The paternally expressed genes in endosperm comfort the kinship theory to explain imprinting in *A. lyrata.*

Audrey Le Veve[1,2], Ömer İltaş[2,3], Julien Y. Dutheil[4], Clément Lafon Placette[2,*]

[1] IHPE, Univ Montpellier, CNRS, IFREMER, Univ Perpignan Via Domitia, Perpignan, France

[2] Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

[3] Max Planck Institute of Molecular Plant Physiology: Potsdam, Brandenburg, DE

[4] RG Molecular Systems Evolution, Department of Theoretical Biology, Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany

[*] to whom correspondence should be addressed; lafonplc@natur.cuni.cz

Abstract

Genomic imprinting is a form of gene regulation resulting in the unequal expression of the two parental alleles. Two main hypotheses are usually mentioned to explain it, the kinship and the sexual antagonism theories, together related to a conflict over resource allocation. Between the different common assumptions expected under these theories, directional selection on imprinted genes could help to discriminate the best theory to explain imprinting because the selection is expected, in regulatory or in coding regions respectively. But, so far, most researches that studied selection in imprinted genes have focused on autogamous species, whose conflict for ressource's allocation is expected to be low. Moreover, kinship theory is expected to promote intense antagonistic coevolution between imprinted genes, but this has never been tested so far.

To address these knowledge gaps, using phylogenetic and population genomics approaches, we searched for signals of selection on imprinted genes across the Brassicaceae family and in allogamous and autogamous populations of *Arabidopsis lyrata*, and we investigated signals of coevolution across imprinted genes. Overall, we found that non-imprinted endosperm genes presented signals of balancing selection, as expected in the case of unresolved intralocus conflict. Moreover, according to kinship theory, the PEGs exhibited signals of directional selection but not in coding regions and this signal was sensitive to mating systems. Finally, we found imprinted genes showing signals of

coevolution, but not related to specific sites, suggesting that the coevolution is driven by other interactions than protein, as predicted by kinship theory.

Key words: genomic imprinting, kinship theory, sexual antagonism theory, endosperm, coevolution.

Introduction

Genomic imprinting is a form of epigenetic gene regulation in flowering plants and mammals in which alleles of genes are expressed in a parent-of-origin dependent manner (Haig,1997; Wilkins, 2011). A variety of theories attempted to explain the existence of genomic imprinting (Haig,1997; Spencer and Clark, 2014; Patten et al,. 2014), but the most frequently invoked are the kinskip and sexual antagonism hypothesis (Patten et al,. 2014 for detailed review). In plants, genomic imprinting is initiated by distinct epigenetic landscapes between male and female gametophytes that are stably inherited after fertilisation (Batista and Köhler, 2020). Interestingly, many imprinted genes appear to be involved in resource accumulation and metabolism (Raissig et al., 2011).

Firstly verbally expressed in 1989 by Haig and Westoby, the kinship hypothesis appeared in a discussion of endosperm development and considered only the evolution of paternally-expressed growth factors and also the possible implication of maternally-expressed growth inhibitors in a review in 1991. According to the theory explained in this review, in species where the maternal parent nourishes the growing progeny from multiple males, the reproductive fitness of the mother is optimal when providing equal resources among all the offspring, while the father's fitness improves with the preferential allocation of maternal resources to his own offspring. The kinship theory posits that imprinting is selected to resolve this conflict over the allocation of resources to the embryo between the maternal and paternal genomes, between growth-limiting genes under maternal control and growth-promoting genes under paternal control (Halg and Graham,1991; Moore and Haig, 1991; Wilkins and Haig, 2001). This theory was then theoretically tested on different simple models of selection on expression levels of one or two genes implicating on resource allocation (Haig, 1997; Wilkins and Haig, 2001), and predicted that 1) an imprinted gene will select a different optimum levels of expression for paternally or maternally gene (Haig, 1997), 2) that paternal expressed gene will select expression that increase the ressource's allocation by offspring whereas the maternally expressed gene will select the opposite (Haig, 2000) and 3) that imprinting at the growth factor locus preceded the evolution of growth suppressor function at the second locus in case of antagonistic coevolution (Wilkins and Haig, 2001). However, these models did not precise the signal of selection expected on these genes and, more importantly, were based on one example, the imprinted genes *IGF* and *IGF2R* implicated on fetal growth and never adapted to endosperm development.

This point is essential to consider because the endosperm is the embryo's nourishing tissue in the seed and that the two parental genomes contribute to its formation. Thus, the

imprinted genes in this tissue could also be under sexual antagonist conflict (Patten et al,. 2014 for review). In this theory, asymmetric relatedness for maternally and paternally inherited alleles is not required and is based on an assumption of intralocus sexual conflict, which occurs when selection at a locus favors different alleles in males vs. females (Day and Bonduriansky, 2004; Bonduriansky, 2007). In the two theories, the hypothesis required 1) a preferential expression on endosperm, 2) a polyandre system, and 3) the resources for the progeny must be essentially distributed by one of the two parents. Moreover, the imprinting genes are expected to present important divergence compared to other endosperm genes, but more specifically in case of intralocus sexual conflict because non-imprinted genes are expected to be associated with a low divergence across species (Day and Bonduriansky, 2004; Kasimatis et al., 2017). One important difference argues between kinship and sexual antagonist theories is that the first is supposed to be apply on region of regulation whereas the second is supposed to be apply on coding region, but this prediction is hardly discussed because coding region could also modify expression efficiency (by bias of amino acid/codon; Patten et al,. 2014 for details). However, the signal of coevolution between maternally and paternally expressed genes is more associated with kinship theory and not mandatory under sexual antagonist theory. Moreover, the imprinted genes are expected to be associated to a variation of expression pattern in endosperm under kinship theory.

Consistent with the kinship theory, in *Arabidopsis thaliana*, an increase in maternal genome dosage (cross between a tetraploid mother and a diploid father) produces small seeds, whereas a reciprocal cross (increased paternal dosage) produces larger seeds (Scott et al., 1998). However, the effects on seed size probably result from an interaction between cytoplasmic effects, dosage effects and genomic imprinting.

Signatures of directional positive selection have been detected for some Paternally Expressed Genes (PEG) of the autogamous Brassicaceae species, *Capsella rubella* and *A. thaliana* (Hatorangan et al., 2016; Tuteja et al., 2019), that is consistent with the two hypothesis. However, these studies did not detect this signal on Maternally Expressed Genes (MEG). These results could be explained by technical limitations to detect selection on non-coding regions if the selection is more intensively applied on expression levels, as predicted under kinship theory, or by the labile polymorphism depending on the mating system predicted by the sexual antagonism theory. Moreover, positive selection on the imprinted *MEDEA* gene was detected only on the allogamous plant *Arabidopsis lyrata* but not in the autogamous sister species *A. thaliana* (Spillane and al., 2007; Miyake and al., 2009).

Another consequence of kinship theory expected at the molecular level is the coevolution of the pairs of imprinted genes having antagonistic effects on offspring growth (Wilkins and Haig, 2001). This coevolution could be explained by direct interactions between the antagonist imprinted genes (Wilkins and Haig, 2001) or indirectly, by interactions with other genes implicated in parental conflict (Willi, 2013). Interestingly, this antagonist coevolution is not mandatory under the sexual antagonism hypothesis (McVean and Hurst,. 1997; Patten et al,. 2014). Against the kinship theory assumption, no phylogenetic signal of coevolution was detected between the imprinted genes *IGF* and *IGF2R* implicated on fetal growth. However, to our knowledge, the signals of coevolution between endosperm imprinted genes were never investigated so far.

Thus, the signature of selection expected on imprinted genes and, more importantly, the contribution of the kinship and sexual antagonism theories on the evolution of the imprinted genes remains unclear in plants.

An important point about the empirical studies previously mentioned is that they concerned essentially autogamous species whereas the allogamy is an important initial condition for the two theories (Haig and Westoby, 1991; Haig, 2000; Day and Bonduriansky, 2004; Brandvain and Haig, 2005; Bonduriansky, 2007; Soliman and Coughlan, 2024 for review). The kinship theory requires different degrees of kinship across progenies induced by the multiple male partners whereas the sexual antagonist theory requires intense inter and intrasexual selection induced by the competition for reproduction between the males. The shift for autogamy could promote important changement on selection on imprinted genes (Soliman and Coughlan, 2024 for review), even if the prediction stayed unclear. Exactly, we must determine if the autogamy promotes important evolutionary change and directional selection on imprinted genes or a relaxation compared to non-imprinted genes.

To discriminate the theory more adapted to imprinted genes in endosperm, we analysed known imprinted genes of *A. lyrata* (Klosinska et al., 2016). This self-incompatible species is close to the autogamous model species *A. thaliana*. Exactly, we will test if the imprinted genes, compared to non-imprinted genes in the endosperm, are associated with : 1) a variation of expression pattern in endosperm, as predicted under kinship theory, 2) an increase of divergence and a decrease of polymorphism in coding regions as expected in case of directional selection, an assumption more expected in case of sexual antagonism excepted if the genes are also associated with 3) codon bias that could modify expression efficiency as predicted for the kinship theory and finally with signal of coevolution predicted for the kinship theory but not mandatory for the sexual antagonism. Population genomic methods may be used to detect coevolving coding sequences. These methods, as genetic

linkage for example, use the variation on allelic frequencies to estimate coevolution. The main advantage of these methods is the possibility of studying all polymorphic genes, whatever the distribution of this gene across species. This point is particularly important for imprinted genes that exhibited low conservation between *A. thaliana* and *C. rubella* (Hatorangan et al., 2016). However, these methods require polymorphism across samples and could be biased by different factors such as genetic distance, demographic events or genetic structuration. More holistic approaches are based on reconstructing the evolutionary history of candidate coevolving genes, assuming that their evolutionary history should be common if they coevolve and also likely diverge from the evolutionary history of the rest of the genome (Dutheil and Galtier, 2007). Most of these methods consist of comparisons of the phylogenies to detect co-occurring changes in a phylogeny.

To test these different assumptions, we will use phylogenetic data across 23 species of *Brassicaceae* and genomic data from nine different allogamous North American populations, and also nine other derived close populations that shift for autogamy several independent times (Ross-Ibarra et al., 2008). With autogamous populations, we will estimate the effect of shift of mating system on imprinted and non-imprinted endosperm genes. Our findings suggest that the sexual antagonism theory is the more adapted to explain the evolution of the imprinted genes and that the shift for autogamy promotes relaxation on these genes.

