooper 1947), and two of the main features in nuclear-mode endosperm development are endosperm proliferation and cellularization. To investigate the role of endosperm cellularization in the hybridization barrier between *A. lyrata* and *A. arenosa*, we
tested if the timing of cellularization was affected in the hybrid seeds by analysing seeds at an early time-point before cellularization (9 DAP) and at a late time-point (15 DAP) after cellularization is expected to have occurred (Paper I Figure 2 and 4). In the diploid control crosses, we observed an uncellularized endosperm at the early time-point and completion of cellularization at the late time-point (Paper I Figure 2). Interestingly, already at the early time-point, the endosperm is cellularized in the diploid hybrid cross with *A. arenosa* mother, showing a correlation between the small seed size (Paper I Figure 1) and early endosperm cellularization observed in these crosses. In contrast, the reciprocal hybrid cross still has uncellularized endosperm at the late time-point (Paper I Figure 2), suggesting that delayed endosperm cellularization leads to embryo arrest and seed abortion. The tetraploid *A. arenosa* and *A. lyrata* control crosses display, like the diploid controls, uncellularized endosperm at the early time-point and cellularization at the late time-point (Figure 5 A-B and C-D, respectively). The successful hybrid crosses between tetraploid *A. lyrata* and diploid *A. arenosa* exhibit this same pattern when tetraploid *A. lyrata* acts as mother (Figure 5 E-F) as well as father (Figure 5 G-H), showing normal timing of endosperm cellularization in the rescued cross. The incompatible cross between tetraploid *A. arenosa* mother and diploid *A. lyrata* father shows early endosperm cellularization and a globular staged embryo (Figure 5 I), indicating that the onset of cellularization is not linked to embryo development. The embryo is still at globular stage at the late time-point (Figure 5 J), however, the uncoupling of embryo and endosperm development supports the role of the endosperm in the post-zygotic hybridization barrier. The reciprocal cross with tetraploid *A. arenosa* father shows no cellularization at either early or late time-point (Figure 5 K-L). Extending the time-frame for investigation, we show that cellularization is still absent after the embryo has reached early heart stage at 17 DAP (Figure 5 M-N), and probably never cellularizes.

To investigate the role of endosperm proliferation in the establishment of the hybridization barrier, endosperm nuclei were scored in 5, 7 and 9 DAP seeds, and the proliferation rate from 5 to 7 DAP and 7 to 9 DAP was calculated (Paper I Figure 2 and 5). In line with the delayed embryo development in diploid *A. lyrata*, there were fewer endosperm nuclei at the first time-point (5 DAP) than in diploid *A. arenosa* (Paper I Figure 2). However, the proliferation rate was similar, indicating that differences in endosperm proliferation happen very early in seed development. The greatest defect is seen for the crosses known to produce very small seeds (2A × 2L, 4A × 2L), where endosperm proliferation arrests at 5 DAP, and the seeds are too shriveled at 9 DAP for the nuclei to be counted (Paper I Figure 2 and 5). Differences in endosperm nuclei counts in the other crosses are also shown to be linked to the endosperm nuclei number at the earliest time-point. However, no clear effect on proliferation rate can be seen for the incompatible crosses with diploid *A. lyrata* mother crossed to diploid or tetraploid *A. arenosa* (2L × 2A, 2L × 4A), suggesting that endosperm proliferation is not coupled to the timing of the endosperm cellularization.
Figure 5 Endosperm cellularization after 9 DAP and before 15 DAP in compatible and incompatible crosses. (A) and (B) 4A × 4A uncellularized endosperm at 9 DAP, which has cellularized at 15 DAP, respectively. (C) and (D) 4L × 4L uncellularized endosperm at 9 DAP, which has cellularized at 15 DAP, respectively. (E) and (F) 4L × 2A uncellularized endosperm at 9 DAP, which has cellularized at 15 DAP, respectively. (G) and (H) 2A × 4L uncellularized endosperm at 9 DAP, which has cellularized at 15 DAP, respectively. (I) and (J) 4A × 2L cellularized endosperm at 9 and 15 DAP, respectively. (K) and (L) 2L × 4A uncellularized endosperm at 9 DAP, which is still uncellularized at 15 DAP, respectively. (M) and (N) 2L × 4A uncellularized endosperm at 17 DAP with focus on embryo and endosperm, respectively. Scale bars 50μm.

With this, we are linking the post-zygotic hybridization barrier between A. arenosa and A. lyrata to the timing of endosperm cellularization, when the endosperm transition from sink to source for the embryo takes place. On the other hand, we found no indications of a significant effect of embryo development or endosperm proliferation. It has previously been shown a link between the alteration on timing of endosperm cellularization and post-zygotic hybridization barriers in intraspecies crosses in A. thaliana and rice, but in contrast to our results in those cases also a change in proliferation rate was found (Scott et al. 1998, Sekine et al. 2013). Our findings are in line with results from interspecies crosses between several rice species (Ishikawa et al. 2011), strengthening the focus on cellularization timing, rather than endosperm proliferation, in interspecies crosses.
In order to test if the endosperm-based post-zygotic hybridization barrier can be found in other interspecies Arabidopsis crosses, we crossed diploid A. thaliana to diploid A. arenosa. In diploid A. lyrata and A. arenosa hybrids, the effect on endosperm cellularization depends on the direction of the cross. Using A. thaliana restricts the cross to one direction as A. thaliana pollen is incompatible with A. arenosa stigmas (Comai et al. 2000), hence, we only use A. thaliana as mother and A. arenosa as father. The diploid A. arenosa population Strecno-1 (SN1) has been shown to largely be incompatible with A. thaliana (Walia et al. 2009), and in order to check if this is connected to endosperm failure we analyzed the endosperm development with focus on cellularization. The A. thaliana self cross at 7 DAP shows very little variation across the seeds analyzed, with mainly late heart stage embryos and cellularized endosperm (Figure 6). The hybrid cross between A. thaliana and the diploid A. arenosa population from Nizke Tatry (2A-NT/MJ09-4) is delayed in development at 7 DAP compared to A. thaliana, but exhibits mainly torpedo embryos and cellularized endosperm at 10 DAP, indicating that as many as 70% of the seeds could be viable (Figure 6). In contrast, the hybrid cross between A. thaliana and the A. arenosa Strecno-1 (SN1) population at 10 DAP mainly has seeds with uncellularized endosperm although the embryos are at heart stage. With only 30% of the seeds cellularized in crosses with the SN1 population, compared to 70% using the 2A-NT population, we suggest that genetic variation between A. arenosa populations influences the success rate of hybridization as previously demonstrated using different accessions of A. thaliana (Burkart-Waco et al. 2012).

Figure 6 Variation in embryo and endosperm developmental stages among hybrid seeds. Quantification of embryo and endosperm stages in selfed A. thaliana and A. thaliana to A. arenosa hybrid seeds (Paper II, Figure 8I). TxT: A. thaliana seeds (n=31), TxA: A. thaliana mother crossed to diploid A. arenosa (2A-NT/MJ09-4) father at 7 DAP (n=34) and 10 DAP (n=81), TxS: A. thaliana mother crossed to diploid A. arenosa Strecno-1 (SN1) father (n=98). DAP: Days after pollination.
The barrier between *A. thaliana* and *A. arenosa* is population and temperature dependent

In contrast to our findings, previous reports on *A. arenosa* SN1 hybrids with *A. thaliana* reported significantly lower germination frequencies (Walia et al. 2009). We reasoned that the discrepancy may be due to a difference in temperature between the experimental setups as our experiments were conducted at 18 °C conditions contrasted to 22 °C in the study by Walia et al. (2009). In order to investigate the effect of temperature, we performed *A. arenosa* hybrid crosses with *A. thaliana* at 18 °C and 22 °C conditions. At 20 DAP, siliques from *A. arenosa* populations SN1 and 2A-NT/MJ09-4 crossed to *A. thaliana* at both temperatures were harvested and seeds analysed (Figure 7), finding that the 4 °C increase in temperature is associated with higher occurrence of shrivelled and dead looking seeds (Figure 7). The germination frequency of the seeds matured at 22 °C is below 10 % and close to 0 % for the hybrids with *A. arenosa* father from the 2A-NT/MJ09-4 and SN1 population, respectively (Paper II Figure 6). Hence, a stronger hybridization barrier is observed when increasing the temperature to 22 °C, and the strength of the barrier is population dependent with hybrids of *A. arenosa* SN1 showing the strongest response to elevated temperature.

Figure 7 The diploid *A. thaliana* *A. arenosa* hybrid barrier is temperature dependent. (A) Siliques with hybrid seeds of *A. thaliana* crossed to plants from two different diploid *A. arenosa* populations (MJ09-4/2A-NT and SN1) matured on 18 °C and 22 °C and harvested 20 days after pollination. (B) Quantification of live seeds from crosses in (A). Each bar represents the average of three replicates, n = [174, 163, 175, 162] respectively. Error bars indicate standard deviation.

It is expected that lowering of temperature slows down processes in the developing seeds. Stress responses could also lead to cell cycle arrest (Komaki and Schnittger 2017), which possibly could be induced by higher temperature. At low temperature, otherwise deleterious outcomes could be bypassed by allowing more time to make it through the cell cycle, or a stress response induced at 22 °C could affect cell division leading to seed development failure. The evolutionary benefits of allowing hybridization under some conditions but not others could be to allow gene flow and use of resources for hybridization when conditions are optimal.

High speciation rates, especially involving polyploidization, have been associated with periods of glaciation (Brochmann et al. 2004, Novikova et al. 2018), suggesting that temperature may play a role
in the formation of new species. Our findings suggest a scenario where both intrinsic variation in genetic pathways in both parents and external abiotic factors such as temperature may act in concert to regulate the strength of post-zygotic species barriers.

MADS box type I genes and their role in the endosperm-based hybridization barrier

The presence of an endosperm based hybridization barrier (Paper I and II), and the finding that the strength of this barrier can be modulated by genetic differences between populations (Paper II and III), call for investigation of the genetic basis of the barrier. A gene family known to be specifically expressed in the endosperm (Parenicova et al. 2003, Bemer et al. 2010), and shown to be deregulated in incompatible hybrid seeds (Walia et al. 2009), is the MADS box type I TFs. This class of TFs have also been demonstrated to be involved in endosperm cellularization (Kang et al. 2008), the major phenotypic cause for the here described post-zygotic barrier. Moreover, genomic imprinting in the endosperm has been suggested to play a role in interspecies barriers (Haig and Westoby 1991), and a large fraction of the MADS box type I gene family has been shown to be imprinted (Paper II) (Köhler et al. 2003, Shirzadi et al. 2011, Zhang et al. 2018). To continue the elucidation of the role of this gene family in endosperm-based hybridization barriers in the Arabidopsis genus, we performed an RNAseq based expression study from 1 to 12 DAP in A. thaliana (Paper II). In addition, a two-time point expression analysis focused on MADS box type I homologs in A. lyrata and A. arenosa and a selection of theirs hybrids (Paper III).

