The effect of parental conflict on imprinted genes in *A. lyrata.*

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

[1] Department of Botany, Faculty of Science, Charles University, CZ-128 01 Prague, Czech Republic

[2] 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. A common assumption is that parental conflict over resource allocation to the offspring promotes directional selection on imprinted genes, but so far, most studies addressing this question have focused on autogamous species, whose parental conflict is expected to be low. Moreover, parental conflict is expected to promote antagonistic coevolution between imprinted genes, but this has never been tested so far.

To address these knowledge gaps, we searched for signals of selection on imprinted genes across the Brassicaceae family based on phylogenetic approaches. We then investigated and compared footprints of selection on imprinted genes in allogamous and autogamous populations of *Arabidopsis lyrata*. Finally, we investigated signals of coevolution across imprinted genes based on phylogenetics and genomics approaches. Overall, we did not find specific signals of selection on imprinted genes in the allogamous populations of *A. lyrata*. Despite some exceptions, these genes were overall not associated with directional selection in allogamous populations and were poorly affected by the mating system, suggesting a limited role of parental conflict. Nevertheless, non-imprinted endosperm genes presented typical signals of balancing selection across the Brassicaceae, as expected in the case of unresolved intralocus sexual conflict. Finally, we found imprinted genes showing signals of recent coevolution, but not related to specific sites, suggesting that drivers other than protein interactions lead to the coevolution between imprinted genes.

Key words: genomic imprinting, parental conflict, positive Darwinian selection, 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), but the most frequently invoked is the parental conflict hypothesis (Haig, 2000; Haig, 2014). 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 parental conflict 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).

In plants, parental conflict is expected at the level of the endosperm, the embryo's nourishing tissue in the seed. Consistent with a conflict over resource allocation, 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. 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).

Parental conflict should lead to positive selection on coding sequences of imprinted genes expressed from the maternally and paternally inherited genomes (Wilkins and Haig, 2001; Mills and Moore, 2004; Haig, 2014). Consistent with this assumption, signatures of 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). However, these studies did not detect this signal on Maternally Expressed Genes (MEG), suggesting that positive selection is not a general phenomenon expected on imprinting genes. Moreover, these observations raised the question of the implication of parental conflict on this positive selection because parental conflict is expected only in allogamous species and not in autogamous species (Brandvain and Haig, 2005). 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). Thus, the signature of selection expected on imprinted genes and the implication of the parental conflict remains unclear in plants.

Another consequence of parental conflict 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). 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 biassed 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. However, to our knowledge, the signals of coevolution between imprinted genes, based on population genomics as on phylogenetics, were never investigated so far.

To determine the selection induced by the genomic imprinting and mating system in allogamous species, we analysed known imprinted genes of *A. lyrata* (Klosinska et al., 2016). This species is close to the model species *A. thaliana* and has both cross-pollinating and selfing populations in North America (Ross-Ibarra et al., 2008). Using phylogenetic data across 23 species of *Brassicaceae* and genomic data from nine different allogamous and nine autogamous populations of *A. lyrata*, we searched: 1) the selective pressure acting on imprinted genes in allogamous populations, 2) the implication of mating type on this selective pressure and 3) evidence of coevolution between imprinting genes and the other genes upregulated in the endosperm. Our findings suggest that parental conflict promotes diversification on non-imprinted genes, specifically in allogamous populations, but not specific selection on imprinted 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_e and MEG_e. The other PEGs and MEGs have been named PEG_o and MEG_o, whereas the 2,703 non-imprinted genes upregulated in endosperm have been named endosperm genes. A set of 100 randomly chosen non imprinted ubiquitous genes will be considered as control genes.

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 Krustal-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 and the PEG_o were significantly smaller than control genes (table S2).

The PEG_e present phylogenetic characteristic of negative selection in Brassicaceae.

We investigated the phylogenetic characteristics of imprinted genes in *A. lyrata* and twenty-one other Brassicaceae (table S3) 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 1A; table S4). This number varied significantly across gene groups (p value KW=2.97^{e-15}), with a significant increase in endosperm genes compared to control (less gene turnover), and a decrease in the PEG_e (higher turnover) compared to endosperm. 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 Krustal-Wallis = 0.6; Fig 1B). 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 mean length of all branches

found in the imprinted genes varied significantly across groups of genes (Fig 1C; p value KW< 2.2e-16; table S4). This mean in all the groups of genes were significantly higher than control genes, but this value was significantly lower for the MEG_o and PEG_e 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 S4) were only a significant increase in endosperm and PEG_o compared to control, and the length of branches found in the imprinted genes was not significantly different than in the endosperm genes (table S4).