Results

Imprinted genes are often preferentially expressed in the endosperm.

In order to disentangle the impact of parental conflict and tissue specificity on selective processes, the genes specifically upregulated in the endosperm of *A. lyrata* were identified based on transcriptomic data obtained for 19,425 genes from 5 tissues (leaf, root, pollen, pistil, endosperm). A total of 2,732 genes were preferentially expressed in the endosperm. By comparison with the list of imprinted genes (table S1), we found that among the 27 MEG and 46 PEG found in *A. lyrata* (Klosinska et al., 2016), 15 MEG and 14 PEG were specifically upregulated in the endosperm. These proportions were significantly higher than those predicted by 10000 random resampling of all the genes studied (p value < 1^{e-04}). For the following analysis, PEG and MEG upregulated in the endosperm will be referred to as PEG and MEG (table S1), whereas the others PEG and MEG will not be considered and the 2,703 non-imprinted genes upregulated in endosperm have been named endosperm genes. Between the MEG and PEG upregulated in endosperm, respectively 6 and 7 were similarly imprinted in *A. thaliana* (table S1). A set of 100 randomly chosen non imprinted ubiquitous genes will be considered as control genes.

An important difference between the kinship and the sexual antagonism theories is that the first suppose selection on expression levels of imprinting genes. The expression of the PEG and MEG were non significantly different than in the non-imprinted endosperm genes (p-value KW test=0.80). However, the expression levels of PEG and MEG results from different numbers of gene's copies expressed (N= one and two respectively). Thus, if we compared the expression levels by gene's copy, we found a significant variations across group of gene (p-value KW test= 1.54^{e-04}), with a significant increase in PEG compared to MEG (p-value Dunn's test=0.02) and non-imprinted endosperm genes (p-value Dunn's test=3.28^{e-05}; Fig 1). The expression levels by gene copy between MEG and non-imprinted endosperm genes were not different (p-value Dunn's test=0.10).

The selection on expression of genes could also imply codon or amino acid bias. Thus, we researched for codon or amino acid bias on imprinted genes compared to endosperm genes. We found significant variations across imprinted and endosperm genes of TGA and CGT (p-values KW test = 0.03 and 0.02), and for the amino acid Thr (p-values KW test = 0.04). Exactly, the PEG are associated with an increase of proportion of CGT for the Arginine, whereas the MEG are associated with a decrease of proportion of TGA for the stop codon (table S2). Moreover, PEGs are associated with an excess of threonine.

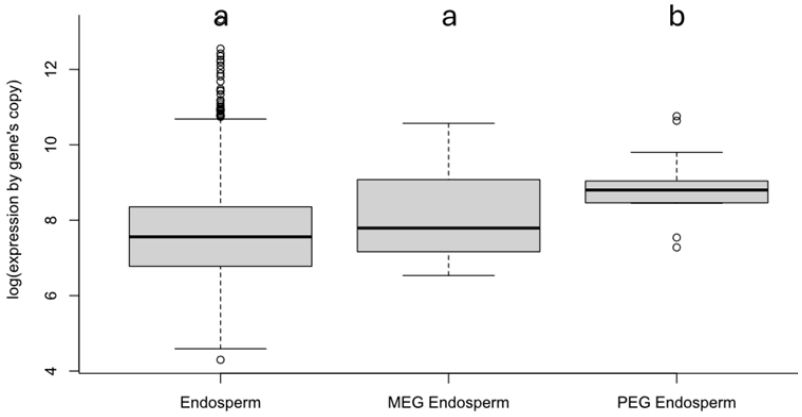

Figure 1: Expression by gene's copies expected to be expressed in endosperm for the imprinted genes compared to endosperm genes across the Brassicaceae . *boxplot of the variations of log of mean read count by gene's copies expected to be expressed in endosperm by group of genes. The number of gene's copies expected to be expressed is N=3, 2 and 1 for the non-imprinted genes, the MEGs and the PEGs respectively. The minuscule letters (a,b,c) signified the groups significantly different following Dunn's test (p-values < 0.05.).*

We compared the size, the coding sequences (CDS) content (%) and the GC content (%) of imprinted genes in *A. lyrata* with the endosperm and the control genes by Kruskal-Wallis tests because these statistics could impact the polymorphism. Only the size of the genes varied significantly across the group of genes (pvalue=1.63^{e-84}). The endosperm genes were significantly smaller than control genes (table S3).

The MEGs and PEGs present phylogenetic characteristics of directional selection in Brassicaceae.

We investigated the phylogenetic characteristics of imprinted genes in *A. lyrata* and twenty-one other Brassicaceae (table S4) compared to other groups of genes. First, we compared the number of species with homologous genes across the group of genes as a proxy for gene turnover (Fig 2; table S5). This number varied significantly across gene groups (p value KW=4.06^{e-15}), with a significant increase in endosperm genes compared to other genes (less gene turnover). However, the Robinson-Foulds (RF) distance based on comparison between the topology of each gene and the topology based on the endosperm genes found in all the Brassicaceae was not significantly different across the imprinted and endosperm genes group of genes (p-value Kruskal-Wallis = 0.6, Fig S1). Thus, this topology was used to estimate the branch lengths, a proxy for the evolution rate, for each gene with homologous genes found in at least three species (including *A. lyrata*). The length of all branches found in the imprinted genes varied significantly across groups of genes (Fig 2; p

value KW=0; table S5). This value in all the groups of genes were significantly higher than control genes, but this value was significantly lower for the MEGs and PEGs compared to endosperm genes. If we consider only the branch length of *A. lyrata*, the significant variations across groups of genes (p value KW= 1.55^{e-115}; table S5) were significant increases in endosperm genes and MEGs compared to control (table S5).

Finally, we estimated the variations of the D_{NS}/D_S between *A. lyrata* and the closer sister allogamous species, *A. halleri*. We expect an increase of this ratio under positive directional selection, and the opposite under negative selection. This ratio varied significantly across groups of genes (p-value Kruskal-Wallis = 0.02; table S5), and was significantly higher in the endosperm genes than in MEGs and control genes.

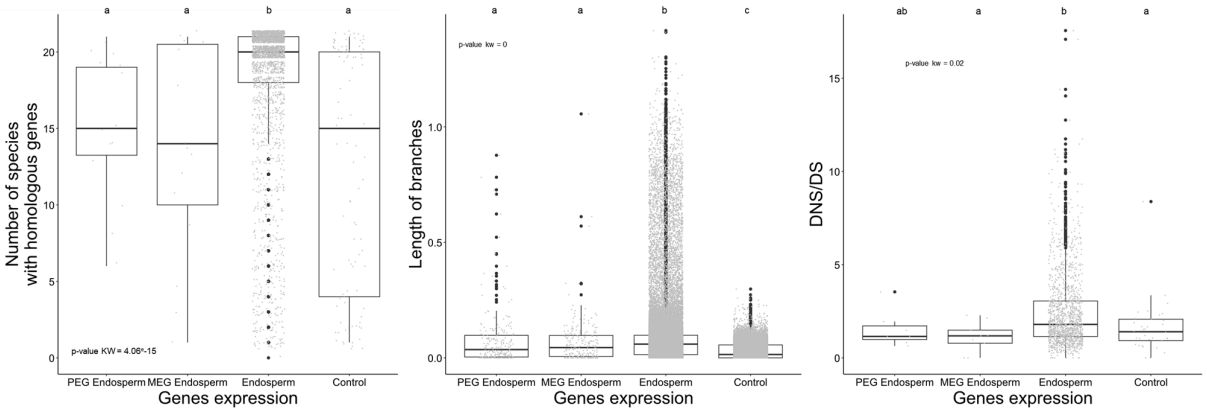

Figure 2: Phylogenetic characteristic of the imprinted genes compared to endosperm and control genes across the Brassicaceae . *Left: boxplot of the variations of the number of species with homologous genes by group of genes. Each grey point represents the value obtained for one gene. Middle: boxplot of the variations of the mean length branch by gene based on the phylogenetic tree obtained by HKY85 model based on the 1055 endosperm genes found in all species. Each grey point represents the value obtained for one gene and one species. Right: boxplot of the variations of the D_{NS}/D_S by gene. Each grey point represents the value obtained for one gene. The minuscule letters (a,b,c) signified the groups significantly different following Dunn's test (p-values < 0.05.).*

Signal of directional selection in population scale and related to the mating system was detected on PEGs

To estimate the effect of mating system and imprinting on polymorphism in populations, we compared Tajima's D and the π_{NS}/π_S of the imprinted genes with the endosperm and control genes in nine allogamous and nine autogamous populations (table S6). In a previous study,

an effect of the mating system on the seed size induced by pollen was reported for these populations (Willi, 2013), suggesting variations on intensity of parental conflict.

In the case of directional selection, we expect a reduction of the Tajima's D, whereas we expect an increase of the Tajima's D in case of balancing selection. However, this method requires some polymorphic positions. Thus, the genes without polymorphic sites in all the populations were removed before analysis. The proportion of endosperm genes conserved for the analysis increased significantly compared to control (binomial tests; table S7). In the allogamous populations, the Tajima's D varied significantly across the gene groups (p-value Kruskal-Wallis = 0.01) with a significant increase for endosperm genes compared to PEGs and control genes (Fig 3; table S8). In the autogamous populations, the Tajima's D varied significantly across the groups (p-value Kruskal-Wallis = 0.04; Fig 3; table S8) with a significant increase in PEGs and MEGs compared to endosperm genes. These different tendencies observed between allogamous and autogamous populations were explained by a significant decrease of Tajima's D in endosperm genes cumulated with a significant increase in PEGs in autogamous populations compared to allogamous populations (table S9).