Regulation of MADS box type I genes

In A. thaliana, we see that most of the type I alpha and gamma genes share an expression pattern with a peak around globular stage and cellularization (3-6 DAP), followed by a decline in expression (Paper II, Figure 1). The MADS box type I genes include a group of mainly alpha and gamma genes that are regulated by FIS-PRC2. In the absence of PRC2 MEDEA (MEA), the MADS box gene expression profiles are affected in the direction of shifting the peak of expression to a later developmental stage or shown as absence of the expected decline (Paper II Figure 4). We find that the type I beta group is least affected by PRC2 MEA with only a few representatives that alter gene expression pattern in the mea mutant (Paper II Figure 4).

Endosperm development is subject to genome dosage effects as the triploid tissue has two maternal and one paternal genome copy, making parental-specific genes of particular interest. In previous studies, imprinting of AGL36 was shown to be regulated in a met1 mutant (met1-4), where paternal AGL36 was derepressed, indicating that MET1 has a role in regulating AGL36 (Shirzadi et al. 2011). In the case of MET1 mediated silencing, it has been found that in the central cell of the female gametophyte the DNA glycosylase DEMETER (DME) demethylates the maternal copy of several genes, leaving the paternal copy silenced, resulting in parental specific expression (Choi et al. 2002). Using a different mutant allele (met1-7), we do not see the same effect, indicating that the two
different MET1 alleles have different effects (Paper II Figure 3). This discrepancy could be due to a different degree of knock-out in the mutant alleles. However, the generation history for met1-7 is known and the line has been kept heterozygous in order to avoid accumulation of hypomethylation. Similar records are not known for the met1-4 mutant line and additional effects due to accumulation of hypomethylation over generations cannot be ruled out. Nevertheless, for AGL28, which we show is imprinted and maternally expressed, we do see an activation of the paternal copy of AGL28 in crosses with the heterozygous met1-7 mutant, suggesting that the paternal allele of AGL28 is derepressed. AGL28 has not previously been shown to be imprinted in A. thaliana, however ecotype specific imprinting has been suggested (Wolff et al. 2011, Pignatta et al. 2014, Gehring and Satyaki 2017). We further demonstrate imprinting and maternal expression of AGL90 and biallelic expression of AGL35 (Paper II Figure 3), in accordance with a recent study (Zhang et al. 2018).

In summary, we demonstrate that MADS box type I TFs are imprinted, putatively maternally activated by DME mediated demethylation (Shirzadi et al. 2011). At later endosperm developmental stages, histone modification by MEA represses and leads to a decline in MADS box type I expression (Paper II Figure 4 and S4) (Shirzadi et al. 2011). An additional layer of regulation of imprinted genes has been suggested to be through RNA directed DNA methylation (RdDM) where Pol IV-siRNAs direct de novo CHH methylation to a nearby area of a PolV transcript. AGL36 has been shown to be upregulated in mutants affecting the PolIV RdDM pathway (Lu et al. 2012). A possibility could be that siRNAs produced near the maternal gene copy of AGL36 could be directed to a PolV transcript near the paternal gene copy, contributing to gene silencing. To test if the upregulation was due to reactivation of the paternal gene copy, we analysed the imprinting status in a PolIV mutant, affecting the RdDM pathway in the same way as Lu et al. (2012) (Paper II Figure 3). However, only maternal expression and no change in the parental contribution to the gene expression was found. Further investigations brought us to repeat the previously published experiment that had shown upregulation (Lu et al. 2012), but we could not repeat the findings (Paper II Figure 3).

In order to investigate the role of imprinting in post-zygotic hybridization barriers, we tested the imprinting status of AGL36 in A. arenosa and in interspecies hybrids. We found that AGL36 is only maternally expressed in A. arenosa, suggesting that the regulation mechanism is conserved in at least one sister species of A. thaliana (Paper II Figure 5). The conservation of the imprinting status of AGL36 could imply that other A. thaliana imprinted genes could be similarly imprinted in A. lyrata and A. arenosa. In addition, AGL36 maintains imprinting and sole maternal expression also in hybrids of diploid A. lyrata and A. arenosa (Paper II Figure 5), suggesting that the imprinting of this gene is not affected in the hybrids. It also indicates imprinting in A. lyrata, although this remains to be tested. Whether this is the case for the interploidy interspecies crosses we do not know, as the parental origin of the overall gene expression in A. arenosa, A. lyrata, and their hybrids has not been determined. It remains to investigate if changes in gene expression in interploidy interspecies hybrids are due to
activation of normally silenced gene copies. Interestingly, the imprinting status was not maintained in *A. thaliana* crossed to *A. arenosa*, where both the maternal *A. thaliana AGL36* allele and the paternal *A. arenosa AGL36* allele are demonstrated to be expressed at the same time around cellularization in the seed (*Paper II Figure 5*). This suggests that silencing of the paternal *A. arenosa AGL36* allele breaks down in the hybrid endosperm leading to derepression.

**MADS box type I genes are regulated in incompatible Arabidopsis hybrids**

In order to elucidate the role of MADS box type I gene expression in post-zygotic hybridization barriers we performed RNAseq on diploid and tetraploid *A. lyrata* and *A. arenosa* as well as diploid and interspecies interploidy hybrid seeds. To focus on the cellularization event two time-points, before and after cellularization were chosen (9 and 15 DAP, compare Paper I). The elaborated MADS box type I expression in *A. thaliana* from 1-12 DAP (Paper II) gives the opportunity to consider the gene expression in a broader sense (*Paper II Figure 1*). The DAP stages in Paper II are the same as described by Shirzadi et al. (2011). Seed development is slowed down in *A. lyrata* and *A. arenosa* compared to *A. thaliana*, and comparing embryo stages, 9 DAP globular stage in *A. lyrata* and *A. arenosa* corresponds to 3-4 DAP globular stage in *A. thaliana*. Likewise, 15 DAP late heart stage to early torpedo in *A. lyrata* and *A. arenosa* corresponds to between 6 - 9 DAP in *A. thaliana* (Shirzadi et al. 2011). The MADS box type I genes found to be expressed in both diploid *A. lyrata* and *A. arenosa* (*Paper III Figure S3*), can be compared to the expression curves in *A. thaliana* (*Paper II Figure 1*) (Figure 8). The *A. thaliana* genes show expression peaks at either 3 DAP or 6 DAP, corresponding to before and after onset of endosperm cellularization (*Paper II Figure 1*).

The obtained transcripts from the RNAseq were matched to the known *A. thaliana* MADS box type I and II genes with a maximum-likelihood phylogeny based on the corresponding amino acid sequences (*Paper III Figure 4*), and checked with sequence alignment analysis. Expressed homologs in diploid *A. lyrata* and *A. arenosa* were found for 60 of the total 107 *A. thaliana* MADS box genes (Parenicova et al. 2003). In addition, five M-alpha genes and a *PHERES* related clade was identified, which we have called *alpha1-5* and *PHErel1-4*. Through expression analysis we found that some genes were only expressed in either diploid *A. lyrata* or *A. arenosa*, which affected our analyses in a way that mainly explained species differences and not expression changes related to the presence or absence of the hybrid barrier. On these grounds, we omitted the species-specific genes from further analysis (*Paper III Figure S3*). The remaining 25 genes were investigated in a Principal Component Analysis (PCA) (*Paper III Figure 6*) and heat-map analysis (*Paper III Figure 7*) using TMM normalized read counts.

The PCA revealed a pattern where the two first components (PC1 47 %, PC2 35%), separates the samples in distinct groups related to whether endosperm cellularization has occurred or not (PC2) in combination with total gene expression levels or a gene expression threshold required for seed survival.
Based on the PCA together with analysis of the heat-map, we propose a set of genes to be important for endosperm cellularization and show that the late cellularizing samples lack this necessary gene expression.

Some genes do not appear to fluctuate noticeably from incompatible to compatible samples, and with no clear difference between the two time-points in the compatible crosses (AGL46, AGL35 and alpha-3, Figure 8, see section I). Compared to A. thaliana expression, AGL46 is expected to have higher
expression at the early time-point than the late. Possibly this could be indications that the *A. thaliana* homolog has a different function. *AGL35* in *A. thaliana* does not show a decline in expression until 6 DAP, which fits with the pattern seen for the sister species and their hybrids, assuming that 15 DAP is the equivalent of 6 - 9 DAP in *A. thaliana*. Lower expression is seen in 2A and 4L2A 15 DAP for *AGL35*, which possibly has a large influence on the position of these two samples in the PCA ([Paper III Figure 6](#)). *Alpha-3* shows similar pattern as *AGL46*, however, with no *A. thaliana* homolog it is difficult to evaluate its role.

A correspondence is seen between *A. thaliana* and the compatible crosses with *A. lyrata* and *A. arenosa*. The 3 DAP peak in *A. thaliana* fits with section II and III, and an *A. thaliana* peak at 6 DAP fits with section IV and V. *AGL87* does not exhibit a decline after 6 DAP which could indicate that it possibly does not fluctuate depending on endosperm cellularization, and is not related to the endosperm transition from sink to source.

In general, the early cellularizing samples exhibit low gene expression (Figure 8, left in heat-map). This fits with the genes that are lowly expressed at 15 DAP, which is seen in the cellularized compatible samples (Figure 8, section II and III). However, for the genes shown to have high expression in compatible cellularized samples in section IV and V, the early cellularized samples (2A2L and 4A2L) have low or no expression (Figure 8). If the expression of M-alpha and M-gamma genes are expected to decline as shown for *A. thaliana*, the low expression seen in the early cellularizing samples could be due to the precocious cellularization. Normal cellularization that have occurred in the compatible 15 DAP samples appears to be associated with low expression of genes in section II and III and higher expression in section IV and V. This expression pattern deviates greatly in the late cellularizing samples (2L2A and 2L4A). The most obvious differences are seen in section III and IV, in which the late cellularizing samples resemble 9 DAP expression in compatible samples. Form this we suggest that the genes in section III and IV are associated with endosperm cellularization and the post-zygotic hybridization barrier. A selection of the genes in III and IV show a uniform expression pattern in the compatible crosses and are differentially expressed in the late cellularizing samples (2L2A and 2L4A) at 15 DAP. These genes are *AGL29*, *alpha-5*, *alpha-4* that are highly expressed, and *AGL102-2*, *PHE-rel1* and *AGL86/92-1* that are lowly expressed in 2L2A and 2L4A (Figure 8). Assuming functional conservation and expression patterns of the homologs in *A. lyrata* and *A. arenosa* to *A. thaliana*, we see that the expected decline or increase is delayed or non-existent in the late cellularizing hybrids (2L2A and 2L4A). In contrast, the early cellularizing hybrids (2A2L and 4A2L) exhibit low gene expression already at the early time point, which could be associated with already declined expression.