Finally, we estimated the variations of the D_{NS}/D_S between *A. lyrata* and the other Brassicaceae species. We expect an increase of this ratio under positive directional selection, and the opposite under negative selection. Before analysis, the genes with homologous genes found in less than three other Brassicaceae and without divergent synonymous sites were removed. This ratio varied significantly across groups of genes (p-value Krustal-Wallis = 7.02^{e-05}; Fig 1D; table S4), and was significantly lower in the MEG_e and the PEG_e than in endosperm genes.

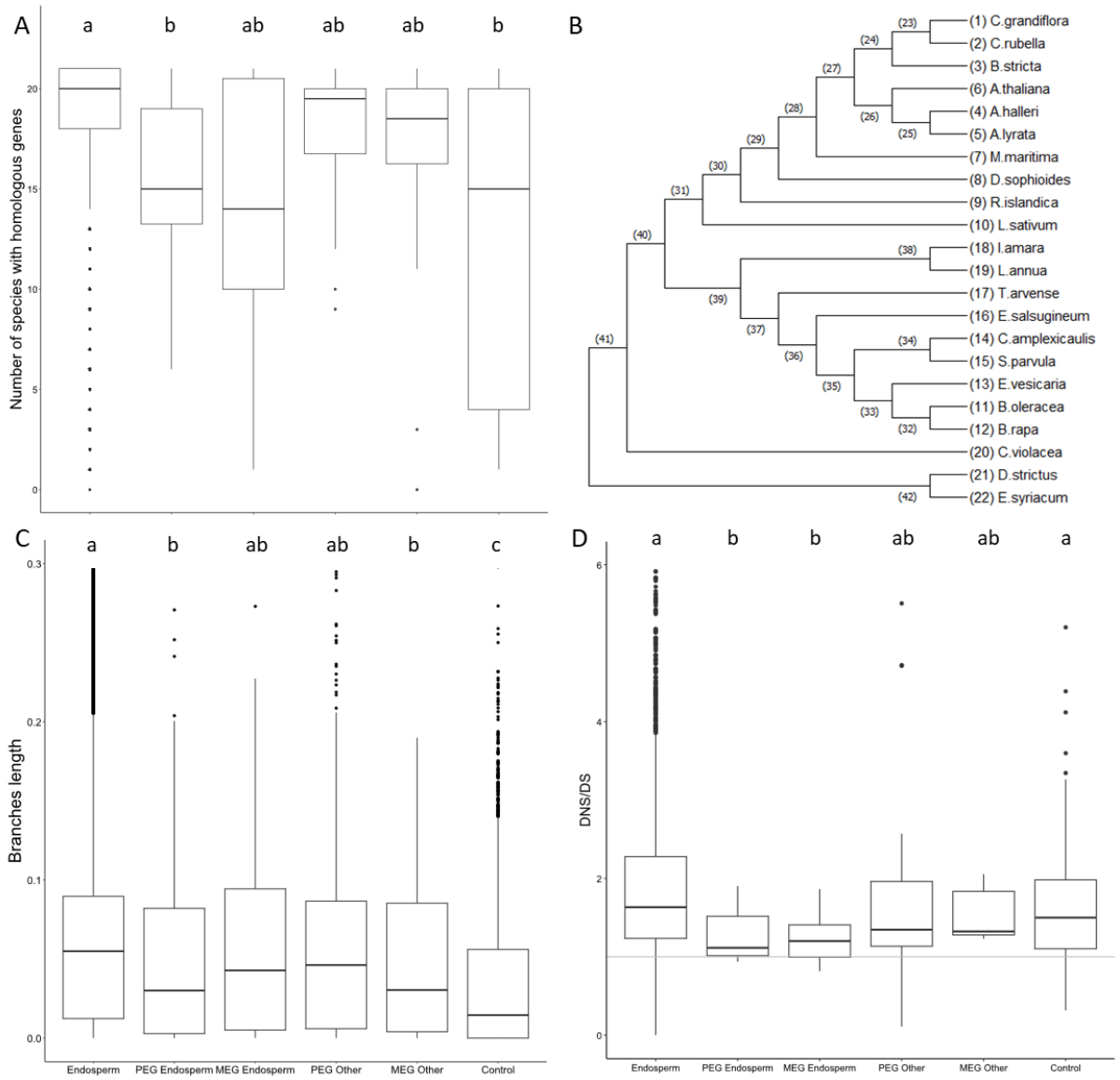

Figure 1: Phylogenetic characteristic of the imprinted genes compared to endosperm genes across the Brassicaceae . *A: boxplot of the variations of the number of species with homologous genes by group of genes. B: phylogenetic tree obtained by HKY85 model[41] based on the 1055 endosperm genes found in 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. C: boxplot of the variations of the mean length branch by gene based on all the species. D: boxplot of the variations of the* D_{NS}/D_S *by gene. The light grey lines represent the threshold value for neutrality. The minuscule letters (a,b,c) signified the groups significantly different following Dunn's test (p-values < 0.05.).*

No signal of positive selection related to the mating system was detected on imprinted genes

To estimate the effect of mating system and imprinting on polymorphism in populations, we compared Tajima's D, the π_{NS}/π_S ratio and the fixation index (= $(D_{ns}/D_s)/(\pi_{NS}/\pi_S)$; FI) of the imprinted genes with the endosperm and control genes in nine allogamous and nine autogamous populations (table S5). 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 but the proportion of MEG_o was significantly lower than endosperm genes (binomial tests; table S6). In the allogamous populations, the Tajima's D varied significantly across the gene groups (p-value Krustal-Wallis = 7.02^{e-05}), but the effect is explained only by a significant decrease for PEG_e compared to MEG_o and is not related to differences compared to endosperm or control genes (Fig S1A; table S7). In the autogamous populations, the Tajima's D did not vary significantly across the groups (p-value Krustal-Wallis = 0.052; table S7). The Tajima's D decreased significantly for endosperm genes in autogamous populations compared to allogamous populations, whereas the same statistics increased significantly on PEG_e (table S8). The Tajima's D for other gene groups did not vary with the mating system.