In the case of directional selection on imprinted genes, we expect a reduction of the π_{NS}/π_S. However, this method requires polymorphism on synonymous sites. Thus, the genes with a π_S to zero in all the populations were removed before analysis. The proportion of endosperm conserved for the analysis increased significantly compared to control but the proportion of imprinted genes were intermediate to control and endosperm genes (binomial tests; table S7). This ratio varied significantly across the groups of genes in allogamous and autogamous (p-values Kruskal-Wallis = 0.02 and 2.44^{e-04} respectively). In the allogamous populations, the π_{NS}/π_S of the PEG increased significantly compared to the endosperm and control genes, and the ratio of control genes was similar than in endosperm genes (Fig S2; table S8), whereas in the autogamous populations, the π_{NS}/π_S of the PEG increased significantly only compared to the endosperm genes, and the ratio on control genes was significantly higher than in endosperm genes (Fig S2; table S8). The π_{NS}/π_S of endosperm genes decreased significantly in autogamous populations compared to allogamous populations (table S9), but did not vary for the other groups of gene with the mating system.

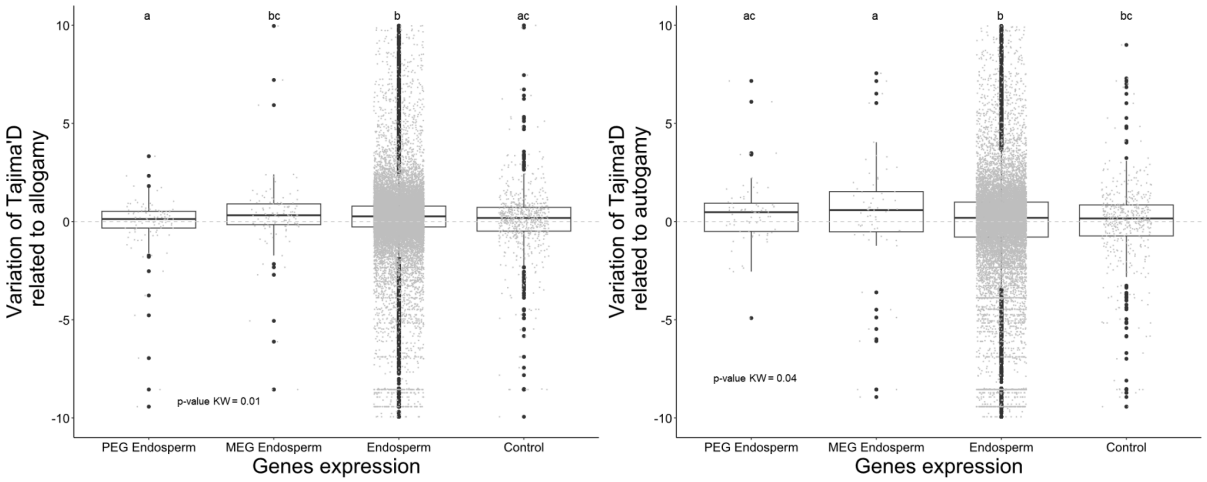

Figure 3: Variations of the Tajima's D in the allogamous and autogamous populations in the imprinted genes compared to endosperm and control genes. *boxplot of the variations of the Tajima's D by gene and population.The dashed grey lines represent the threshold value for neutrality (=0). The grey points represent the value of each gene in each population. The minuscule letters (a,b,c) signified the groups significantly different following Dunn's test (p-values < 0.05.).*

A limited number of imprinted genes presented intense signals of directional selection.

We searched for imprinted genes associated with strong signals of directional selection compared to endosperm genes by detecting imprinted genes with a lower mean of Tajima's D and/or mean π_{NS}/π_S in allogamous populations than 5% of these same values in the endosperm genes.

One MEG and one PEG were associated with a lower mean Tajima's D in allogamous populations (table S1). Based on homologous genes in *A. thaliana*, the PEG encodes for a chromatin remodelling protein implicated in resistance to oxidative stress tolerance of seed (Sujeeth et al., 2020), in seed germination (Dekkers et al., 2013) and in female gametophyte development (Wuest et al., 2010). No homologous gene was found for the MEG. Moreover, two MEG encoding a Dof-type zinc finger domain-containing protein implicated in seed coat formation (Zou et al., 2012) and one F-box domain containing protein that is regulated by non-CG DNA methylation were associated with a reduced mean π_{NS}/π_S in allogamous populations (table S1).

Finally, to determine if the proportion of candidate imprinted genes with signal of directional selection was significantly higher than in endosperm genes, we tested if this proportion was higher than 5% by binomial test for each statistic independently (Tajima's D and π_{NS}/π_S).

However, these proportions were never significantly higher than 5% (p values binomial tests >= 0.17).

Clear trend of coevolution between imprinted genes in Brassicaceae

The last important expectation about imprinting genes is that kinship theory promotes coevolution between genes favouring paternal or maternal fitness in an arms race scenario, whereas this tendency is not required for sexual antagonist theory. To test this assumption, different coevolution signals between PEG, MEG and the other endosperm genes were investigated. More precisely, we assessed whether some genes presented correlations based on 1) the π in the populations, 2) the length of branches or 3) Robinson-Foulds distance between pairs of topologies of the phylogenetic trees. First, the correlations of the π and the length of branches between each pair of PEG, MEG and endosperm genes were estimated by Pearson correlation. Then the top 1% negative and positive correlations of each distribution obtained for the pairs of endosperm genes were used as limits for significant correlations with imprinted genes. These values were -0.66 and 3.20 for the polymorphism and -0.64 and 0.89 for the length of the branches.

We estimated coevolution by correlation of π, and not correlation of π_{NS}/π_S, because the coevolution hypothesis is more related to kinship theory that is also expected to target non-coding regions. Based on π, 6 PEG and 9 MEG presented 12 significant correlations of polymorphism together. The number of correlations between PEG and MEG (e.g. 13) and the number of genes correlated (e.g. 15) were significantly higher than in 1000 random samplings of 29 endosperm genes (p-values= 4.5e-02 and <1e-03 respectively). Considering the endosperm genes significantly correlated to these MEG and PEG, and the imprinted genes correlated to these endosperm genes (e.g. indirect correlation), the complete network included 14 MEG, 13 PEG and 1643 endosperm genes (Fig 4; table S10). This network did not present enrichment on specific GO terms than expected in random samples of endosperm and imprinted genes of the same size (1,670 genes). One PEG encoding chloroplast localized protein required for embryonic development (Bryant et al,. 2010) was correlated with more genes than 99% of the endosperm genes (table S1 and S10).

For the length of the branches, 6 PEG and 9 MEG presented 16 significant correlations together. The number of correlations between PEG and MEG (*e.g.* 16) and the number of genes correlated (*e.g.* 15) were significantly higher than in 1000 random samplings of 29 endosperm genes (p-values = 8^{e-03} and 3^{e-03} respectively). Considering the endosperm genes significantly correlated to these MEG and PEG, and the imprinted genes correlated to these

endosperm genes (e.g. indirect correlation), the complete network included 14 MEG, 14 PEG and 1027 endosperm genes (Fig 4B, table S11). The complete network was significantly associated with the negative regulation of catalytic activity and syntaxin binding by GO (table S12), compared to GO expected in random samples of endosperm and imprinted genes of same size (1,055 genes). One MEG encoding an AGAMOUS-like protein expressed during proliferative endosperm development (Day et al., 2008) and implicated in female gametophyte, seed and early endosperm development (Bemer et al., 2010; Zhang et al., 2018) was correlated with more genes than 99% of the endosperm genes (table S1 and S11).

Finally, we aimed to detect gene pairs that displayed correlated deviation from the species topology. We first select genes whose genealogy deviated from the species tree, by removing the genes whose genealogy had a RF distance with the global topology of the endosperm genes was lower than the top 5% of the distribution obtained with endosperm genes (Fig S1; =0.71). Then, discriminate pairs of genes including imprinted genes with a null RF distance as pairs with a significant correlation on topology. Based on these criterias, the topology of one PEG (*AL6G33630*), a F-box protein, was significantly related to six endosperm genes (Fig 4C; table S13). This small network of genes was significantly associated with DNA-binding transcription factor activity by GO (table S12), compared to GO expected in random samples of seven endosperm and imprinted genes. Moreover, the PEG presented a significantly higher number of associations compared to 99% of non-imprinted endosperm genes (table S1 and S13).

Thus, within the possible pairs of imprinted genes, 29 including 22 PEG/MEG were directly associated in the networks based on π or on the length of branches. To detect coevolution between specific mutations within these pairs of imprinted genes, we used the clustering method of the pipeline CoMap package (Dutheil and Galtier, 2007). More precisely, we searched for groups compound by 2 to 10 sites in coevolution in each candidate pair of imprinted genes. To avoid false association, we compared the number of groups with two to ten coevolving sites found in the candidates independently with the number of groups with equivalent number of coevolving sites for all the possible pairs of imprinted genes never associated in the previous networks and found in three or more species. We never found significant increases in the number of groups detected in candidates compared to non candidate imprinted genes, whatever the number of coevolving sites in the group considered (table S14). Thus, these pairs of imprinted genes directly correlated were not associated with an increase on the number of coevolving sites.