High expression of *AGL62* and *AGL96-1* is tolerated in the 2A4L cross, indicating that their high expression alone is not deleterious at this level (Figure 8). *AGL62* is suggested to have a role in
suppressing endosperm cellularization, as the \textit{agl62} mutant cellularize early and expression is observed until the peripheral endosperm is cellularized (Kang et al. 2008). No specific function has been reported for AGL96 in \textit{A. thaliana}, but it has been shown to be also expressed in the embryo (Bemer et al. 2010), which is interesting how we observe a similar expression pattern for the two genes, and an elevated expression in the compatible hybrid 2A4L at 15 DAP (Figure 8). Furthermore, cellularization has occurred in this cross at 15 DAP (\textit{Paper I} Figure 4), and based on the expression level we see, cellularization presumably can occur despite high levels of \textit{AGL62}.

**Possible interaction between MADS box type I TFs in \textit{A. arenosa} and \textit{A. lyrata}**

AGL36 has been shown to interact with AGL28 and AGL62 (de Folter et al. 2005). Analysing the interaction map of the proteins relevant to our study of MADS box type I genes, we find that many of the interaction partners also are present in these sister species of \textit{A. thaliana} (Figure 9). Other genes are not found or is not expressed in our study set up, for instance no AGL40 homolog in diploid \textit{A. lyrata} is detected. In \textit{A. thaliana}, AGL40 is shown to interact with several other AGLs during seed development (de Folter et al. 2005, Bemer et al. 2010). AGL40 is suggested to be expressed in both embryo and endosperm in dimers of the different interaction partners (Bemer et al. 2010). There are differences in seed development between diploid \textit{A. lyrata} and \textit{A. arenosa} (\textit{Paper I}), and it would be interesting in future experiments to investigate if AGL40 has a central role in the differences observed.

\textbf{Figure 9 Protein interaction map of selected \textit{A. thaliana} MADS box type I TF.} The selection is based on relevance for the expression analysis of \textit{A. lyrata} and \textit{A. arenosa} type I MADS box genes. The TF with white filling are not detected in diploid \textit{A. lyrata} and \textit{A. arenosa} (protein name in brackets), or just in one of them (no brackets). Interactions based on yeast two hybrid (Y2H) study by de Folter et al. (2005).

\textit{AGL48} is detected in both \textit{A. arenosa}, \textit{A. lyrata} and the hybrids, but no homolog for the interaction partner AGL64 is detected. In both \textit{A. lyrata} and \textit{A. arenosa}, only one transcript was found with high sequence similarity to the clade containing \textit{AGL34}, \textit{AGL36} and \textit{AGL90}. The expected interaction partners for the TF encoded by this transcript can therefore possibly be as for AGL36 or AGL90, or a combination of the interactions suggested by the \textit{A. thaliana} study (Figure 9). The same observation was made for AGL86 and AGL92 as the \textit{A. lyrata} and \textit{A. arenosa} transcripts share common ancestry.
but the detected transcripts could not be assigned to specifically one of the two *A. thaliana* genes. Interestingly some of the interaction links are between proteins, which at gene level are shown to be expressed with different expression patterns. For instance, AGL62 interacts in Y2H with AGL92, that in *A. thaliana* is shown to have an expression peak when AGL62 is on the decline (Figure 8). The different expression curves of the MADS box type I genes (*Paper II Figure 1*), in light of the shown protein interactions (de Folter et al. 2005), may suggest that the proteins dimerize with different partners during different stages in seed development.

**Not just a triploid block, but also a triploid bridge**

The MADS box type I gene expression pattern observed in incompatible hybrids between diploid *A. lyrata* and tetraploid *A. arenosa* (Figure 8, *Paper III Figure 7*) is associated with endosperm cellularization failure and seed abortion (Figure 5, *Paper I Figure 3 and 4*). The endosperm-based post-zygotic hybridization barrier functions as a triploid block as previously described (Schatlowski and Köhler 2012). The compatible hybrids between diploid *A. arenosa* and tetraploid *A. lyrata* on the other hand, show a MADS box type I gene expression similar to the expression in the parents (2A and 4L) (Figure 8, *Paper III Figure 7*), and endosperm cellularization and seed germination same as the parents (Figure 5, *Paper I Figure 3 and 4*). The compatible diploid *A. arenosa* and tetraploid *A. lyrata* hybrids grow into mature plants, demonstrating that the hybridization barrier observed between the diploids has been bypassed. Ploidy testing of the plants revealed that they all had a relative DNA content corresponding to triploid plants, and that hybrid success was not due to unreduced gametes leading to tetraploid offspring (*Paper III Figure 8*). The mature triploid F1 hybrids (2A4L and 4L2A) were backcrossed to both the parental populations (diploid *A. arenosa* and tetraploid *A. lyrata*), demonstrating that the hybrids are semi-fertile as viable seeds were produced. In a scenario where aneuploidy is detrimental and euploid gametes was more successful in producing viable offspring than aneuploidy gametes, mainly diploid, triploid and tetraploid offspring would be expected. However, ploidy testing of the backcrossed offspring revealed a continuum of relative DNA content, indicating that gametes from the triploids resulting in mature offspring represented a range of different DNA contents, corresponding to a chromosome number ranging from eight (haploid) to 16 chromosomes (diploid) (*Paper III Figure 8*). As suggested by Ramsey and Schemske (1998), we also observed offspring of backcrossed triploids to have DNA contents fitting with gametes from triploids of euploid and near-euploid cytotypes, rather than the possible main gamete cytotype containing approximately half of the somatic chromosome number (3x/2 ± 1). In addition, we show that backcrossed F1 hybrids (BC1) can mature to adult plants with the full range of aneuploidy, and that they are semi-fertile as they can be backcrossed to the parents (*Paper III Figure 8*). The tolerance of aneuploidy that we observe is in line with previous a previous study with *A. thaliana* interploidy hybrids (Henry et al. 2005). Our results suggests that the compatible interspecies interploidy hybrids can be backcrossed and function as a triploid bridge, allowing gene flow between the species.
The Endosperm Balance Number (EBN) theory and the observed role of introgression is conflicting

The mimicry of parental excess phenotypes in reciprocal equal-ploidy crosses indicates a difference in effective ploidy related to the Endosperm Balance Number (EBN) theory (Johnston et al. 1980). EBN can be used to predict crossing outcomes, and knowing the EBN of a species could be used to predict the EBN of other species based on the outcome of a cross. A clue to endosperm development being a major cause for the post-zygotic barrier was seen in the diploid interspecies crosses between *A. lyrata* and *A. arenosa*, with non-reciprocal seed phenotypes, with too early or to late cellularization of the endosperm (Figure 3A, *Paper I Figure 1*), indicating a possible difference in effective ploidy.

The seed phenotypes of hybrids between *A. lyrata* and *A. arenosa*, appears to be related to an imbalance between the two species where *A. arenosa* has a stronger impact than *A. lyrata*, as *A. arenosa* dictates the seed phenotype in crosses with *A. lyrata*. The seeds are smaller when *A. arenosa* acts as mother in the cross mimicking maternal excess, and shrivelled and collapsed seeds are produced when *A. arenosa* acts as father, mimicking paternal excess, as if *A. arenosa* has a higher ploidy than *A. lyrata*. Following the EBN theory this could suggest that *A. arenosa* has a higher EBN than *A. lyrata*. EBN is a relative number assigned to different species, based on the know crossing outcome of related species. If EBN in diploid *A. arenosa* is set to 2, diploid *A. lyrata* would have a lower number, for instance 1. Given this relationship, the expected outcome of a cross with tetraploid *A. arenosa* and diploid *A. lyrata* would be inviable seeds if the tetraploid has a proportional increase in EBN, making the difference in EBN even greater (for instance 4EBN vs. 1EBN). In accordance with the scenario described here, the outcome of crosses between diploid *A. lyrata* and tetraploid *A. arenosa* is more extreme, with lower germination frequency (~0 %, *Paper I Figure 3*), compared to the diploid interspecies cross (~0-10%, *Paper I Figure 1*). Given that the tetraploids have the double EBN of the diploids, or that the difference in EBN between the tetraploids is the same as between the diploid, the tetraploid interspecies cross is expected to be incompatible in the same manner as the diploid interspecies cross. However, the tetraploid interspecies cross, with the tetraploid plants from the populations used to cross with the diploids, the species barrier was absent as more than 90 % of the seeds germinated (4A-KER crossed to 4L-WAC) (*Paper I Figure 3*). A possibility could be that the EBN relationship is not linear and that the relationship between the diploids cannot be directly extrapolated to the tetraploids. Alternatively, higher ploidy in itself affects the barrier, given extra genetic material. However, such a scenario would indicate a different mechanism than what EBN propose. Although often referred to, the proposed buffering effect of extra genome copies have few studies that can back up this hypothesis (Mable and Otto 2001, Ha et al. 2009), and following the trail to a commonly used reference (Matzke and Matzke 1998) ends up in a reference to an unpublished dataset (Matzke and Matzke 1998).
The EBN theory indicates that endosperm factors should be doubled in the case of tetraploidization if no mutations have occurred. Johnston et al. (1980) suggests that in the case of mutations to these endosperm factors, rapid changes in the compatibility of crosses can occur. To this end, the unexpected compatibility of the tetraploid *A. lyrata* and *A. arenosa* could be due to genetic variation affecting the endosperm factors in tetraploids compared to the diploids. In light of results presented in paper III, tetraploid interspecies crosses between plants of non-introgressed populations show crossing success more in line with the EBN theory. The contrasting results between the crosses with introgressed and non-introgressed plants indicate that the introgressed genetic material has affected the endosperm factors and resulted in a shift in EBN, making the introgressed plants of the two species of the same EBN and compatible. As a consequence, it would be expected that in taxa that diverged long ago, thus accumulating mutations, or in populations known to have experienced introgression, EBN cannot be predicted. A possibility could be that introgression and mutations affects the proposed linear relationship that is expected from the original EBN theory. If genetic variation affecting the endosperm linked to EBN is changed, without causing a complete hybridization block, the result may be an intermediate barrier.