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, except the proportion of MEG_o that was significantly lower than endosperm genes (binomial tests; table S6). This ratio varied significantly across the groups of genes in allogamous and autogamous (p-values Krustal-Wallis = 7.47^{e-04} and 7.48^{e-05} respectively). However, the tendencies observed were different: in the allogamous populations, the ratios of the PEG_e and the MEG_o increased significantly compared to the endosperm genes, and the ratio of control genes was similar than in endosperm genes (Fig 2A; table S7). In the autogamous populations, only the ratio of the PEG_e increased significantly compared to the endosperm genes, and the ratio on control genes was significantly higher than in endosperm genes (Fig 2B; table S7). The π_{NS}/π_S of endosperm genes decreased significantly in autogamous populations

compared to allogamous populations (table S8). The π_{NS}/π_S of other gene groups did not vary with the mating system.

We attempted to detect signals of selection by a variation of the FI as in Tuteja et al. (2019). We expected an increase of the FI in case of positive selection. However, this method requires divergent synonymous positions and a π_{NS}/π_S higher than 0. Thus, the genes that did not respect these criterias were removed before analysis. The proportion of endosperm and PEG_o genes conserved for the analysis increased significantly compared to control but the proportion of MEG_o was significantly lower than in the endosperm genes (binomial tests; table S6). The FI varied significantly across the groups of genes in allogamous and autogamous (p-values Krustal-Wallis = 2.54^{e-07} and 4.16^{e-07} respectively), consistently with the observations on π_{NS}/π_S. In the allogamous populations, the FI of the PEG_e and the MEG_o decreased significantly compared to the endosperm and control genes (Fig 2C; table S7), whereas only the FI of the PEG_e decreased significantly compared to the endosperm genes in the autogamous populations (Fig 2D; table S7). Consistent with the π_{NS}/π_S, only the FI increased significantly for endosperm genes in autogamous populations compared to allogamous populations (table S8).