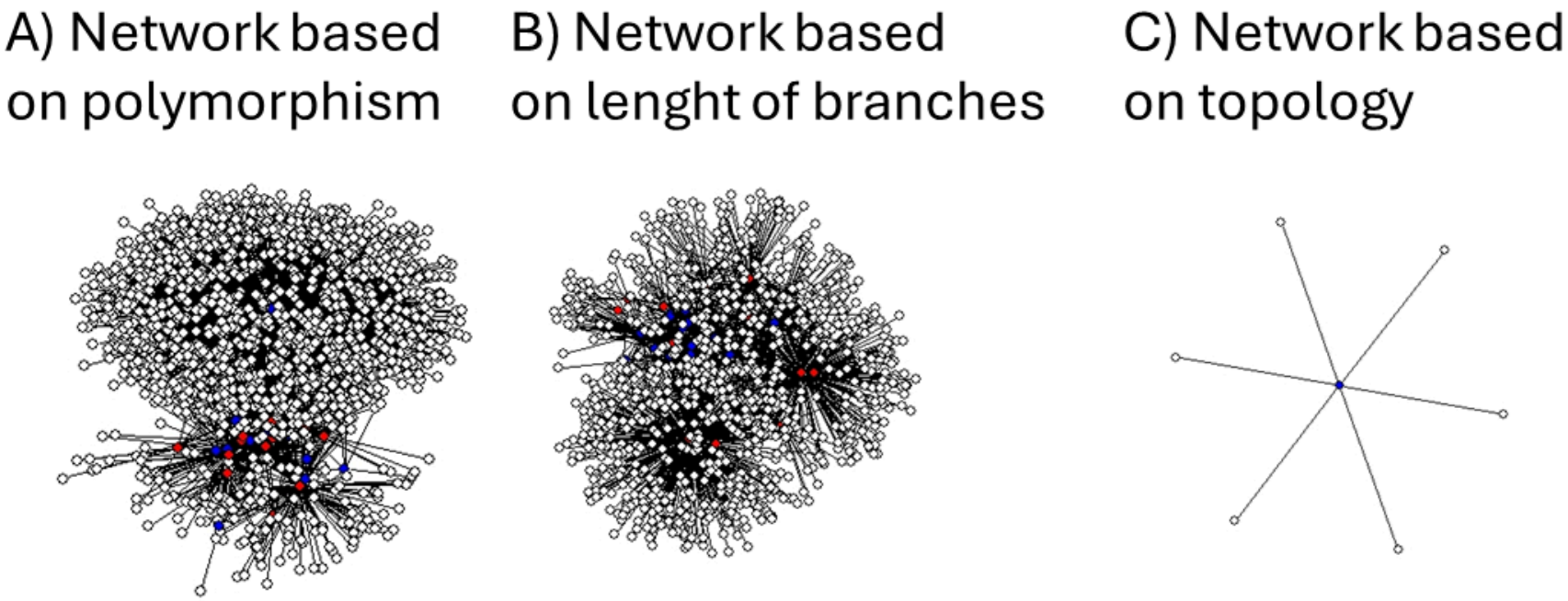

Figure 4: Networks of the imprinted and endosperm genes based on A) polymorphism in population , B) length of branches and C) phylogenetic topologies across the Brassicaceae. *A) Network based on the Pearson's correlations of π found on genes in the populations. B) Network based on the Pearson's correlations of the length of branches for each gene estimated with the global topology. C) Networks based on the RF distance of topologies. The black and grey lines represent the negative and positive correlations respectively. Blue circles: PEG. Red circles: MEG. White circles: non-imprinted endosperm genes.*

Discussion

Genomic imprinting is a form of epigenetic gene regulation in which alleles of genes are expressed in a parent-of-origin dependent manner (Haig, 1997; Wilkins, 2011). The most frequently invoked selective explanations for the evolution of imprinting is the kinship and the sexual antagonism hypothesis (Patten et al,. 2014 for detailed review). Under the two theories, the imprinted genes must be implicated on ressource's allocations and in a system with multiple mating (Haig, 2000). Moreover, in the two hypotheses, the imprinted genes are supposed to solve the conflict between offspring and mother or mother and father respectively, promoting common phylogenetic signatures of directional selection on these genes compared to non-imprinted endosperm genes because (Day and Bonduriansky, 2004; Kasimatis et al., 2017). However, contrary to sexual antagonism, this signature is not expected on coding regions under kinship theory, which implicates more regulatory regions. This detection of directional selection on imprinted genes is also expected in population scale, specifically with Tajima's D that considered all sites, whereas the π_{NS}/π_S is expected efficient only in case of selection on coding regions, so under sexual antagonism theory. However, the lability of this signal with the mating system is expected higher in population under sexual antagonism theory, even if the expectations on genes was not clearly defined. Moreover, the kinship theory predicts the coevolution between imprinted genes (Wilkins and Haig, 2001; Willi, 2013), that is not mandatory under the sexual antagonism hypothesis. We summarized all the population genomics and phylogenetics assumptions investigated in our study in populations of *A. lyrata* in table 1.

Signals of balancing selection promoted by kinship theory in endosperm genes.

Without genomic imprinting, the genes under conflict for ressource's allocation are expected to be under balancing selection, a footprint of intralocus sexual conflict (Day & Bonduriansky, 2004; Sylvestre et al., 2023). Consistently, for the non-imprinted endosperm genes, we detected all the expected phylogenetic signals of balancing selection: compared to control genes, these genes were more found across the Brassicaceae, suggesting low turn over, and these genes presented longer branches (Edwards, 2009). Moreover, these genes were associated with an increase of Tajima's D in the allogamous populations. However, all these phylogenetic and population genomics signals disappeared when we considered coding regions (e.g. D_{NS}/D_S and π_{NS}/π_S). These observations could be explained if the selection happens more on regulatory regions than in coding regions, as predicted by kinship theory, even if it could be also explained by a decrease of statistical power by the reduced number of sites considered.

Table 1: Summary of the expectations on phylogenetics characteristics and polymorphism on non-imprinted and imprinted endosperm genes under kinship or sexual antagonist theories. The expectations on non-imprinted endosperm genes are relative to control genes. The expectations on imprinted endosperm genes are relative to non-imprinted endosperm genes.

	Gene	Kinship theory	Sexual antagonism theory
Expression levels	Imprinted endosperm	Upregulated	No
Codon and amino acid bias	Imprinted endosperm	Possible	No
Number of homologous genes across species	Non-imprinted endosperm	Increase (less turn over with intralocus conflict)	Increase (less turn over with intralocus conflict)
	Imprinted endosperm	Decrease (more turn over by high evolutionary rate)	Decrease (more turn over by high evolutionary rate)
Length of branches	Non-imprinted endosperm	Increase by the intralocus conflict	Increase by the evolutionary rates increases
	Imprinted endosperm	Increase by evolutionary rates increases	Increase by the intralocus conflict
DNS/DS	Non-imprinted endosperm	No variation (selection not in coding regions)	No variation by intralocus conflict
	Imprinted endosperm	No variation (selection not in coding regions)	Increase or decrease if positive or negative selection respectively
Tajima's D in allogamous population	Non-imprinted endosperm	Increase (intralocus conflict)	Increase by intralocus conflict
	Imprinted endosperm	Decrease (directional selection)	Decrease by directional selection
πNS/πS in allogamous population	Non-imprinted endosperm	No variation (selection not in coding regions)	No variation by intralocus conflict
	Imprinted endosperm	No variation (selection not in coding regions)	Decrease by directional selection
Variation of πNS/πS and Tajima's D with mating system	Non-imprinted endosperm	Yes (decrease with selfing)	Yes (decrease with selfing)
	Imprinted endosperm	No (already fixed) or Yes (increase with relaxed selection)	No (already fixed) or Yes (increase with relaxed selection)
Coevolution between genes	Non-imprinted endosperm	Yes	Not necessarily
	Imprinted endosperm	Yes	Not necessarily

The conflict for ressource's allocation is expected only in case of multiple mating (Haig, 2000), and the autogamy, by "resolving" the conflict, would allow directional selection on genes implicated in resource allocation. Consistent with our expectations, in the autogamous populations, the Tajima's D, but also the π_{NS}/π_S in endosperm genes compared to allogamous populations, suggesting that the process of selection on these genes is sensitive to mating type. These observations suggested also that the effect of mating system shift on conflict in endosperm genes is detectable in autogamous populations, despite that the transitions to selfing of our populations are recents and partial (Ross-Ibarra et al., 2008; Foxe et al., 2010; Hough et al., 2013; Willi et al., 2018). Thus, in autogamous populations, the relaxation of conflict for ressource's allocation and the balancing selection it exerts on endosperm genes may allow directional selection on these genes.

Phylogenetic signals of high evolutionary rates on imprinted genes.

The two theories predict that many genes under imprinting should affect growth (Haig, 2000; Patten et al,. 2014). Thus, whatever the hypothesis considered, the imprinted genes must be preferentially expressed on endosperm tissue. Consistent with the two hypotheses, the imprinted genes were associated with a significant excess of genes preferentially expressed in the endosperm. However, over half of the imprinted genes were also upregulated in one or more other tissues. We assumed that this differential pattern of expression across imprinted genes could affect selection on these genes, because the genes preferentially expressed in other tissues than endosperm could be under different selection processes, not only related to ressource's allocations, and we did not consider them in the rest of the analysis.

In *C. rubella* and *A. thaliana*, imprinted genes presented a rapid evolution compared to the rest of the genome (Hatorangan et al., 2016; Tuteja et al., 2019). Consistent with these studies, based on phylogenetic approaches, we found that the PEG and the MEG present a higher length of branches across Brassicaceae compared to control genes, translating a fast evolutionary rate. However, these branches were also smaller than in endosperm genes. This observation could be easily explained if intralocus conflict for resources in endosperm genes promoted an important increase of length of branches by balancing selection (Kasimatis et al., 2017; Llaurens et al., 2017). Interestingly, the imprinted genes exhibited also less conservation of homologous genes across species compared to endosperm genes, which could be explained by a higher turnover induced by the high evolutionary rate. However, imprinting is barely conserved across *A. lyrata* and *A. thaliana* (less than the half in our study) and more generally between species (Hatorangan et al., 2016), so the fast evolution of these genes across Brassicaceae may not be related to their imprinting. Notably,

this tendency to fast evolution was conserved for MEG when we considered only the *A. lyrata* species, where these genes are imprinted.

PEGs are associated with signals expected under kinship theory.

An important prediction of the kinship theory that we don't expect with sexual antagonism theory is that imprint modifies the expression pattern of the genes. According to this theory, the PEGs are associated with an increase of expression by gene copy that permits to obtain equivalent expression for these genes than other endosperm genes, but we did not detect this effect on expression for the MEGs. Moreover, even if we expected selection preferentially on the regulatory region under kinship theory (Patten et al,. 2014), we could also expect selection on coding regions because the PEGs and MEGs are associated with a codons bias compared to endosperm genes, cumulated with an amino acid bias for the PEGs.