Based on our findings, we suggest that for the populations used in our studies, EBN and the crossing outcome can be predicted for crosses where the populations are non-introgressed, or the divergence of EBN is substantial, as between diploid *A. lyrata* and the other species. Moreover, in the tetraploid populations that have experienced introgression, a weak hybridization barrier is expected (Paper I and III). With our current dataset, we cannot resolve whether the populations are introgressed because they were completely compatible in the first place, thus allowing introgression, or if they are compatible to the observed extent due to one or many events of introgression.

4L-LIC is indicated to have a lower effective ploidy than the three tetraploid *A. arenosa* populations used in this work, since the seed appearance seems to be dictated by *A. arenosa* in all crosses (*Paper III Figure 2*). When 4A-KRO and 4A-KER crossed as mother to 4L-LIC, the hybrid seeds appear smaller than the parent seeds, resembling a maternal excess cross. However, many of the seeds look viable and germination frequency is around 50%. The reciprocal crosses on the other hand, the seeds appear collapsed, possibly due to delayed timing of endosperm cellularization, mimicking a paternal excess cross. In these crosses, where 4L-LIC acts as seed donor the seed germination frequencies are significantly lower, than when 4L-LIC is pollen donor. In relation to the EBN theory, this asymmetry can suggest that one or more of the endosperm factors are parent-of-origin specific (imprinting) or dosage dependent (endosperm contributing maternal to paternal ratio).

On observation of seed appearance, assuming that *A. arenosa* dictates the seed phenotype, I propose the following EBN relationship for the populations used in Paper III:
Genetic variation in one or more loci affecting the endosperm in line with the EBN theory may be the cause for the varying strength of the hybridization barriers between the tetraploids. In the proposed relationship where some populations are suggested to be similar or greater than (≥), the further away from each other two populations are, the greater the chance of hybrid failure.

Based on backcrossing of the diploid *A. arenosa* to *A. lyrata* hybrids we find support to the suggestion that the hybridization barrier between *A. arenosa* and *A. lyrata* is due to a complex multilocus genetic basis (Paper I Figure 6). In crosses between diploid *A. lyrata* mother and diploid *A. arenosa* father we find that, depending on the diploid *A. arenosa* population, the hybridization barrier is based on two or three paternal loci that interact with one maternal locus in *A. lyrata*. In the reciprocal cross, we find that the barrier is based on three loci from each parent. This suggested complex multiple loci interaction is in line with other studies in *A. thaliana* (Burkart-Waco et al. 2012), between *Solanum* species three unlinked loci are suggested (Ehlenfeldt and Hanneman 1988), *Capsella* (Rebernig et al. 2015), and in *Mimulus* where loci with parent-of-origin effects are suggested to be part of the hybridization barrier loci (Garner et al. 2016).

A study investigating hybrids between diploid tomato species it was discovered a species with possibly higher effective ploidy compared to two other diploid tomato species (Roth et al. 2018b). Roth et al. (2018b) find that the main part of differentially expressed genes in maternal excess crosses are involved in nutrient allocation, whereas in the paternal excess crosses they are largely involved in transcription regulation and hence growth. In addition they find that transcription factors of the MADS box and WRKY gene families are shown to be affected in hybrid seeds. Tomato as well as e.g. *Mimulus* have cellular endosperm development separating them from species with syncytial endosperm development. However, recent work show that the endosperm is affected in interspecies crosses, and imprinted genes analysed, indicate that the mode of endosperm development possibly is not crucial with regards to interspecies hybridization (Florez-Rueda et al. 2016, Kinser et al. 2018, Roth et al. 2018a). In this regard, investigations towards the transition where the embryo start feeding on the endosperm, would be a fruitful future research area.

### Possible regulatory mechanism for deleterious gene expression in incompatible hybrids

The regulatory mechanism for the deleterious gene expression we observe in the incompatible *A. arenosa* and *A. lyrata* hybrids have not been investigated in our studies. However, the cause for the prolonged 9 DAP expression in late cellularizing samples could be due to a delay in the regulatory mechanism in charge of deregulating these genes. In addition, or linked, it could be a delay in the upregulation of the 15 DAP genes. In the study by Walia et al. (2009), the hybridization barrier
between diploid *A. thaliana* and diploid *A. arenosa* is bypassed by using a tetraploid *A. thaliana* mother. They show that several MADS box type I genes are upregulated in the incompatible hybrids, along with a low expression of *FIS2*, compared to the compatible hybrids. The incompatible cross is between diploid *A. thaliana* mother and diploid *A. arenosa* father, which is an equal-ploidy cross. As the tetraploid *A. thaliana* rescue the hybrid cross, it suggests that diploid *A. thaliana* has a lower EBN than diploid *A. arenosa*, and that this diploid hybrid cross with *A. arenosa* as father mimics a paternal excess cross. Our results further supports that an upregulation of for instance *AGL36* and *AGL62* is associated with a paternal excess cross rather than a maternal excess cross. However, Walia et al. (2009) also show upregulation of *PHE1* and *PHE2* in their incompatible hybrids, which we do not find in our results. Although, there are obvious differences between the sister species when it comes to the *PHERES* genes, as an extra *PHERES* clade was identified in *A. lyrata* and *A. arenosa*, it is suggested that the relationship of these genes across taxa is complex. The upregulation in the incompatible crosses is shown to correlate with a low expression of *FIS2*, which is in line with the shown role of the FIS-PRC2 complex in down regulating MADS box type I genes (Paper II) (Zhang et al. 2018). One example is *AGL36*, which is upregulated 40 times in a FIS-PRC2 mutant *mea* at a late time-point (6 DAP) (Shirzadi et al. 2011). To this end, it suggests that FIS-PRC2 components are expressed at a low level in our paternal excess crosses, both actual (2L4A) and effective excess crosses (2L2A), resulting in the continued 9 DAP observed.
Conclusion

The main aim of our studies has been to investigate the molecular and genetic basis of the post-zygotic hybridization barriers between *Arabidopsis* species. We have succeeded in describing the molecular basis by showing that the barrier is connected to the transition of the endosperm going from a sink of resources to supplying the embryo during seed development. The transition is marked by the event of cellularization and we have shown how incompatible hybrids between *A. thaliana* and *A. arenosa*, and *A. arenosa* and *A. lyrata*, have endosperm that has cellularized too early or too late. This is in line with similar studies in other species. We show how this barrier is bypassed in the cross between diploid *A. arenosa* and tetraploid *A. lyrata* with normal endosperm cellularization, suggesting that the tetraploidization of *A. lyrata* is the reason for this. Implying that the doubling of genetic elements in the *A. lyrata* genome makes it compatible with *A. arenosa*, gives support to the Endosperm Balance Number (EBN) theory. In crosses between diploid *A. arenosa* and diploid *A. lyrata* we see a non-reciprocal effect on these inviable hybrid seeds, which in light of the EBN theory, suggest that they have different EBNs and that *A. lyrata* has a lower number than *A. arenosa*. Increasing the ploidy of *A. lyrata* and therefore increasing the endosperm factors, results in a correct interplay between the maternal and paternal contribution to the endosperm. We find that the hybridization barrier between diploid *A. lyrata* and *A. arenosa* is based on multiple loci, in line with investigations in other species systems.

But what are these endosperm factors and important loci? In the search for specific genes important for the post-zygotic hybridization barrier, we have investigated MADS box type I transcription factor (TF) genes due to their previously shown expression change in incompatible *Arabidopsis* hybrids and the fact that the gene family is predominantly expressed in the endosperm. Only some of the MADS box type I TF genes have previously been investigated in detail and by showing elaborate expression profiles in *A. thaliana* we have improved the foundation for the research on their role in the endosperm and the endosperm-based hybridization barrier. In hybrids between *A. lyrata* and *A. arenosa* we find that the expression patterns are similar to the parents in compatible hybrids. In contrast, we find that the early cellularized crosses possibly already have finished most of their MADS box type I TF expression at the early time-point. While in the late cellularizing samples the expression still resembles 9 DAP expression when the samples are at 15 DAP. Comparing to results presented by Walia et al. (2009), it is suggested that low expression of FIS-PRC2 components in the late cellularizing samples could be the cause for the deleterious MADS box type I expression. It would be interesting to test if in the reciprocal cross, where the endosperm cellularize early, an elevation is seen for the same regulatory components.

We also find that imprinting of *AGL36* is conserved in diploid *A. arenosa* and that the imprinting status is maintained in *A. arenosa* and *A. lyrata* hybrids, but not in a hybrid cross between *A. thaliana*
and *A. arenosa*. From hybrid crosses between *A. thaliana* and *A. arenosa* we discovered that the hybridization barrier is temperature dependant, resulting in lower germination frequencies at a higher temperature. The mechanism for this difference when increasing the temperature 4 °C remains to be uncovered.

Finally, we have found that the post-zygotic hybridization barrier is stronger in crosses between non-introgressed tetraploid populations of *A. lyrata* and *A. arenosa*, than between two introgressed populations. This could be due to natural variation between populations, or possibly due to an allelic variation specifically in the endosperm factors important for the here described endosperm-based post-zygotic hybridization barrier.
Future perspectives

Determine seed set affecting factors

As a measure of fitness, knowing the expected seed set for the different populations would have been most useful. From small scale trials we found this difficult to determine due to large variation from plant to plant, in addition to individual within-plant variation (results not shown). For future projects, factors like expected number of ovules, crossing partner compatibly, pollen viability, and ideal timing of pollination should be established.

Describe *A. lyrata* and *A. arenosa* differences

*A. arenosa* is the more dominant of the two species, which is central to the EBN theory, but it remains unknown what the underlying mechanisms are. One possibility is differences in imprinted genes leading to a possible deleterious gene dosage, causing the hybrid offspring to abort. In order to investigate the role of gene dosage on the hybridization barrier, mapping of imprinted genes in the two species should be performed.

On observations, diploid *A. lyrata* seems more prone to fungus and insect attacks, suggesting differences in stress responses. Thus when searching for genetic differences, it may be expected large variations in genes involved in stress-related pathways.

Investigate allelic differences in the different tetraploid *A. arenosa* populations.