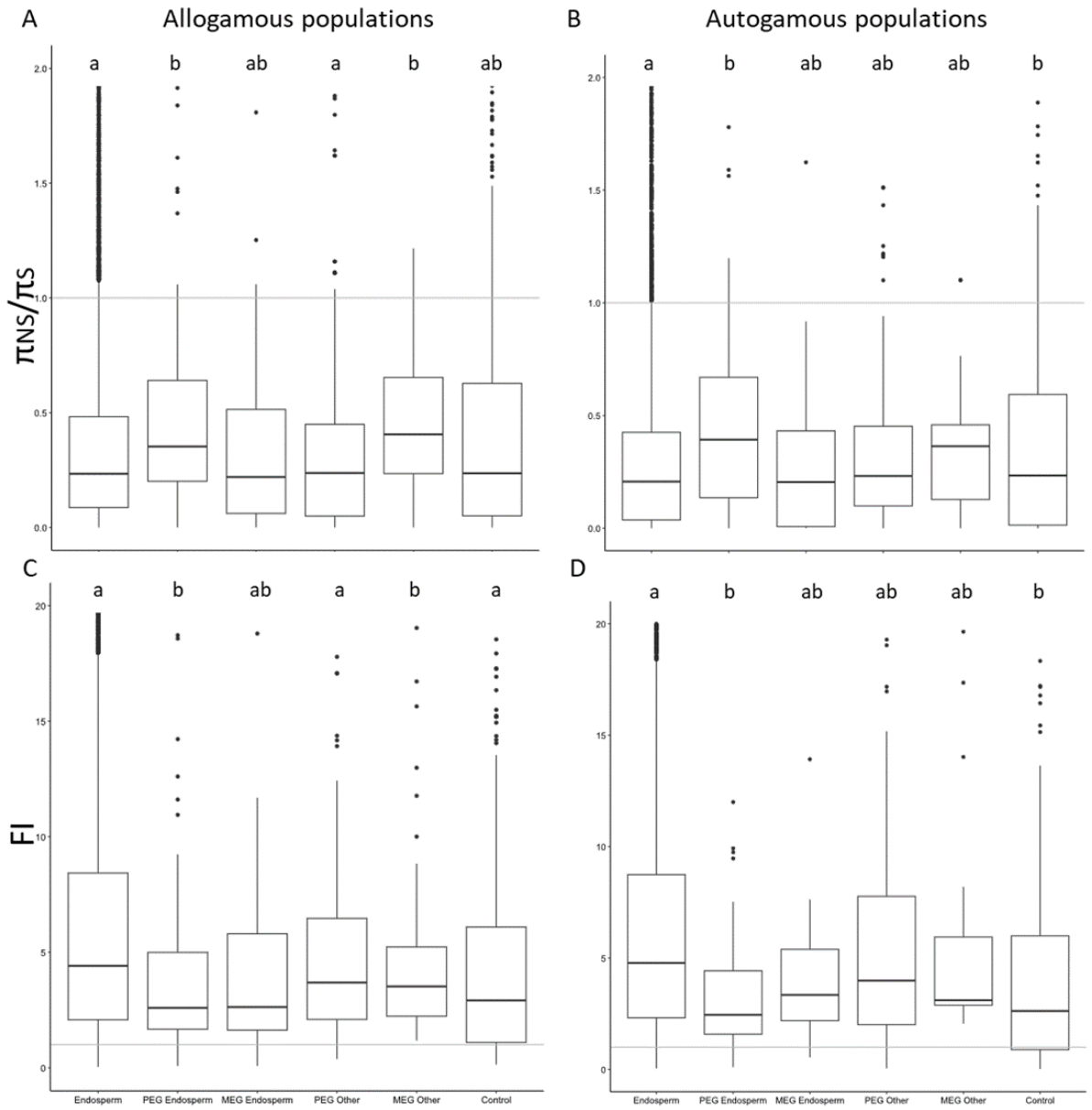

Figure 2: Variation of the polymorphism in the allogamous and autogamous populations in the imprinted genes compared to endosperm and control genes. *A and B: boxplot of the variations of the median π_{NS}/π_S by gene. C and D: boxplot of the variations of the median FI by gene. The light grey lines represent the threshold value for neutrality. A and C: allogamous populations. B and D: autogamous populations. 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 length of branches across the Brassicaceae, and/or mean of Tajima's D in allogamous populations