The kinship and sexual antagonism theories are together expected to promote directional selection on imprinted genes compared to other genes under intralocus conflict by the fixation or the loss of the advantageous or disadvantageous allele for the parent that expresses the genes (Wilkins and Haig, 2001; Mills and Moore, 2004; Day and Bonduriansky, 2004; Kasimatis et al., 2017). However, this effect on divergence is expected to be less detectable by D_{NS}/D_S because kinship theory is expected to modify essentially non coding regions. According to this assumption, we did not detect variations of this ratio on PEG, whereas we detected global effects on evolutionary rate. Moreover, in the allogamous population of *A. lyrata*, the PEGs were associated with a reduction of Tajima's D compared to endosperm genes, as expected in case of directional selection (Haig, 2014). However, this signal of directional selection disappeared for the π_{NS}/π_S, as expected if selection is applied on non-coding regions. More interestingly, the π_{NS}/π_S increase in PEGs, that could be explained by a relaxation of selection on coding regions. However, this relaxation of selection could be also explained by the recent colonisation of *A. lyrata* of North America from Europe (Ross-Ibarra et al,. 2008; Otto et al., 2015; Willi et al., 2018), that could change the environmental pressures (biotic and abiotic) but also could decrease the efficiency of sexual selection by the general reduction of polymorphism and the increase of genetic load observed in these populations (Willi et al., 2018). To test this hypothesis, it would be required to compare the signal of selection in endosperm and imprinted genes in allogamous populations with different levels of polymorphism in future studies.

Contrary to PEGs, the MEGs exhibited a clear decrease of D_{NS}/D_S compared to endosperm genes, translating negative directional selection (or purged). However, this signal was not

detected in the population scale. However, three of the four imprinted genes that exhibited the more intense signal of directional selection (e.g. Tajima's D and) were MEGs.

It is also important to emphasise that the list of imprinted genes used was from a previous study on *A. lyrata* lines (Klosinska et al., 2016) and not directly from the populations studied. Indeed, we assumed that the imprinting was conserved across the species. However, the set of imprinted genes largely varies even within a species such as in *A. thaliana* (Gehring et al., 2011; Wolff et al., 2011; Hsieh et al., 2011; Pignatta et al., 2014). It is therefore possible that the "imprinted genes" studied are actually not imprinted in our populations, and thus are under different selective pressure than that expected if they were imprinted. A population-specific survey of genomic imprinting in future studies is required to solve this issue.

Imprinted genes present a strong signal of coevolution on polymorphism.

Another expected consequence of parental conflict is the coevolution of imprinted genes having antagonistic effects on offspring growth by direct or indirect interactions (Wilkins and Haig, 2001; Mills, and Moore, 2004). To detect coevolution between imprinted and endosperm genes, population genomic and phylogenetic methods were used (Dutheil and Galtier, 2007). For the first time, we show that the majority of the PEG/MEG were associated directly or indirectly by signals of coevolution and the number of direct associations based on polymorphism found was significantly higher than expected by chance. Together, these imprinted genes are parts of coevolution networks bearing important functions for endosperm development. One of these networks, based on the correlation of branch lengths, showed enrichment for enzymatic activities related to cell wall formation and modification. In Arabidopsis, this process is a crucial component of endosperm development, namely cellularization, where the endosperm transitions from a syncytium to multicellular tissue (Hehenberger et al., 2012). This transition impacts seed size (Pignatta et al., 2018; Zhang et al., 2020), which is a likely object of conflict for ressource's allocation. Another coevolution network involving genes related to syntaxin binding protein, associated to membrane activity, essential for transport of molecules, and storage of protein in seed. Finally, the network based on topology showed enrichment in DNA-binding transcription factor activity, those having developmental roles such as seed development or tissue differentiation and those involved in regulation of metabolism, lipid and protein storage (Noguero et al., 2013). Further research may explore the functional significance of this coevolution signal. Interestingly, the coevolution between imprinted genes did not involve more site-by-site coevolution in our study, suggesting that protein-protein molecular interactions, often a cause of gene-gene coevolution (Green et al., 2021), was a limit effect

here. Thus, this coevolution may act on antagonistic genes yet not directly interact at the protein level, comforting again the kinship theory that implicates more regulatory regions and selection on dosage than on product (e.g. protein).

Autogamy modifies the signal of selection detected in imprinted genes compared to endosperm genes.

Conflict for ressource's allocation is expected to be relaxed in autogamous lineages compared to allogamous ones (Haig, 2000; Brandvain and Haig, 2005). This phenomenon explained that the imprinted *MEDEA* locus was associated with an increase of polymorphism in the autogamous species *A. thaliana* but not in the allogamous sister plant *A. lyrata* (Spillane et al., 2007; Miyake et al., 2009). However, the research of genetic and phylogenetic signals of selection induced by imprinting is usually on autogamous species. In our study, we found that the MEGs and more specifically the PEGs are falsely associated to an increase of polymorphism in autogamous populations compared to endosperm genes because the selection on PEGs but, more importantly, on the endosperm genes is drastically modify by mating system. Moreover, our observations were not in line with the positive selection detected on PEGs on the autogamous species *A. thaliana* and *C. rubella* (Hatorangan et al., 2016; Tuteja et al., 2019), and raised the question of the general effect of the mating system, and more specifically of the selfing rate, on the selection of imprinted genes in autogamous species or populations. Moreover, the absence of an effect of mating type on the polymorphism of the MEGs in our study suggests a limited effect of the parental conflict on selection on these genes.

Conclusions.

Here, we studied the phylogenetic and genomic signatures of selection on imprinted and endosperm genes in *A. lyrata*. By phylogenetic approaches, we found evidence for intense directional selection on imprinted genes expressed in endosperm. Moreover, the PEGs exhibited the signals of polymorphism related to directional selection in allogamous populations, but not if we considered the coding regions. This observation is according to kinship theory. This theory is also comforted by the intense signal of phylogenetics and population genomics coevolution, even if future studies are required to further investigate the cause for this coevolution. However, for the MEGs, despite that these genes exhibited phylogenetic signals of negative selection in coding regions, we did not detect any clear signals of selection favoring one of the two hypotheses tested to explain imprinting. Thus, kinship theory appears to be the best hypothesis to explain the evolution of imprinted genes in *A. lyrata* populations.

Materials and Methods

We expected that kinship and sexual antagonist theories promote different genomic signatures of selection on the imprinted genes than in the other endosperm genes. To detect signals of selection on imprinted genes and the implication of each theory on the selection, we developed an approach based on the comparison between the phylogenetic pattern and the variation of polymorphism of the 27 maternally expressed genes (MEG) and the 46 paternally expressed genes (PEG; Klosinska et al., 2016) clearly identified in the reference genome of *A. lyrata* (Hu et al., 2011) with the 2703 other genes upregulated in endosperm and a set of 100 randomly chosen ubiquitous genes.

Expression data

The RNAseq expression data for tissues, including leaf, root, pistil, and pollen, used in this study were acquired from populations of European and North American lineages of *A. lyrata* generated by Iltas et al. (2024). Additionally, endosperm RNAseq data from five outcrossed *A. lyrata* individuals were incorporated from a previous study by Klosinska et al. (2016). The endosperm data was integrated into the data processing pipeline as described by Iltas et al. (2024). Genes that showed significant upregulation (log2FC > 1 and adjusted P < 0.05) in all pairwise comparisons with other tissues were considered as endosperm-specific for further analysis. The RNAseq data for leaf, root, pistil, and pollen tissues can be found in the NCBI BioProject database under submission number PRJNA1067428. The endosperm RNAseq data is available under submission number GSE76076.

Genes clusters

We used the list of the 49 PEG and 35 MEG found in *A. lyrata* (Klosinska et al., 2016). We filtered the genes with conserved coordinates in the reference genome of *A. lyrata* V1.0.23 (Hu et al., 2011) to finally study 27 MEG and 46 PEG. The previous studies in *C. rubella* (Hatorangan et al., 2016) and in *A. thaliana* (Tuteja et al., 2019) suggested different signatures of selection between the PEG and the MEG. Thus, we distinguished the two types of imprinted genes in our study.

To discriminate the signature of selection related to endosperm, considered under conflict for ressources's allocation, to the signature associated with other possible selection process, we distinguished the imprinted genes upregulated specifically in the endosperm (14PEG/15MEG) with the other imprinted genes not specifically upregulated in the endosperm (32PEG/12MEG) using differential expression analysis. The significance of the number of PEG and MEG upregulated in endosperm between all the PEG and MEG were

tested by permutation tests based on 10,000 resamples of the genes with transcriptomic data. For all the following analyses, we compared only the imprinted genes upregulated in endosperm with the 2703 other genes upregulated in the endosperm (*e.g.* endosperm genes) and a set of 100 randomly chosen ubiquitous genes (*e.g.* control genes).

Then, we compared the normalised expression (read counts) of the non-imprinted endosperm genes with the expression of PEG and MEG upregulated in endosperm in European allogamous populations using Kruskal–Wallis test as implemented in R version 4.3.2, and completed by Dunn's test of multiple comparisons using rank sums implemented in the R package dunn.test (version 1.3.6). Exactly, we considered the exact expression, and the expression by the number of copies expressed (N), three for the non-imprinted genes, two for the MEG and one for the PEG.

Based on the sequences on the reference genome of *A. lyrata* (Hu et al., 2011), we tested if the size of the gene, the GC content and the CDS content (%) on the gene were significantly different across the group of genes. To test for significant differences across the group of genes, we performed Kruskal–Wallis test as implemented in R version 4.3.2. Then for the traits with significant variations across groups of genes, we performed Dunn's test of multiple comparisons using rank sums implemented in the R package dunn.test (version 1.3.6).

Estimation of codon and amino acid bias

Based on the sequences on the reference genome of *A. lyrata* (Hu et al., 2011) and a custom python code (https://github.com/leveveaudrey/codon-bias), we estimated the proportion of each amino acid on gene and the proportion of each possible codon for each amino acid.

Then, we tested if these proportions were significantly different across MEG, PEG and endosperm genes using Kruskal–Wallis test as implemented in R version 4.3.2, and completed by Dunn's test of multiple comparisons using rank sums implemented in the R package dunn.test (version 1.3.6).

Read mapping and variant calling

We worked on natural accessions from *A. lyrata* of North America represented by 18 samples of 25 individuals from allogamous and autogamous populations (table S6), using available pool-sequencing data (Willi et al., 2018).

Raw reads from Willi et al (2018) datasets were mapped onto the complete *A. lyrata* reference genome V1.0.23 (Hu et al., 2011) using Bowtie2 v2.4.1 (Langmead and Salzberg,

2012). File formats were then converted to BAM using samtools v1.3.1 (Li et al., 2009) and duplicated reads were removed with the MarkDuplicates program of picard-tools v1.119 (http://broadinstitute.github.io/picard). Biallelic SNPs in these regions were called using the Genome Analysis Toolkit v. 3.8 (GATK, DePristo et al., 2011) with the option GVCF and a quality score threshold of 60 using vcftool v0.1.15 (Danecek et al., 2011). We excluded sites with either less than 15 reads aligned. Sites covered by at least 15 reads but containing no SNP were considered as monomorphic. The final number of sites in each sample set is summarised in table S6.