The tetraploid *A. arenosa* populations used in paper III when crossed to tetraploid *A. lyrata*, resulted in seed germination frequency from 0 to 95 percent, indicating that differences between the populations have an effect on the hybridization barrier. The difference in seed viability could be due to allelic or gene dosage variation, which can be investigated by doing a differential expression analysis and SNP calling on early developing seed. Based on transcriptome analysis in paper III, were we saw indications of detrimental gene expression already at 9 DAP, the time point of collecting seed tissue should be prior to this.


Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe


Based on the biological species concept, two species are considered distinct if reproductive barriers prevent gene flow between them. In Central Europe, the diploid species *Arabidopsis lyrata* and *Arabidopsis arenosa* are genetically isolated, thus fitting this concept as “good species.” Nonetheless, interspecific gene flow involving their tetraploid forms has been described. The reasons for this ploidy-dependent reproductive isolation remain unknown. Here, we show that hybridization between diploid *A. lyrata* and *A. arenosa* causes mainly inviable seed formation, revealing a strong postzygotic reproductive barrier separating these two species. Although viability of hybrid seeds was impaired in both directions of hybridization, the cause for seed arrest differed. Hybridization of *A. lyrata* seed parents with *A. arenosa* pollen resulted in failure of endosperm cellularization, whereas the endosperm of reciprocal hybrids cellularized precociously. Endosperm cellularization failure in both hybridization directions is likely causal for the embryo arrest. Importantly, natural tetraploid *A. lyrata* was able to form viable hybrid seeds with diploid and tetraploid *A. arenosa*, associated with the reestablishment of normal endosperm cellularization. Conversely, the defects of hybrid seeds between tetraploid *A. arenosa* and diploid *A. lyrata* were aggravated. According to these results, we hypothesize that a tetraploidization event in *A. lyrata* allowed the production of viable hybrid seeds with *A. arenosa*, enabling gene flow between the two species.

According to the biological species concept, reproductive isolation is a major criterion for defining two distinct species (1). Nevertheless, hybridization has long been recognized to play an important role in plant evolution, with many reported cases of related species that, either punctually or continuously in their life history, have experienced gene flow, leading to introgression of adaptive traits or hybrid speciation (2-9). Thus, the “tree of life” is considered more as a complex network than an actual tree (10, 11). Whereas gene flow between previously isolated species can occur after a secondary contact (5, 12), the mechanisms contributing to this gene flow are largely unknown. The evolutionary relationships of the *Arabidopsis* genus may also be depicted as a network rather than a dichotomizing tree (9). For example, the two tetraploid species *Arabidopsis suecica* and *Arabidopsis kamchatka* both have a hybrid (allopolyploid) origin involving *Arabidopsis thaliana* and *Arabidopsis arenosa* as parental species in *A. suecica* and *Arabidopsis lyrata* and *Arabidopsis halleri* ssp. *gemmifera* in *A. kamchatka* (13, 14). Gene flow has also been observed between *A. lyrata* and *A. arenosa* (15, 16). These two species, which are believed to have originated and radiated in Central Europe around 2 Mya (17), both exist in a diploid and tetraploid form. Diploid *A. lyrata* colonized central and northern Europe as well as northern America, whereas the tetraploid form is limited to eastern Austria (15). For *A. arenosa*, the diploids are mainly found in the Carpathians and southeastern Europe, whereas the tetraploids occupy central and northern Europe (15, 18, 19). Diploid *A. lyrata* and *A. arenosa* are considered “good species” because they are genetically and phenotypically distinct and no past or recent gene flow can be detected between them (15, 16). Although the geographical separation between these species can partly explain this observation, it is likely that other reproductive barriers also contributed to the divergence between diploid *A. lyrata* and *A. arenosa*. In contrast, two main hybridization zones between tetraploid *A. lyrata* and tetraploid *A. arenosa* have been described in Austria (15, 16). It has been suggested that the gene flow between tetraploid *A. lyrata* and *A. arenosa* is bidirectional (16). Another study, however, found that, although tetraploid *A. lyrata* was highly introgressed by *A. arenosa* in the contact zones, the tetraploid *A. arenosa* populations showed a low hybrid index in the same region (15). Based on these data, unidirectional gene flow from *A. arenosa* to tetraploid *A. lyrata* was proposed, and the authors concluded that tetraploid *A. lyrata* originated and spread in Austria as a consequence of an original hybridization of *A. arenosa* with *A. lyrata* (15). Whether this original

Significance

Hybrid seed lethality has been recognized and addressed by a long-standing tradition of plant-breeding research. Nevertheless, its role in evolution and speciation has been underestimated. In this study, hybrid seed lethality between *Arabidopsis lyrata* and *Arabidopsis arenosa*, two model species of growing interest in the scientific community, was investigated. This study shows that endosperm defects are sufficient to explain the direction of gene flow between the two wild species, suggesting an important role of this hybridization barrier in plant speciation. In addition, we show that natural polyploidization is involved in breaking down hybridization barriers, not only establishing them, as implied by the traditional “triploid block” concept. Our data suggest that polyploidy-mediated hybrid seed rescue, long known in artificial crosses, could play an important role in plant evolution.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.
hybridization event occurred between diploids, tetraploids, or in an interploid manner remains unresolved.

The presence of interspecific gene flow involving tetraploid, but not diploid A. lyrata, raises the question whether polyplodization of A. lyrata actually facilitated the hybridization success with A. arenosa. Indeed, even though new polyploids are often reproductively isolated from their diploid congeners, referred as the “triploid block” (20), polyploidization is also associated with the bypass of interspecific reproductive barriers (21–25). Postzygotic interspecific reproductive barriers rely on similar mechanisms as the triploid block; both affect hybrid seed survival as a consequence of endosperm developmental failure. The endosperm of most angiosperms is a triploid tissue that develops after fertilization of the homodiploid central cell to support embryo growth (26). It is a dosage-sensitive tissue, which strictly requires a relative maternal/paternal genome ratio of 2:1 to develop correctly (27, 28). Interploidy hybridizations disrupt the 2:1 ratio in the endosperm, leading to asymmetric endosperm developmental failure depending on the direction of hybridization. Increased ploidy of the paternal parent causes delayed or failed endosperm cellularization, whereas increased maternal ploidy induces precocious endosperm cellularization (29–31). Endosperm cellularization is a major developmental transition that, if impaired, leads to embryo arrest and seed lethality (32–34). As a consequence, interploidy hybrid seeds are usually inviable (29–31, 35, 36). Interestingly, very similar nonreciprocals seed phenotypes are also observed in interspecies hybrid seeds, even between species of same ploidy levels (36–39). This finding suggests that species, despite having the same ploidy level, differ in their relative genome dosage, resulting in unbalanced contributions of cellularization factors leading to cellularization failure. Increasing the ploidy level of the species with predicted lower genome dosage leads to increased hybrid seed viability, confirming the quantitative character of the interspecies hybridization barrier (21, 25).

According to the endosperm balance number (EBN) hypothesis, the effective ploidy of a species must be in a 2:1 maternal to paternal ratio in the endosperm to allow normal endosperm development (21–23, 40). It has been proposed that species can have different effective ploidies, explaining the quantitative nature of hybrid seed failure. In the Solanum genus (21) and in several other genera, like Impatiens (22) and Trifolium (23), each species was assigned a numerical effective ploidy (or EBN). Using this concept, the hybridization success between related species can be predicted; species with similar effective ploidy or EBN can successfully hybridize, whereas species with different EBN give inviable hybrid seeds, unless the ploidy of the lower EBN species is increased to match the EBN of the other species (21, 40). Based on the assigned EBNs, these studies were accurately able to predict the viability of artificial interspecific crosses. It remains unresolved, however, whether EBN divergence and its bypass by polyploidization is important for hybrid seed viability in the wild and whether it regulates gene flow between wild species.

Reciprocal crosses between diploid A. lyrata and diploid A. arenosa exhibit impaired hybrid seed viability in a nonreciprocal manner (41); however, the cause for this barrier has not been identified. In this work, we tested the hypothesis that both species are reproductively isolated by an endosperm-based reproductive barrier and thus differ in their EBN. Our data add strong support to this hypothesis; in incompatible hybrid cross combinations, endosperm cellularization is either precocious, delayed, or does not occur at all. Increasing the ploidy of A. lyrata is sufficient to overcome this hybridization barrier, assigning A. lyrata a lower EBN compared with A. arenosa. Our data let us to propose that the tetraploidization of A. lyrata initiated the gene flow between A. lyrata and A. arenosa observed in Central Europe.

Results

Hybridizations of Diploid A. lyrata and A. arenosa Lead to Mostly Inviable Seeds. To test the hypothesis that the central European diploid populations of A. lyrata and A. arenosa are reproductively isolated by a postzygotic barrier acting in the endosperm, we performed manual reciprocal crosses between these two diploid species (for population details, see Materials and Methods and Fig. S1). Seeds derived after hybridizations of A. lyrata seed parents with A. arenosa pollen donors (denoted A. lyrata × A. arenosa) were dark and shriveled, whereas the reciprocal hybrid seeds were severely reduced in size but appeared normally shaped compared with nonhybrid seeds (Fig. 1A). However, independent of the seed phenotype, hybrid seeds of both cross directions largely failed to germinate (Fig. 1B). A. lyrata × A. arenosa seeds were of similar size compared with the average of parental seeds (Fig. 1C, midparent value) (42), but weighed significantly less (Fig. 1D). In the reciprocal A. arenosa × A. lyrata cross, seeds were significantly smaller and weighed significantly less compared with the midparent value (Fig. 1C and D).

Nonreciprocal Endosperm Defects Cause Diploid A. lyrata/A. arenosa Hybrid Seed Inviability. The cross-direction–dependent phenotype of these hybrid seeds resembles those observed in interploidy crosses (29–31) as well as other interspecies crosses (36, 38, 43). Delayed or absence of endosperm cellularization or precocious cellularization has been linked to the inviability of interploidy and interspecific hybrid seeds (36, 38, 43). We therefore performed a detailed analysis of seed development and the endosperm cellularization phenotype in A. lyrata × A. arenosa reciprocal hybrids.

Diploid (2x) A. lyrata seeds developed at a slower pace compared with 2x A. arenosa seeds, as revealed by monitoring embry development at defined time points (Fig. S2). However, the
endosperm of both species was still syncytial at 9 d after pollination (DAP) (Fig. 2A). In contrast, at 9 DAP the A. arenosa × A. lyrata hybrid endosperm was fully cellularized, suggesting a shortened syncytial phase and an earlier onset of cellularization (Fig. 2A). In contrast, the endosperm of A. lyrata × A. arenosa hybrid seeds was still syncytial at 15 DAP, when the endosperms of both parents were cellularized. Cellularization failure correlated with a delay in embryo development: when the embryo of the parents had reached at least the early bent cotyledon stage, the A. lyrata × A. arenosa hybrid embryos were still at the heart stage.