and/or mean π_{NS}/π_S in allogamous populations than 5% of these same values in the endosperm genes. We compared the list of candidate genes found with a reduced Tajima's D and/or π_{NS}/π_S in allogamous populations with those found with the same method in autogamous populations to distinguish the genes under directional selection because of the parental conflict. One MEG_o, two MEG_e and one PEG_o were associated with a lower mean length of branches across the Brassicaceae (table S1). Based on homologous genes in *A. thaliana*, one MEG_e encodes for 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), and the PEG_o encodes a non-catalytic subunit common to Nuclear DNA-dependent RNA polymerases IV and V implicated on the RNA-Directed DNA Methylation pathway (Ma et al., 2015). The second MEG_e was also associated with a lower mean Tajima's D in allogamous populations, as well as one PEG_e (table S1). Finally, one MEG_e encoding a Dof-type zinc finger domain-containing protein implicated in seed coat formation (Zou et al., 2012) and one PEG_o were associated with a reduced mean π_{NS}/π_S in allogamous populations (table S1). None of these genes with reduced Tajima's D and π_{NS}/π_S in allogamous populations were found with similar signals in autogamous populations

Then, we searched for imprinted genes associated with strong signals of negative and positive selection in *A. lyrata* compared to endosperm genes by searching imprinted genes with a lower and higher mean D_{NS}/D_S than 5% of the endosperm genes respectively. Two MEG_e previously mentioned were also associated with a reduced D_{NS}/D_S, as well as two other PEG_o (table S1). Moreover, one MEG_o and one PEG_o were associated with an increased mean D_{NS}/D_S (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 (Length of branches, Tajima's D, π_{NS}/π_S, increase and decrease of D_{NS}/D_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 parental conflict promotes coevolution between genes favouring paternal or maternal fitness in an arms race scenario. 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.

Based on these values, 25 PEG_o, 7 MEG_o, 7 PEG_e and 10 MEG_e presented 85 significant correlations of polymorphism together. The number of correlations between PEG and MEG (*e.g*. 85) and the number of genes correlated (*e.g.* 49) were significantly higher than in 1000 random samplings of 73 endosperm genes (p-values= 6^{e-03} and $<1^{e-03}$ respectively). Moreover, some of these imprinted genes also presented significant correlation of polymorphism with 145 endosperm genes that were significantly correlated by polymorphism to 5 PEG_o, 4 MEG_o, 6 PEG_e and 4 MEG_e (Fig 3A; table S9). The complete network, including the genes previously mentioned and the 1,865 other endosperm genes with a polymorphism correlated with one or more imprinted genes, did not present enrichment on specific GO terms than expected in random samples of endosperm and imprinted genes of same size (2,078 genes). Three imprinted genes were correlated with more genes than 99% of the endosperm genes (table S1 and S9).

For the length of the branches, 15 PEG_o, 5 MEG_o, 6 PEG_e and 10 MEG_e presented 46 significant correlations together. The number of correlations between PEG and MEG (*e.g*. 46) and the number of genes correlated (*e.g.* 36) were not significantly higher than in 1000 random samplings of 73 endosperm genes (p-values = 0.08 and 0.051 respectively). Moreover, some of these imprinted genes also presented significant correlation of length of the branches with 282 endosperm genes, themselves significantly correlated to 17 PEG_o, 5 MEG_o, 8 PEG_e and 4 MEG_e (Fig 3B, table S10). The complete network, including the genes previously mentioned and the 1,012 other endosperm genes with a polymorphism correlated with one or more imprinted genes, was significantly associated with the negative regulation of catalytic activity by GO (table S11), compared to GO expected in random samples of endosperm and imprinted genes of same size (1,364 genes). All genes (*AT2G47670*, *AT3G05741*, *AT3G12880*, *AT4G00872* and *AT4G15750* for *AL4G47240*, *AL3G16450*, *AL3G24550, AL6G52570,* and *AL7G39610* in *A. lyrata* respectively) with this GO term (GO:0043086) were involved in pectin methylesterase inhibition and cell wall formation. One MEG_e was correlated with more genes than 99% of the endosperm genes (table S1 and S11). Interestingly, this AGAMOUS-like protein was previously mentioned because it is associated with reduced length of branches and D_{NS}/D_S.