Effect of mating-type on global population polymorphism

We decomposed the signal of selection across the imprinted genes into two elementary statistics: the Tajima's D and the ratio of average polymorphism between the non-synonymous and synonymous sites π_{NS}/π_S. For the allogamous and autogamous populations, we compared the values obtained for the imprinted genes with the values obtained across the endosperm genes and the control genes.

First, for each gene of each population, we estimated the π obtained for all positions of each gene using vcftools[37] with –site-pi option. If a position of the *A. lyrata* genome was covered but not polymorphic, the π was set to 0. Then, the 0-fold and 4-fold degenerate sites were identified and extracted from the reference genome and the gene annotation using the script NewAnnotateRef.py (Williamson et al., 2014). Because all mutations at 0-fold degenerate sites alter the sequence of the encoded protein, we assumed that these mutations are non-synonymous (NS). In contrast, mutations at the 4-fold degenerate sites never alter the encoded amino acid, so we assumed that these mutations are synonymous (S). For the sake of simplicity, we discarded mutations at 2- or 3-fold degenerate sites. Based on the π estimated for each site, we estimated, for each gene in each population, the mean π_{NS} and π_S, to finally estimate the π_{NS}/π_S. We removed the genes with a π_S to 0.

Secondly, for each gene of each population, we estimated the Tajima's D using vcftools (Danecek et al., 2011) with --TajimaD option. We removed the genes without polymorphic sites. To test for significant differences across the group of genes, for the allogamous and autogamous populations independently, we performed Kruskal–Wallis test as implemented in R version 4.3.2. Then for the statistics with significant variations across groups of genes, we performed Dunn's test of multiple comparisons using rank sums as implemented in the R package dunn.test (version 1.3.6).

Distribution of the genes across Brassicaceae

We researched if the imprinted genes present specific signals of selection compared to the endosperm and control genes in the phylogeny of 22 Brassicaceas. All the reference genomes were extracted from the Phytozome database (Goodstein et al., 2012; table S4). Thus, we extracted the sequences of the genes in the *A. lyrata* reference genome (Hu et al., 2011) using bedtools V2.30 (Quinlan et al., 2010). Then, we researched the best hit of these sequences with an e-value lower than 1^{e-5} on 22 other Brassicaceae by blast V2.10 (Altschul et al., 1990). If more of one hit were found for a gene in one specie, we filtered the best hit as the hit with: the higher length alignment, the higher % of identity, then the higher coverage on the gene of *A. lyrata* and the higher reciprocal coverage on the gene of the other species. Then we extracted the sequence of the best hit for each species of each gene.

Based on this data, to test for significant differences on the number of species with homologous genes across the group of genes, we performed Kruskal–Wallis test, as implemented in R version 4.3.2. Then, we performed Dunn's test of multiple comparisons using rank sums, with two-sided P values adjusted using the Bonferroni method as implemented in the R package dunn.test (version 1.3.6).

Divergence of the imprinted genes across Brassicaceae

The rates of nonsynonymous mutations (D_{NS}) and of synonymous mutations (D_S) is assumed to follow the neutral evolutionary process and the ratio D_{NS}/D_S is therefore approximate to the selective pressure on the protein product of a gene. An increase of the D_{NS}/D_S in genes signifies positive selection, while a decrease indicates negative selection. To test if imprinted genes display accelerated evolutionary rates compared with endosperm genes, using the annotation of the 0-fold and 4-fold degenerate sites in *A. lyrata*, we estimated the mean evolutionary rates D_{NS}/D_S for each gene between the reference genome of *A. lyrata* and the species with an homologue found across the 22 other Brassicaceae. Only the genes found in at least 3 species and with a D_S higher than 0 were considered.

To test for significant differences on the D_{NS}/D_S we performed Kruskal–Wallis test, as implemented in R version 4.3.2. Then, we performed Dunn's test of multiple comparisons using rank sums as implemented in the R package dunn.test (version 1.3.6).

Topology of the genes across Brassicaceae

For each gene, we computed the gene and all the best hits found in the different species in fasta files to do alignment with Muscle V3.8.155 (Edgar, 2004). Then, we build a phylogenetic tree following the HKY85 model for each gene with phyml V3.3 (Guindon et al., 2010). Moreover, we concatenated the alignment of the 1055 endosperm genes found in the

22 others Brassicaceae in a fasta file to build a phylogenetic tree without branch length of the endosperm using phyml V3.3 (Guindon et al., 2010). This tree was used as a proxy of the topology of the endosperm. Then, we estimated the Robinson-Foulds distance between the topology of this tree with the topology of the tree of each gene obtained previously using the R package TreeDist. Then, to test for significant differences on the RF distance across the group of genes, we performed Kruskal–Wallis test, as implemented in R version 4.3.2.

Variation of the length of branches across Brassicaceae

We researched if the imprinted genes present a variation of the length of branches compared to the endosperm and control genes. For each gene, we estimated the length of branches for each species with the topology of the phylogenetic tree obtained with the 1055 endosperm genes using phyml V3.3 (Guindon et al., 2010). We extracted the length of branches of each gene with the R package 'ape' version 5.7-1. If the gene was not present in a species, the value was replaced by "NA". To test for significant differences on the length of branches across the group of genes, we performed Kruskal–Wallis test, as implemented in R version 4.3.2. Then, we performed Dunn's test of multiple comparisons using rank sums as implemented in the R package dunn.test (version 1.3.6).

Genes in coevolution

To detect coevolving coding sequences, we used three different methods based on the following assumptions: the coevolution could create 1) related polymorphic patterns between genes in populations, 2) related evolutionary rate across species and/or 3) similarity in the topologies across species (Dutheil and Galtier, 2007).

To detect coevolution based on polymorphism in population and on evolutionary rates across species, we estimated Pearson's linear correlations between each pair of imprinted and endosperm genes for the mean π and the length of branches on each species. Based on these matrices of correlations, we extracted the correlations between pairs of endosperm genes to define the thresholds of extreme values in the distributions. Then, all pairs of genes included imprinted genes with correlations lower than the value at 1% of the distributions or higher than the value at 99% of the distributions were considered as significant. We used these significant correlations of π and length of branches to build networks of coevolution.

To detect coevolution based on topology with other genes, first, we used the distribution of RF distance between endosperm genes and the global endosperm topology to estimate the value obtained in 95% of the distribution. Then, we removed all the imprinted and endosperm genes with lower RF distance than this value to remove genes with a topology

explained by global topology. Finally, we compared the topology of each remaining imprinted gene and endosperm genes. All pairs of genes with a null RF distance were considered as significantly similar and used to build the last coevolution network.

For each network of coevolution, we compared the number of pairs of imprinted genes directly correlated with the number found between genes in 1000 random samples of 29 endosperm genes. Moreover, we tested for gene enrichment on each GO term in each network based on the occurrences of each term on 1000 random samples of equivalent size of all the endosperm and imprinted genes.The complete list of GO term associated to each gene was downloaded from Gene Ontology online platform for *A. thaliana* (Ashburner et al,. 2000).

Identification of sites in coevolution

We used a model-based approach for detecting coevolving positions of each pair of genes with a coevolution signal. The pairs of imprinted genes studied were all the pairs directly associated within at least one network previously mentioned. To use this method, we concatenated each sequence of pairs of genes for each species. If one species was not found in the other gene, the sequence associated with this species was removed.

Coevolving positions within each pair of genes were predicted using the clustering method implemented in the CoMap package (version 1.5.2; Dutheil and Galtier, 2007). The built-in algorithm was then used to correct for multiple testing and estimate the false discovery rate (FDR). In this study, 1,000 simulations per tested pair were performed, and the groups of sites with an adjusted p value $< 1\%$ were considered coevolving. Simple correlation was used as a measure of coevolution. For each pair comparison, we extracted the coevolving groups of substitutions and classified them by size (between 2 and 10).

To test the significant excess of correlations, we compared the results obtained with the results for the same number of pair comparisons of imprinted genes not directly associated in any network, for each cluster of coevolving groups of substitutions of equivalent size by Kruskal-Wallis tests. If the number of correlations is significantly higher than the number expected between two random genes by chance or by alignment errors, we expect a significant increase of these numbers in pairs of associated imprinted genes.

Data availability

The RNAseq data for leaf, root, pistil, and pollen tissues can be found in the NCBI BioProject database under submission number PRJNA1067428. The endosperm RNAseq data is available under submission number GSE76076.

Acknowledgements

This work has been supported by the Czech Science Foundation (GACR 22-20240S granted to CLP). Computational resources were supplied by the project 'e-Infrastruktura CZ' (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Author contributions

O.I developed and performed the transcriptional analysis. A.L analysed and interpreted the data. AL and C.L-P wrote the manuscript. All authors edited the manuscript.

Declaration of interests

The authors declare no competing interests.

Bibliography

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

2) Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29. https://doi.org/10.1038/75556

3) Batista, R.A., Köhler, C., 2020. Genomic imprinting in plants-revisiting existing models. Genes Dev 34, 24–36. https://doi.org/10.1101/gad.332924.119

4) Bemer, M., Heijmans, K., Airoldi, C., Davies, B., Angenent, G.C., 2010. An Atlas of Type I MADS Box Gene Expression during Female Gametophyte and Seed

Development in Arabidopsis. Plant Physiology 154, 287–300. https://doi.org/10.1104/pp.110.160770

5) Bonduriansky, R. 2007. The genetic architecture of sexual dimorphism: the potential roles of genomic imprinting and condition dependence. In Sex, Size And Gender Roles: Evolutionary Studies of Sexual Size Dimorphism. (D.J. Fairbairn, W.U. Blanckenhorn and T. Szekely, editors). Oxford University Press, Oxford, UK.