To test whether differences in endosperm cellularization were connected to differences in endosperm proliferation, we assessed endosperm proliferation in parental seeds by counting endosperm nuclei at 5, 7, and 9 DAP. Consistent with phenotypic observations, A. lyrata had fewer endosperm nuclei, corresponding to one cycle of endosperm nuclear division less at 5 DAP in A. lyrata compared with A. arenosa (Fig. 2B). However, after 5 DAP the rate of proliferation was similar between the two species (Fig. 2C), suggesting that the initial endosperm proliferation in A. lyrata is either delayed or slower than in A. arenosa. In contrast, the number of endosperm nuclei in A. lyrata × A. arenosa hybrid seeds was substantially higher than that of the maternal A. lyrata parent at all analyzed time points (Fig. 2B), whereas the proliferation rate in hybrid seeds was similar to the maternal parent (Fig. 2C), suggesting that the difference in nuclei number is a consequence of a delayed or slower onset of endosperm proliferation in the maternal A. lyrata parent compared with the hybrid. In contrast, A. arenosa × A. lyrata hybrid endosperm started to degenerate after 5 DAP and proliferation could thus not be observed. The total nuclei number at 5 DAP was comparable to the paternal A. lyrata parent. Taken together, these data suggest that hybrid endosperm cellularization timing is not coupled to endosperm proliferation, and that both processes are independently genetically controlled.

Previous studies have implicated endosperm cellularization failure as a primary cause for hybrid embryo and seed arrest (29–31, 37, 38, 43). To test the hypothesis that A. lyrata × A. arenosa hybrid embryo arrest is causally connected to abnormal endosperm cellularization, we rescued hybrid embryos by dissecting them from the seeds and growing them in vitro in embryo-supporting culture. Dissected embryos survived at a higher rate than when germinated from seeds (54% vs. <10%, respectively; n = 29 dissected embryos). Seedlings regenerated from this experiment were viable and developed into healthy adult hybrids (Fig. S3A and B) with a phenotype that was intermediate between the dark green, waxy serrate leaves of the parental A. arenosa accession and the lighter green, lobed leaves of the A. lyrata accession used for

![Fig. 2. Impaired viability of diploid A. lyrata × A. arenosa reciprocal hybrid seeds is correlated with asymmetric endosperm cellularization defects, but not with endosperm proliferation. (A) Feulgen-stained seeds of A. lyrata, A. arenosa, and reciprocal hybrids are depicted. Early (9 DAP) and late (15 DAP) time points are displayed. At 15 DAP, A. arenosa seeds (2A × 2A) are very large and mostly occupied by the embryo. For a better display of the remaining endosperm, the larger inlay represents the entire seed and the smaller inlay represents the zoomed-in endosperm region. Images shown in this figure are representative for a total number of observed seeds of at least 20 per biological replicate (three replicates). (Scale bars, 50 μm.) (B) Endosperm proliferation was assessed by measuring endosperm nuclei number at 5, 7, and 9 DAP, represented as absolute endosperm nuclei number. (C) Proliferation rate was calculated as the ratio of nuclei number between 5 and 7 DAP and between 7 and 9 DAP. 2A: 2x A. arenosa; 2L: 2x A. lyrata. Parental species are represented by dark red (A. arenosa) and dark blue (A. lyrata) colors and reciprocal hybrids by light shades (light red and light blue for A. arenosa × A. lyrata and A. lyrata × A. arenosa, respectively). The cross (†) at 9 DAP indicates that endosperm nuclei for A. arenosa × A. lyrata were not detectable. In B, error bars represent SD between three biological replicates (n = 5 seeds per sample). In C, the error bars were calculated as follows: ratio value × (coefficient of variation×SEM^2 + (coefficient of variation×SEM^2))/2.
this cross (Fig. S3 B–D). In vitro survival of hybrid embryos strongly supports the hypothesis that endosperm failure is the primary cause for hybrid seed lethality.

We thus conclude that endosperm cellularization defects are responsible for the failure of 2x A. arenosa × 2x A. arenosa reciprocal hybrid seeds and that endosperm cellularization and endosperm proliferation can be uncoupled.

Increased Ploidy of A. lyrata and A. arenosa Impacts Postzygotic Reproductive Barrier Strength. In contrast to diploid populations, natural populations of tetraploid A. lyrata and A. arenosa hybridize (15, 16). We aimed to test the hypothesis that polyploidization impacts the postzygotic hybridization barrier and thus may explain gene flow between the two species. We therefore performed reciprocal crosses between diploids and natural tetraploids of the two species in all possible combinations (Fig. 3A and Fig. S4).

Strikingly, when A. lyrata natural tetraploids (4x) were reciprocally crossed with 2x A. arenosa, seeds were fully viable in both cross directions, as judged by seed shape and germination tests (Fig. 3A and B). The size of 2x A. arenosa × 4x A. lyrata hybrid seeds was similar to the midparent value, whereas 4x A. lyrata × 2x A. arenosa seeds were significantly smaller (Fig. 3D). Hybrid seed viability was associated with restored endosperm cellularization, occurring around 15 DAP in both cross directions, similar to the tetraploid parents (Fig. 4 A–H and Fig. S4). In both cross directions the endosperm was syncytial at 9 DAP, revealing similar to the tetraploid parents (Fig. 4).

In conclusion, the sole tetraploidization of A. lyrata was sufficient to bypass the endosperm-based reproductive barrier between A. lyrata and A. arenosa. Therefore, we propose that a tetraploidization event in A. lyrata allowed the production of viable hybrid seeds, enabling gene flow between the two species.

The Hybridization Barrier Separating Diploid A. lyrata and A. arenosa Has a Multigenic Basis. We inferred the genetic basis of the postzygotic incompatibility between A. lyrata and A. arenosa by observing the phenotypic segregation in diploid F1 backcrosses. Given the nonreciprocity of hybrid seed defects, we analyzed A. lyrata × A. arenosa F1 hybrids in both cross directions. Thus, we backcrossed F1 hybrids (♂) to A. lyrata (♀) to infer the number of paternal A. arenosa loci, and F1 hybrids (♀) × A. arenosa (♂) backcross allowed to assess the number of maternal A. lyrata loci
underlying the A. lyrata (♀) × A. arenosa (♂) hybrid seed lethality (Fig. 6). To infer the genetic loci underlying incompatibility in the A. arenosa × A. lyrata cross, we analyzed F1 hybrids (♀) × A. lyrata (♂) and A. arenosa (♀) × F1 hybrids (♀). We generated two different A. lyrata × A. arenosa F1 hybrids (F1A and F1B) using A. arenosa individuals from two different populations (Table S1) in backcrosses with parental plants. Strikingly, A. lyrata × F1 seed abortion rates varied from 15 to 69% depending on the F1 plant used, suggesting segregating alleles in the parental species (Fig. 6A). Backcrosses using F1A produced 15% (n = 101 seeds) aborted seeds, which is not significantly different (P = 0.17, χ² test) from a theoretical expectation of 12.5% for a combination of three A. arenosa paternal loci required to induce seed abortion (38). However, F1B backcrosses resulted in 69% (n = 142 seeds) seed abortion, suggesting two independent A. arenosa paternal loci being involved with each of them sufficient to give hybrid seed abortion (expected seed abortion 75%, P = 0.15, χ² test) (Fig. 6A). Because F1A and F1B hybrids were generated with A. arenosa individuals from two different populations, these data suggest that the genetic basis of the A. lyrata × A. arenosa incompatibility differs between A. arenosa populations, but nevertheless results in the same hybrid seed phenotype.

Crossing the A. lyrata × A. arenosa F1 individuals as seed parent with A. arenosa produced similar seed abortion rate independently of the F1 origin [50% for F1A (n = 230 seeds) and 43% for F1B (n = 126 seeds), P = 0.26 for F1A and P = 0.31 for F1B, χ² test] (Fig. 6B). These data suggest that one A. lyrata maternal locus is sufficient to induce seed abortion upon interacting with one or multiple A. arenosa loci.

Hybrid seeds derived from A. arenosa × A. lyrata crosses have a normal shape but are extremely small compared with parental seeds (Fig. 1C). Therefore, seed size was measured in the F1 backcrosses and the proportion of seeds mimicking the A. arenosa × A. lyrata phenotype (seed size < 0.21 mm²), thus considered as “abnormal,” was estimated (Fig. 6C and D). In A. arenosa × F1 and F1 × A. lyrata, the rate of abnormal seeds reached 15.6% (n = 109 seeds) and 12% (n = 50 seeds), respectively. This was not significantly different from the theoretical expectation of 12.5% in both cases (P = 0.34 and P = 1, respectively, χ² test), suggesting that interaction of three paternal and three maternal loci underlies genetic incompatibility of A. arenosa × A. lyrata.

In conclusion, our results suggest that the genetic basis of the A. lyrata × A. arenosa incompatibility is not fixed within A. arenosa and differs between A. arenosa populations. Depending on the A. arenosa population, this hybridization barrier is built by two to three A. arenosa paternal loci that interact with one A. lyrata maternal locus to produce hybrid seed lethality. In contrast, the A. arenosa × A. lyrata incompatibility involves three parental loci on each side of the cross. Together, our data show...
that the endosperm-based hybridization barrier between *A. lyrata* and *A. arenosa* is based on a complex multiple loci interaction.

**Discussion**

In this study we found that endosperm-based hybridization barriers and bypassing of such by natural polyploidization can explain the gene flow between *A. lyrata* and *A. arenosa*. Until now, endosperm-based hybridization barriers have not been considered to play a major role in plant speciation, but the present study strongly suggests otherwise. We furthermore show that polyploidization can break down hybridization barriers and not only establish them, as implied by the classic triploid block concept. Polyploidy-mediated restoration of hybrid seed viability has been used in plant breeding to obtain artificial hybrids. This study, however, suggests that this phenomenon exists in the wild as well, playing an important role in reenabling gene flow between species. Although this work focuses on the case study of two sister species, *A. lyrata* and *A. arenosa*, the growing interest of the scientific community for endosperm-based hybridization barriers suggests that similar findings will be discovered in other plant species over the next years.