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 (Fig 1B) lower than 95% of the distribution of the RF distance found in endosperm genes (=0.71). Based on these criterias, the topology of one PEG_e was significantly related to six endosperm genes. Three of them presented topologies significantly similar to one PEG_o. The topology of these imprinted genes was also similar with two other endosperm genes (Fig 3C; table S12). Moreover, three endosperm genes presented significant similar topologies with two PEG_o, and one topology of these endosperm genes was also significantly similar to one other PEG_o. These two networks of genes were significantly associated with proteolyse and protein deubiquitination according to GO (table S11). Moreover, the PEG_e (*AL6G33630*) presented a significantly higher number of associations compared to 99% of non-imprinted endosperm genes (table S12).

Thus, within the 2,628 possible pairs involving imprinted genes, 131 including 58 PEG/MEG (77%) 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 S13). Thus, these pairs of imprinted genes directly correlated were not associated with an increase on the number of coevolving sites.

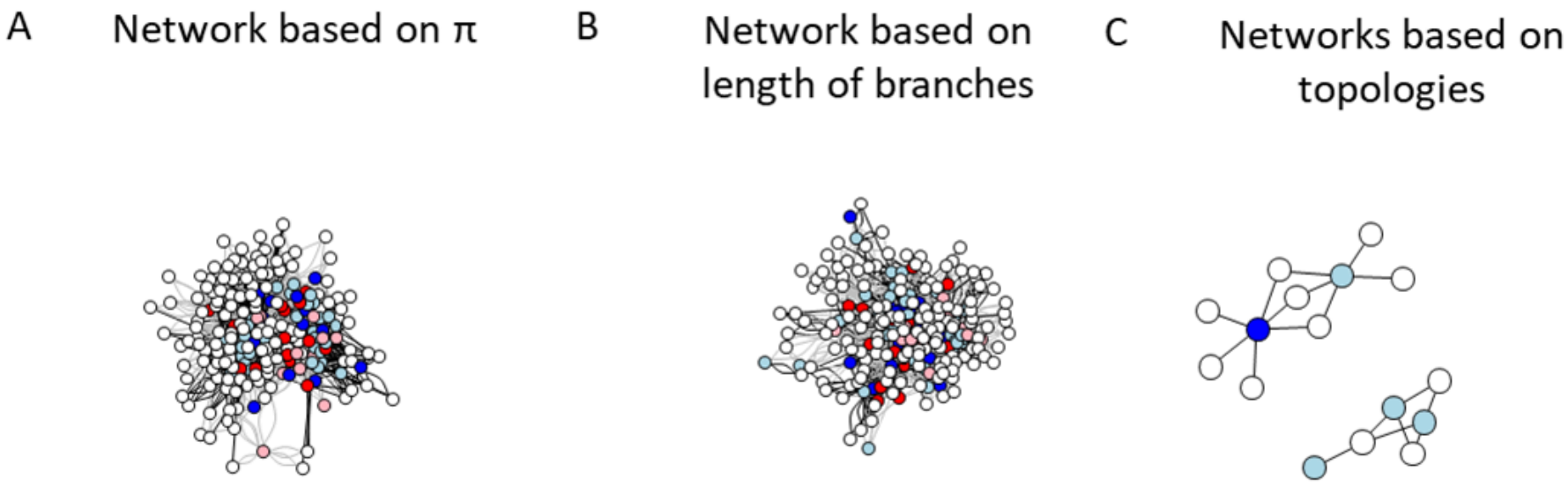

Figure 3: Networks of the imprinted and endosperm genes based on A) polymorphism in population , B) length of branches and C) phylogenetic topologies across the Brassicaceae. *A) Network based on the Pearson's correlations of π found on genes in the populations. B) Network based on the Pearson's correlations of the length of branches for each gene estimated with the global topology. C) Networks based on the RF distance of topologies. Only the correlations that associate all the PEG/MEG of the networks together are represented. The correlations between PEG/MEG directly correlated and the endosperm genes are masked to simplify. The black and grey lines represent the negative and positive correlations respectively. Blue circles: PEG_e. Light blue circles: PEG_o. Red circles: MEG_e. Pink circles: MEG_o. White circles: endosperm genes required to correlate all the PEG/MEG.*

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 explanation for the evolution of imprinting is the parental conflict hypothesis (Haig, 2000). This conflict is expected to lead imprinted genes to directional selection in allogamous populations and to promote the coevolution of imprinted genes having antagonistic effects on offspring growth (Wilkins and Haig, 2001; Mills and Moore, 2004). This coevolution could be explained by direct or indirect interactions between the antagonist imprinted genes (Wilkins and Haig, 2001; Willi, 2013). Based on population genomic and phylogenetic approaches, we investigated whether imprinted genes are associated to specific signals of selection in allogamous populations of *A. lyrata*, if these signals are sensitive to mating systems and if these genes present signals of coevolution.