6) Brandvain, Y., Haig, D., 2005. Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. Am Nat 166, 330–338. https://doi.org/10.1086/432036

7) Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., Meinke, D., 2010. Identification of Nuclear Genes Encoding Chloroplast-Localized Proteins Required for Embryo Development in Arabidopsis. Plant Physiology 155, 1678. https://doi.org/10.1104/pp.110.168120

8) Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

9) Day, T., Bonduriansky, R., 2004. Intralocus sexual conflict can drive the evolution of genomic imprinting. Genetics 167, 1537. https://doi.org/10.1534/genetics.103.026211

10) Day, R.C., Herridge, R.P., Ambrose, B.A., Macknight, R.C., 2008. Transcriptome Analysis of Proliferating Arabidopsis Endosperm Reveals Biological Implications for the Control of Syncytial Division, Cytokinin Signaling, and Gene Expression Regulation. Plant Physiology 148, 1964–1984. https://doi.org/10.1104/pp.108.128108

11) Dekkers, B.J.W., Pearce, S., van Bolderen-Veldkamp, R.P., Marshall, A., Widera, P., Gilbert, J., Drost, H.-G., Bassel, G.W., Müller, K., King, J.R., Wood, A.T.A., Grosse, I., Quint, M., Krasnogor, N., Leubner-Metzger, G., Holdsworth, M.J., Bentsink, L., 2013. Transcriptional Dynamics of Two Seed Compartments with Opposing Roles in Arabidopsis Seed Germination. Plant Physiology 163, 205–215. https://doi.org/10.1104/pp.113.223511

12) DePristo, M.A., Banks, E., Poplin, R.E., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using

next-generation DNA sequencing data. Nat Genet 43, 491–498. https://doi.org/10.1038/ng.806

13) Dutheil, J. ,Galtier, N., 2007. Coevolution within and between Genes. https://doi.org/10.1159/000107599

14) Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

15) Edwards, S.V., 2009. Natural selection and phylogenetic analysis. Proceedings of the National Academy of Sciences 106, 8799–8800. https://doi.org/10.1073/pnas.0904103106

16) Foxe, J.P., Stift, M., Tedder, A., Haudry, A., Wright, S.I., Mable, B.K., 2010. Reconstructing Origins of Loss of Self-Incompatibility and Selfing in North American Arabidopsis Lyrata: A Population Genetic Context. Evolution 64, 3495–3510. https://doi.org/10.1111/j.1558-5646.2010.01094.x

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

18) Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40, D1178-1186. https://doi.org/10.1093/nar/gkr944

19) Green, A.G., Elhabashy, H., Brock, K.P., Maddamsetti, R., Kohlbacher, O., Marks, D.S., 2021. Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. Nat Commun 12, 1396. https://doi.org/10.1038/s41467-021-21636-z

20) Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307–321. https://doi.org/10.1093/sysbio/syq010

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

22) Haig, D., Westoby, M., 1991. Genomic Imprinting in Endosperm: Its Effect on Seed Development in Crosses between Species, and between Different Ploidies of the Same Species, and Its Implications for the Evolution of Apomixis. Philosophical Transactions: Biological Sciences 333, 1–13.

23) Haig, D., 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. Proceedings of the Royal Society of London. Series B: Biological Sciences 264, 1657–1662. https://doi.org/10.1098/rspb.1997.0230

24) Haig, D., 2000. The Kinship Theory of Genomic Imprinting. Annual Review of Ecology and Systematics 31, 9–32. https://doi.org/10.1146/annurev.ecolsys.31.1.9

25) Haig, D., 2014. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. Heredity 113, 96–103. https://doi.org/10.1038/hdy.2013.97

26) Halg, D., Graham, C., 1991. Genomic imprinting and the strange case of the insulin-like growth factor II receptor. Cell 64, 1045–1046. https://doi.org/10.1016/0092-8674(91)90256-X

27) Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T., Köhler, C., 2016. Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. The Plant Cell 28, 1815–1827. https://doi.org/10.1105/tpc.16.00304

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

29) Hough, J., Immler, S., Barrett, S.C.H., Otto, S.P., 2013. Evolutionarily stable sex ratios and mutation load. Evolution 67, 1915–1925. https://doi.org/10.1111/evo.12066

30) Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., Fischer, R.L., 2011. Regulation of imprinted gene expression in Arabidopsis endosperm. Proceedings of the National Academy of Sciences 108, 1755–1762. https://doi.org/10.1073/pnas.1019273108

31) Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F.X., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., Guo, Y.-L., 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet 43, 476–481. https://doi.org/10.1038/ng.807

32) İltaş, Ö., Čertner, M., Placette, C.L., 2024. Natural variation in sexual traits and gene expression between selfing and outcrossing Arabidopsis lyrata suggests sexual selection at work. Plant and Cell Physiology pcae090. https://doi.org/10.1093/pcp/pcae090

33) Kasimatis, K.R., Nelson, T.C., Phillips, P.C., 2017. Genomic Signatures of Sexual Conflict. Journal of Heredity 108, 780–790. https://doi.org/10.1093/jhered/esx080

34) Klosinska, M., Picard, C.L., Gehring, M., 2016. Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nature Plants 2, 1–8. https://doi.org/10.1038/nplants.2016.145

35) Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

36) Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

37) Llaurens, V., Whibley, A., Joron, M., 2017. Genetic architecture and balancing selection: the life and death of differentiated variants. Molecular Ecology 26, 2430–2448. https://doi.org/10.1111/mec.14051

38) McVean, G.T., Hurst, L.D., 1997. Molecular Evolution of Imprinted Genes: No Evidence for Antagonistic Coevolution. Proceedings: Biological Sciences 264, 739–746.

39) Mills, W., Moore, T., 2004. Polyandry, Life-History Trade-Offs and the Evolution of Imprinting at Mendelian Loci. Genetics 168, 2317–2327. https://doi.org/10.1534/genetics.104.030098

40) Miyake, T., Takebayashi, N., Wolf, D.E., 2009. Possible Diversifying Selection in the Imprinted Gene, MEDEA, in Arabidopsis. Molecular Biology and Evolution 26, 843–857. https://doi.org/10.1093/molbev/msp001

41) Moore, T., Haig, D., 1991. Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet 7, 45–49. https://doi.org/10.1016/0168-9525(91)90230-N

42) Noguero, M., Atif, R.M., Ochatt, S., Thompson, R.D., 2013. The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. Plant Science 209, 32–45. https://doi.org/10.1016/j.plantsci.2013.03.016

43) Otto, S.P., Scott, M.F., Immler, S., 2015. Evolution of haploid selection in predominantly diploid organisms. Proceedings of the National Academy of Sciences 112, 15952–15957. https://doi.org/10.1073/pnas.1512004112

44) Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R., Wolf, J.B., 2014. The evolution of genomic imprinting: theories, predictions and empirical tests. Heredity 113, 119–128. https://doi.org/10.1038/hdy.2014.29

45) Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., Gehring, M., 2014. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife 3, e03198. https://doi.org/10.7554/eLife.03198

46) Pignatta, D., Novitzky, K., Satyaki, P.R.V., Gehring, M., 2018. A variably imprinted epiallele impacts seed development. PLOS Genetics 14, e1007469. https://doi.org/10.1371/journal.pgen.1007469

47) Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

48) Raissig, M.T., Baroux, C., Grossniklaus, U., 2011. Regulation and Flexibility of Genomic Imprinting during Seed Development. The Plant Cell 23, 16–26. https://doi.org/10.1105/tpc.110.081018

49) Ross-Ibarra, J., Wright, S.I., Foxe, J.P., Kawabe, A., DeRose-Wilson, L., Gos, G., Charlesworth, D., Gaut, B.S., 2008. Patterns of Polymorphism and Demographic History in Natural Populations of Arabidopsis lyrata. PLOS ONE 3, e2411. https://doi.org/10.1371/journal.pone.0002411

50) Scott, R.J., Spielman, M., Bailey, J., Dickinson, H.G., 1998. Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125, 3329–3341. https://doi.org/10.1242/dev.125.17.3329

51) Soliman, H.K., Coughlan, J.M., 2024. United by conflict: Convergent signatures of parental conflict in angiosperms and placental mammals. J Hered 115, 625–642. https://doi.org/10.1093/jhered/esae009

52) Spencer, H.G., Clark, A.G., 2014. Non-conflict theories for the evolution of genomic imprinting. Heredity 113, 112–118. https://doi.org/10.1038/hdy.2013.129

53) Spillane, C., Schmid, K.J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., Gagliardini, V., Page, D.R., Wolfe, K.H., Grossniklaus, U., 2007. Positive darwinian selection at the imprinted MEDEA locus in plants. Nature 448, 349–352. https://doi.org/10.1038/nature05984

54) Sujeeth, N., Mehterov, N., Gupta, S., Qureshi, M.K., Fischer, A., Proost, S., Omidbakhshfard, M.A., Obata, T., Benina, M., Staykov, N., Balazadeh, S., Walther, D., Fernie, A.R., Mueller-Roeber, B., Hille, J., Gechev, T.S., 2020. A novel seed plants gene regulates oxidative stress tolerance in Arabidopsis thaliana. Cell. Mol. Life Sci. 77, 705–718. https://doi.org/10.1007/s00018-019-03202-5

55) Sylvestre, F., Mérot, C., Normandeau, E., Bernatchez, L., 2023. Searching for intralocus sexual conflicts in the three-spined stickleback (Gasterosteus aculeatus) genome. Evolution 77, 1667–1681. https://doi.org/10.1093/evolut/qpad075

56) Tuteja, R., McKeown, P.C., Ryan, P., Morgan, C.C., Donoghue, M.T.A., Downing, T., O'Connell, M.J., Spillane, C., 2019. Paternally Expressed Imprinted Genes under Positive Darwinian Selection in Arabidopsis thaliana. Molecular Biology and Evolution 36, 1239–1253. https://doi.org/10.1093/molbev/msz063

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

58) Willi, Y., Fracassetti, M., Zoller, S., Van Buskirk, J., 2018. Accumulation of Mutational Load at the Edges of a Species Range. Molecular Biology and Evolution 35, 781–791. https://doi.org/10.1093/molbev/msy003

59) Wilkins, J.F., Haig, D., 2001. Genomic imprinting of two antagonistic loci. Proceedings of the Royal Society of London. Series B: Biological Sciences 268, 1861–1867. https://doi.org/10.1098/rspb.2001.1651