**Endosperm-Based Hybridization Barriers Play an Important Role in Controlling Interspecific Gene Flow in *A. lyrata* and *A. arenosa***

Species divergence is a highly dynamic process, associated with the accumulation of multiple reproductive barriers over time (12). Consequently, untangling the life history that led to the current state of gene flow, geographical distributions, or ecological preferences between species is rather complex, especially in the case of species with several ploidy levels, such as *A. lyrata* and *A. arenosa*. Therefore, measuring gene flow and the strength of specific reproductive barriers is a first important step toward understanding the life history of species. In this study, we focused on the endosperm-based postzygotic barrier preventing the formation of viable hybrid seeds between *A. lyrata* and *A. arenosa* and compared these results with documented gene flow between these species.

Two previous studies found no evidence for gene flow between diploid *A. lyrata* and *A. arenosa* (15, 16), consistent with our data demonstrating low viability of hybrid seeds in both hybridization...
directions. Our data thus suggest that hybrid seed lethality has contributed to the long-lasting complete isolation between the two diploid species. Nevertheless, hybrid seed viability rates reported in another study involving controlled crosses between diploid *A. lyrata* and *A. arenosa* were substantially higher (41). This difference could be because of genetic variation between natural populations, consistent with our data showing that the genetic basis of *A. lyrata × A. arenosa* hybrid seed lethality differs among *A. arenosa* populations. The presence of distinct incompatibility loci in *A. arenosa* populations suggests a recent intraspecific diversification of those loci and therefore a fast evolving genetic basis of hybrid incompatibility, similar to that proposed for other species (38, 44, 45). Genomic imprinting, the epigenetic phenomenon by which genes are expressed in a parent-of-origin manner, has been causally linked to endosperm-based hybridization barriers and is disturbed in interspecific hybrid seeds (24, 46–48). Moreover, genomic imprinting has been shown to vary among populations of the same species (49, 50). Together with parent-of-origin defects observed in the hybrid seeds, these facts make genomic imprinting a promising candidate molecular mechanism to explain interspecific hybrid seed lethality.

Tetraploid *A. arenosa* and diploid *A. lyrata* do not exhibit any gene flow (16), even though their geographical distributions overlap (51, 52). The strong reproductive barrier identified in our study provides an explanation as to how the two species can coexist in sympathy, suggesting that this reproductive barrier is maintained as a reinforcement mechanism, as previously proposed (39, 53). Tetraploids of both species coexist and hybridize in the wild (15, 16) and bidirectional gene flow was observed between the two species (16). These observations are fully supported by our data showing that hybrid seeds between tetraploid *A. lyrata* and *A. arenosa* are viable in both cross directions. Contrasting these observations, another study reported that gene flow between tetraploids was mainly unidirectional from *A. arenosa* to *A. lyrata* (15), suggesting that additional reproductive barriers act in a nonreciprocal way. We did not find any apparent difference in fertilization success depending on the cross direction, and therefore hypothesize that either additional postzygotic barriers impair 4x*A. arenosa × 4x*A. lyrata* hybrid fitness, or that 4x*A. lyrata × 4x*A. arenosa* hybrids are favored. In support of the latter hypothesis, 4x*A. lyrata × 4x*A. arenosa* seeds were bigger with later endosperm cellularization than the reciprocal hybrid seeds, which may promote early seedling growth similar as reported for viable paternal excess triploid seedlings in *A. thaliana* (54).

In conclusion, the strong association between documented gene flow and the viability of hybrid seeds suggests an important role for endosperm-based reproductive barriers in controlling gene flow between *A. lyrata* and *A. arenosa*.

**Gene Flow Between *A. lyrata* and *A. arenosa* May Have Been Enabled by *A. lyrata* Tetraploidization.** Endosperm cellularization has been suggested to regulate resource accumulation by shifting the resource sink from the endosperm to the embryo, and is therefore crucial for the embryo feeding and survival (55, 56). Thus, impairing this developmental transition leads to embryo arrest and seed lethality (33, 34). In this study, we showed that the inviability of hybrid seeds between 2x*A. lyrata* and 2x*A. arenosa* is due to nonreciprocal endosperm cellularization defects. In 2x*A. lyrata × 2x*A. arenosa* hybrid seeds, the endosperm fails to cellularize, whereas in the reciprocal cross direction the endosperm cellularizes precociously. Similar defects were observed in interplioidy hybrid seeds originating from paternal and maternal genome excess, respectively (29–31). Abnormal endosperm development in 2x*A. lyrata × 2x*A. arenosa* reciprocal hybrid seeds thus mimics a genome dosage imbalance, with *A. arenosa* behaving as a higher ploidy parent. According to the EBN hypothesis, this would imply that at the same ploidy level the *A. arenosa* genome contribution to the endosperm has a higher effective ploidy than the *A. lyrata* genome, or a higher EBN (40). In this scenario, we therefore assign an EBN of 1 to the diploid *A. lyrata* and 2 to the diploid *A. arenosa*. In reciprocal diploid interspecies crosses, this produces a deviation from the required 2:1 EBN ratio in the endosperm (4:1 and 2:2 maternal: paternal ratio), leading to hybrid seed inviability. The EBN hypothesis predicts that the effective ploidy in the endosperm increases with the ploidy of the species. Increasing the ploidy of a low EBN species matches its effective ploidy with a higher EBN species, and thus allows the production of viable hybrid seeds (21, 25). Aligned with our observations, an EBN expectation increased to 2 in both natural and synthetic tetraploid *A. lyrata* restores endosperm cellularization and produces viable seeds when crossed with the diploid *A. arenosa* (EBN = 2). Furthermore, reciprocal crosses with tetraploid *A. arenosa* (EBN = 4) and diploid *A. lyrata* (EBN = 1) increases the severity of hybrid defects compared with diploid hybrids, because the difference of EBN between the two species is even higher (8:1 and 1:2 maternal: paternal ratio).

In the tetraploid species, the same difference in EBN as between diploids is expected (EBN = 2 and 4 for tetraploid *A. lyrata* and *A. arenosa*, respectively). As the phenotype of the synthetic tetraploids is similar to the seed phenotype obtained from the diploid × diploid hybrids, suggesting that the synthetic tetraploid *A. arenosa* has indeed a higher EBN than the synthetic tetraploid *A. lyrata*. The molecular mechanism underlying the EBN remains to be discovered. To this end, deregulated imprinted genes underpin endosperm-based hybridization barriers (46–48), and are thus likely to form the molecular basis of the EBN. Consistent with this idea, *Solanum* species with different EBNS exhibit different genomic imprinting patterns (48). Additional studies are needed to determine whether this difference not only correlates with but is indeed causative for the EBN.

Altogether, we propose the following scenario: facilitated by the autotetraploidization of *A. lyrata*, previously reproductively isolated *A. lyrata* and *A. arenosa* became able to form viable hybrid seeds. The initial hybridization event(s) took place between autotetraploid *A. lyrata* and diploid *A. arenosa*. This conclusion is in agreement with the pattern of gene flow between *A. lyrata* and *A. arenosa* that only involves the tetraploid *A. lyrata*, whereas the diploid *A. lyrata* is still reproductively isolated from *A. arenosa* (15, 16). Following the initial hybridization, introgression of 1.3% loci into *A. lyrata* allowed its compatibility with tetraploid *A. arenosa*, as supported by the viability of hybrid seeds between natural tetraploids, explaining the ongoing interspecific gene flow observed at the tetraploid level (15, 16). The bypass of hybridization barriers by polyploidization has been shown in artificial hybridization experiments (21, 25) but, to our knowledge, it has never been documented in wild populations. Polyploidization may thus play a prominent role in plant evolution, not only by mere reproductive isolation (57), but also by weakening hybridization barriers between related species.

**Materials and Methods**

**Plant Material and Growth Conditions.** All accessions used in this study originate from wild populations located in Central Europe, previously described by Jørgensen et al. (16) (see Fig. S1 and Table S2 for details). Diploid F1 hybrids were obtained from the 2x*A. lyrata* population MJ09-11 and 2x*A. arenosa* populations MJ09-1 and MJ09-5. For parental backcrosses of hybrids, *A. lyrata* MJ09-11 and *A. arenosa* MJ09-1 and MJ09-5 were used. Two to six F1 hybrid individuals per cross type were backcrossed to the respective parental species. All seeds were surface-sterilized using 5% (vol/vol) sodium hypochloride solution under the fume hood. After sterilization, seeds were plated on Petri dishes with Murashige and Skoog Medium [MS25] and solidified with 1% bactoagar. After stratification for 3 wk in the dark at 4 °C, germinated seedlings were grown in a growth room under long-day photoperiod (16-h light and 8-h darkness) at 22 °C temperature and a light intensity of 100 µE. Seedlings at the four-to-six leaf stage were transferred to soil and plants were grown in plant growth chambers at 60% humidity and days of 16-h light at 10 °C and 21 °C.

Lafon-Placette et al.
8-h darkness at 18 °C. After 3 wk, rosette-stage plants were transferred to winter conditions (10-h light and 14-h darkness at 8 °C) for 8–10 wk.

Plant Crosses. For all crosses, designated female partners were emasculated, and the pistils were hand-pollinated 2 to 3 d after emasculation. At least three biological replicates were made for each cross unless specified otherwise. Each biological replicate consisted of two or more siliques unless otherwise specified.

Scoring of Seed Phenotypes. Seeds were arranged on white plastic dishes and imaged using a Nikon D90 camera and Leica Z16apoA microscope. Seed size was measured by converting images to black and white using the “threshold” and “Analyze Particles” functions in ImageJ (https://imagej.nih.gov/ij/). Seed weight of 30–150 seeds per replicate was determined with an Ohaus GA1110 or Mettler AT200 scale and the total weight was divided by the number of seeds. Hybrid seeds that weighed less than the minimum range of the scale were given the value 0.0001 g per 50 seeds (smallest scorable weight). Noncollapsed seeds were scored from three replicates of imaged seed pools and characterized as brown and plump with variable size. Collapsed seeds were black or black and shriveled, brown and shriveled or not round, or very small seeds (less than 0.2 mm²).

Germination Assays. One month after harvesting, seeds were surface-sterilized described and plated on MS2 media without sucrose. To break seed dormancy, plates were then kept at 4 °C for 6 wk and put to normal growth conditions (see above) for 1 mo. Germination rates were assessed by counting seeds with ruptured seed coat and protruding radicles.

In vitro Embryo Rescue. Hybrid embryos derived from seeds of crosses of A. lyrata × A. arenosa were rescued at 27 DAP by in vitro cultivation. After a short incubation of siliques in 70% (vol/vol) ethanol, the embryos were isolated by dissection using hypodermic needles and placed on MS media containing 2% (vol/vol) sucrose. Plates were incubated in a light chamber. Surviving seedlings were transferred to soil after 14 d.