Mixed phylogenetic signals of selection on imprinted genes.

The parental conflict theory predicts that many genes under imprinting should affect growth (Haig, 2000). Thus, if the imprinted genes are implicated in parental conflict, these genes must be preferentially expressed on endosperm tissue. Consistent with the parental conflict hypothesis, 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 parental conflict.

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_e and the MEG_o present a reduced length of branches across Brassicaceae compared to endosperm genes, translating a fast evolutionary rate. However, imprinting is barely conserved across 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 not found when we considered only the *A. lyrata* species, where these genes are imprinted. Moreover, contrary to previous studies, the reduction of the D_{NS}/D_S on these PEG_e and the MEG_e suggest more a process of negative selection than a process of positive selection. Interestingly, the PEG_o did not exhibit this acceleration of the evolutionary rate and the MEG_o were not associated with a reduction of D_{NS}/D_S. These observations tend to confirm that the PEG and MEG

specific to endosperm are under singular and different evolutionary processes compared to the imprinted genes expressed in other tissues.

Generally no specific signal of selection on imprinted genes in allogamous populations, with a few exceptions.

Parental conflict is expected to promote directional selection on imprinted genes compared to other genes under parental 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). In the allogamous population of *A. lyrata*, consistent with the study on *C. rubella* (Hatorangan et al., 2016*)*, we didn't find a homogenous signal of selection based on population genomics. In detail, we did not find a specific signal of selection compared to endosperm genes on MEG_e and on PEG_o, whereas the PEG_e and MEG_o presented a decrease of the FI, consistent with negative selection previously mentionned, but also an increase of the π_{NS}/π_S, and an increase of the Tajima's D on MEG_o. These signals, totally opposite from the directional selection expected under the parental conflict theory (Haig, 2014), could translate a recent relaxation of selection. This relaxation of selection could be 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. However, for the MEG_o, the signal of relaxation compared to endosperm genes is maybe overestimated because the higher proportion of these genes removed because of an absence of polymorphism, that could be explained by an intense directional selection in population.

Moreover, it is 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.

Parental conflict is expected to be relaxed in autogamous lineages compared to allogamous ones (Haig, 2000; Brandvain and Haig, 2005). This phenomenon could explain 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, these 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. Consistently, the absence of an effect of mating type on the polymorphism of the imprinted genes in our study suggests a limited effect of the parental conflict on selection on these genes.

Nevertheless, we found some exceptions (eleven) of imprinted genes with a signal of directional selection in *A. lyrata* allogamous populations. Interestingly, these genes were never associated with signals of directional selection in the autogamous populations, suggesting parental conflict at work. Among the eleven imprinted genes with signals of directional selection, four of them were associated with negative selection and two with positive selection. Interestingly, one of these genes, the MEG_e that encodes an AGAMOUS-like protein implicated on female gametophyte, seed and endosperm development (Bemer et al., 2010; Zhang et al., 2018), presented also significant footprints of coevolution with a network of endosperm genes, suggesting a pivotal role of this gene in endosperm development associated with allogamy.

Potential evolutionary signals of parental conflict on non-imprinted endosperm genes.

Genomic imprinting, enabling parent-specific expression in a biparental tissue, is expected to allow directional selection under parental conflict, and thus to resolve such conflict. Without genomic imprinting, genes with antagonistic effects on maternal and paternal fitness 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: these genes were more found across the Brassicaceae than the control genes and these genes presented longer branches (Edwards, 2009).