60) Wilkins, J.F., 2011. Genomic imprinting and conflict-induced decanalization. Evolution 65, 537–553. https://doi.org/10.1111/j.1558-5646.2010.01147.x

61) Williamson, R.J., Josephs, E.B., Platts, A.E., Hazzouri, K.M., Haudry, A., Blanchette, M., Wright, S.I., 2014. Evidence for Widespread Positive and Negative Selection in Coding and Conserved Noncoding Regions of Capsella grandiflora. PLOS Genetics 10, e1004622. https://doi.org/10.1371/journal.pgen.1004622

62) Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M., Köhler, C., 2011. High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLOS Genetics 7, e1002126. https://doi.org/10.1371/journal.pgen.1002126

63) Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., Mering, C. von, Grossniklaus, U., 2010. Arabidopsis Female Gametophyte Gene Expression Map Reveals Similarities between Plant and Animal Gametes. Current Biology 20, 506–512. https://doi.org/10.1016/j.cub.2010.01.051

64) Zhang, S., Wang, D., Zhang, H., Skaggs, M.I., Lloyd, A., Ran, D., An, L., Schumaker, K.S., Drews, G.N., Yadegari, R., 2018. FERTILIZATION-INDEPENDENT SEED-Polycomb Repressive Complex 2 Plays a Dual Role in Regulating Type I MADS-Box Genes in Early Endosperm Development. Plant Physiology 177, 285–299. https://doi.org/10.1104/pp.17.00534

65) Zhang, B., Li, C., Li, Y., Yu, H., 2020. Mobile TERMINAL FLOWER1 determines seed size in Arabidopsis. Nat Plants 6, 1146–1157. https://doi.org/10.1038/s41477-020-0749-5

66) Zou, H.-F., Zhang, Y.-Q., Wei, W., Chen, H.-W., Song, Q.-X., Liu, Y.-F., Zhao, M.-Y., Wang, F., Zhang, B.-C., Lin, Q., Zhang, W.-K., Ma, B., Zhou, Y.-H., Zhang, J.-S., Chen, S.-Y., 2012. The transcription factor AtDOF4.2 regulates shoot branching and seed coat formation in Arabidopsis. Biochemical Journal 449, 373–388. https://doi.org/10.1042/BJ20110060

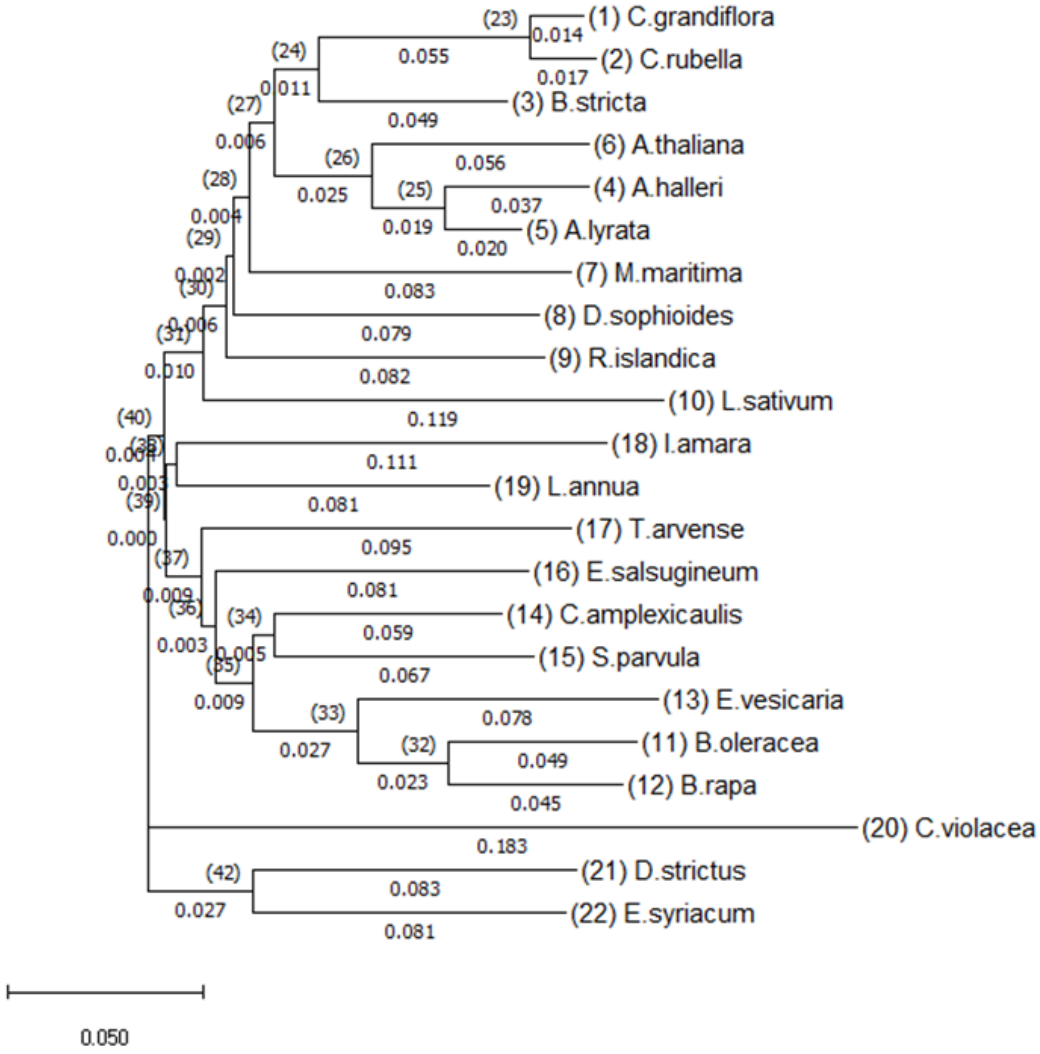

Figure S1: phylogenetic tree obtained by HKY85 model based on the 1055 endosperm genes found in all species after 1000 bootstrap. *The percentage of trees in which the associated haplotypes clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in mean number of substitutions per site.*

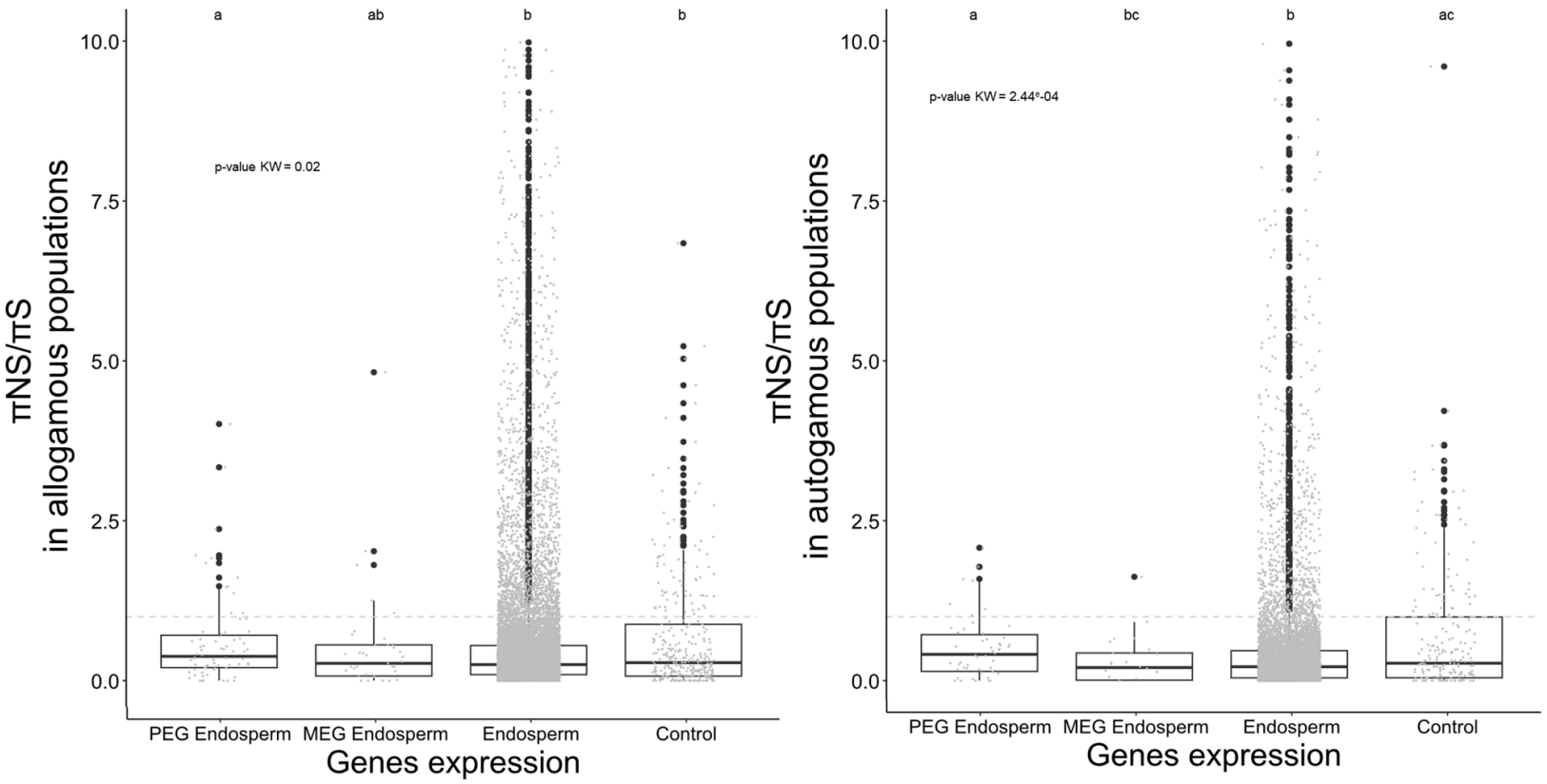

Figure S2: Variations of the polymorphism in the allogamous and autogamous populations in the imprinted genes compared to endosperm and control genes. *boxplot of the variations of the π_{NS}/π_S by gene and populations. The dashed grey lines represent the threshold value for neutrality (=1). The grey points represent the value of each gene in each population. The minuscule letters (a,b,c) signified the groups significantly different following Dunn's test (p-values < 0.05.).*