Generation of Synthetic 4x A. lyrata and 4x A. arenosa. To obtain synthetic tetraploid plants from the diploid A. lyrata and A. arenosa, 10 µL of 0.25% colchicine solution was applied on the shoot apical meristem of 1-mo-old hybrid embryos derived from seeds of crosses of A. lyrata × A. arenosa. A total of 20 seeds per biological replicate (three biological replicates) were sampled at 9 and 15 DAP for investigation of endosperm cellularization.

Microscopy. Clearing analysis was performed as previously described (59, 60) and imaged using a Zeiss axioplan Imaging2 microscope system equipped with Nomarski optics and cooled LCD imaging facilities.

Endosperm nuclei counts were conducted on cleared seeds at defined time points. Seeds were fixed in 9:1 ETOH:Acetic acid. After 1-d incubation in chloral hydrate solution [1:8 glycerol/chloral hydrate/water (vol/vol)] at 4 °C, a stack of pictures were taken throughout the whole seed and endosperm nuclei were subsequently counted in each picture using ImageJ cell counter (https://imagej.nih.gov/ij/).

A total of 20 seeds per biological replicate (three biological replicates) were sampled at 9 and 15 DAP for investigation of endosperm cellularization. Preparation of seeds was performed using Feulgen staining as described by Braselton et al. (61). Imaging of optical sections of the endosperm was performed using a multiphoton Zeiss LSN NLO microscope with an excitation wavelength of 770 nm and emission from 518 nm and onwards or an Olympus Fluoview 1000 Confocal Laser Scanning microscope (BX61W) with an excitation of 488 nm and emission from 500 to 600 nm.

Statistical Analysis. Seed size and weight of hybrid seeds were compared with the midparent value (42) using Student’s t-test. The endosperm proliferation rate was calculated as the ratio of nuclei number between two time points. Finally, χ² test was used to test for differences from expected frequencies according to Mendelian segregation.

Acknowledgments. We thank Roswitha Schmickl for fruitful discussion. This research was supported by a Funding Scheme for Independent Projects (FIPRO) Grant 214052/F20 from the Norwegian Research Council (to P.E.G.) and a European Research Council Starting Independent Researcher grant (to C.K.). C.L.P. was supported by a grant from the Nilsson-Ehle Foundation. K.S.H and I.M.J were supported by the Norwegian Ministry of Education and the Centre for Epigenetics, Development and Evolution.

References

Fig. S1. Locations of populations used in this study. Red and blue symbols represent populations of diploid and tetraploid Arabidopsis arenosa and diploid and tetraploid Arabidopsis lyrata, respectively. For more information, see Table S2. The map was modified from Google Maps.
Fig. S2. Early seed development in diploid A. lyrata, A. arenosa, and reciprocal hybrids. The microscopy pictures show cleared seeds where the embryo and endosperm nuclei are visible. The directionality of the cross is indicated seed parent (♀) × pollen donor (♂). 2A: 2x A. arenosa; 2L: 2x A. lyrata. (Scale bars, 50 μm.)
Fig. S3. Viable A. lyrata × A. arenosa hybrid plants obtained from in vitro embryo rescue. (A) Three-week-old A. lyrata × A. arenosa seedlings grown on MS plate. (B) Fully grown A. lyrata × A. arenosa plant transferred to soil. (C) Fully grown A. arenosa plant. (D) Fully grown A. lyrata plant.

Fig. S4. Early embryo development in diploid and tetraploid A. lyrata, A. arenosa, and reciprocal hybrids. The microscopy pictures show cleared seeds where the embryo and endosperm nuclei are visible. Embryo stages depicted are representative for the average embryo stage observed. The directionality of the cross is indicated seed parent (♀) × pollen donor (♂). 2A: 2x A. arenosa; 2L: 2x A. lyrata; 4A: 4x A. arenosa; 4L: 4x A. lyrata. (Scale bars, 20 μm.)
Fig. S5. Synthetic tetraploid *A. lyrata* restores hybrid seed viability similar as natural tetraploids. (A) Pictures of seeds obtained from controlled crosses between synthetic 4x *A. lyrata* and diploid *A. arenosa*. The directionality of the cross is indicated seed parent (♀) × pollen donor (♂). (Scale bars, 1 mm.) (B) Seed viability of interploidy interspecific crosses, as measured by the rate of noncollapsed and germinated seeds. (C) Seed size for indicated crosses. (D) Seed weight for indicated crosses. Error bars represent SD between cross replicates (>2 replicates for each sample). 2A: 2x *A. arenosa*; 4Lsynth: synthetic 4x *A. lyrata*.

Fig. S6. Seventeen DAP endosperm cellularization phenotype (in 2L4A and 4L4A). Feulgen-stained seeds of *A. lyrata*, *A. arenosa* reciprocal hybrids at 17 DAP. (A and C) Embryo phenotype. (B and D) Endosperm cellularization phenotype. Images shown in this figure are representative for a total number of observed seeds of at least 20 per biological replicate. (Scale bar, 50 μm.) 4A: 4x *A. arenosa*; 4L: 4x *A. lyrata*; 2L: 2x *A. lyrata*. 
Fig. S7. Hybrid seeds between synthetic tetraploid *A. lyrata* and *A. arenosa* resemble the hybrids between diploids, not between natural tetraploids. (A) Images of seeds obtained from controlled crosses between natural tetraploid, diploid and synthetic tetraploid *A. arenosa* and *A. lyrata*. The directionality of the cross is indicated seed parent (♀) × pollen donor (♂). (Scale bars, 1 mm.) (B) Seed viability of interspecific crosses, as measured by the rate of noncollapsed and germinated seeds. (C) Seed size for indicated crosses. (D) Seed weight for indicated crosses. Error bars represent SD between cross replicates (>2 replicates for each sample). 2A: 2x *A. arenosa*; 2L: 2x *A. lyrata*; 4A: natural 4x *A. arenosa*; 4L: natural 4x *A. lyrata*; 4Asynth: synthetic 4x *A. arenosa*; 4Lsynth: synthetic 4x *A. lyrata*.
Fig. S8. Ploidy level assessment of the synthetic 4x A. lyrata and 4x A. arenosa. Ploidy levels were estimated by measuring the relative nuclear DNA content of each sample by flow cytometry. The A. lyrata and A. arenosa diploid controls are shown in A and B. On top of nuclear DNA, content matching with diploid cells (2C), tetraploid (4C), and octoploid (8C) cells can be seen because of different mitotic stages and endoreduplication. The relative nuclear DNA content of colchicine treated positive stems (4x) is shown for A. lyrata (C) and A. arenosa (D). In these samples, as in diploids, an octoploid level (8C) is seen because of cells in G2 mitotic phases and endoreduplication, but tetraploidy (4C) is the lowest ploidy level observable. Progenies obtained from synthetic 4x A. lyrata × 2x A. arenosa (E) and 2x A. arenosa × synthetic 4x A. lyrata (F) are triploid as shown by the intermediate relative nuclear DNA content between diploids and tetraploids (3C, 6C, and 12C). Progenies obtained from synthetic 4x A. arenosa × synthetic 4x A. lyrata (G) are tetraploid, as their relative nuclear DNA content was similar to the tetraploid parents. For each biological sample, n = 5 replicates. The directionality of the cross is indicated seed parent () × pollen donor (_). 2A: 2x A. arenosa; 4L: 2x A. lyrata; 4A_syn: synthetic 4x A. arenosa; 4L_syn: synthetic 4x A. lyrata.
Table S1. Origin of parents used to produce F1 hybrids

<table>
<thead>
<tr>
<th>Type</th>
<th>Seed parent</th>
<th>Pollen donor</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1A</td>
<td>A. lyrata pop. MJ09-11</td>
<td>A. arenosa pop. MJ09-1</td>
</tr>
<tr>
<td>F1B</td>
<td>A. lyrata pop. MJ09-11</td>
<td>A. arenosa pop. MJ09-5</td>
</tr>
<tr>
<td>Backcrossed A. lyrata</td>
<td>A. lyrata pop. MJ09-11</td>
<td>A. lyrata pop. MJ09-11</td>
</tr>
<tr>
<td>Backcrossed A. arenosa</td>
<td>A. arenosa pop. MJ09-1</td>
<td>A. arenosa pop. MJ09-1</td>
</tr>
</tbody>
</table>

The table also indicates the origin of backcrossed parents.

Table S2. Sampling of A. arenosa and A. lyrata included in this study

<table>
<thead>
<tr>
<th>Collection no.</th>
<th>Taxon</th>
<th>Ploidal level</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Jørgensen et al. (16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MJ09-1</td>
<td>A. arenosa</td>
<td>2x</td>
<td>SVK: Vysoké Tatry; Prešovský kraj; Belianske Tatry; Zadné Med'odoly Valley; Kopské Sedlo</td>
<td>N 49.2569</td>
<td>E 20.1431</td>
<td>a2_SVK1</td>
</tr>
<tr>
<td>MJ09-8</td>
<td>A. lyrata</td>
<td>4x</td>
<td>AUT: Lower Austria; Dunkelstein Forest; Wachau; N Bachamsdorf</td>
<td>N 48.4000</td>
<td>E 15.5377</td>
<td>I4_AUT3</td>
</tr>
<tr>
<td>MJ09-9</td>
<td>A. arenosa</td>
<td>4x</td>
<td>AUT: Lower Austria; Eastern Alps; SSW St. Aegyd am Neuwalde; Kernhof</td>
<td>N 47.8178</td>
<td>E 15.5350</td>
<td>a4_AUT1</td>
</tr>
<tr>
<td>MJ09-11</td>
<td>A. lyrata</td>
<td>2x</td>
<td>AUT: Lower Austria; street from Pernitz to Pottenstein</td>
<td>N 47.9190</td>
<td>E 15.9755</td>
<td>I2_AUT1</td>
</tr>
<tr>
<td>MJ09-12</td>
<td>A. lyrata</td>
<td>2x</td>
<td>AUT: Lower Austria; S Vienna; Bad Vöslau; rocks near Vöslauer Hütte</td>
<td>N 47.9798</td>
<td>E 16.1637</td>
<td>I2_AUT2</td>
</tr>
</tbody>
</table>

Country names are abbreviated as follows: AUT, Austria; CZE, Czech Republic; GER, Germany; SVK, Slovakia. Collected by Marte H. Jørgensen July 23–28, 2009. Population number used in Jørgensen et al. (16) listed in the last column.