Parental conflict is expected only in case of multiple mating (Haig, 2000), and the selection on allocation ressources should be equivalent between the male and the female in autogamous populations. Thus, we could expect that autogamy, by "resolving" parental conflict, would allow directional selection on genes implicated in resource allocation which

may have been under balancing selection due to their antagonistic effect on maternal and paternal fitness. Consistent with our expectations, in the autogamous populations, the π_{NS}/π_S in endosperm genes decreased compared to control genes, suggesting a general directional selection on these genes. Moreover, the autogamy was associated with a decrease in the π_{NS}/π_S and in the Tajima's D on endosperm genes compared to allogamous populations, suggesting that the process of selection on these genes is sensitive to mating type. These observations suggested that the absence of parental conflict on 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 parental conflict and the balancing selection it exerts on endosperm genes may allow directional selection on these genes.

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 parental conflict. Another coevolution network involving imprinted genes, based on topology, showed enrichment in proteolyse and protein deubiquitination, a pathway also regulating seed size (Li & Li, 2014). 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. If parental conflict-driven, this coevolution may thus act on antagonistic genes yet not directly interacting at the protein level.

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 more efficient and ancient negative selection on imprinted genes expressed in endosperm. However, we did not detect any clear impact of the mating type on selection on these genes, except for a few exceptions. Thus, parental conflict appears to have limited effect on the evolution of imprinted genes in *A. lyrata* populations. Nevertheless, for the first time, we found signals of coevolution between imprinted genes, an expectation from the parental conflict theory. These signals may or may not be due to parental conflict, and future studies are required to further investigate the cause for this coevolution.

Materials and Methods

We expected that parental conflict promotes 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 the parental conflict 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 parental conflict, to the signature associated with other possible selection process, we distinguished the imprinted genes upregulated specifically in the endosperm ($14PEG_e/15MEG_e$) with the other imprinted genes not specifically upregulated in the endosperm ($32PEG_o/12MEG_o$) 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 these clusters of imprinted genes 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).

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, with two-sided P values adjusted using the Bonferroni method as implemented in the R package dunn.test (version 1.3.6; Dinno, 2024).

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 S5), 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 S5.

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. Because parental conflict is expected to promote positive directional selection on imprinted genes and balancing selection on non-imprinted genes in allogamous populations, we expected a general decrease of π_{NS}/π_S and of the Tajima's D on imprinted genes compared to endosperm genes

in the allogamous populations. However, these differences of statistics between imprinted and non-imprinted genes were not expected for the autogamous populations. In fact, in the case of selfing, we expect that non-imprinted endosperm genes present signals of positive directional selection because of the absence of parental conflict.

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)[37] 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, with two-sided P values adjusted using the Bonferroni method as implemented in the R package dunn.test (version 1.3.6; Dinno, 2024).

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 S3). 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; Dinno, 2024).

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 (table S6).

Then, we further compared the fixation index (FI) on imprinted genes with endosperm and control genes based on the ratio of the between-species divergence (D_{NS}/D_S) to the within-species polymorphism (π_{NS}/π_S) on the populations of allogamous and autogamous as used in Tuteja et al (2019).

Based on these data, to test for significant differences on the D_{NS}/D_S and FI 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; Dinno, 2024).

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 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, with two-sided P values adjusted using the Bonferroni method as implemented in the R package dunn.test (version 1.3.6; Dinno, 2024).

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 73 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 Krustal-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) Ascencio-Ibáñez, J.T., Sozzani, R., Lee, T.-J., Chu, T.-M., Wolfinger, R.D., Cella, R., Hanley-Bowdoin, L., 2008. Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol 148, 436–454. https://doi.org/10.1104/pp.108.121038

3) 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

4) 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

5) 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

6) Bollier, N., Sicard, A., Leblond, J., Latrasse, D., Gonzalez, N., Gévaudant, F., Benhamed, M., Raynaud, C., Lenhard, M., Chevalier, C., Hernould, M., Delmas, F., 2018. At-MINI ZINC FINGER2 and Sl-INHIBITOR OF MERISTEM ACTIVITY, a Conserved Missing Link in the Regulation of Floral Meristem Termination in Arabidopsis and Tomato. The Plant Cell 30, 83–100. https://doi.org/10.1105/tpc.17.00653

7) 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

8) 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

9) Chen, I.-P., Haehnel, U., Altschmied, L., Schubert, I., Puchta, H., 2003. The transcriptional response of Arabidopsis to genotoxic stress – a high-density colony array study (HDCA). The Plant Journal 35, 771–786. https://doi.org/10.1046/j.1365-313X.2003.01847.x

10) 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

11) 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

12) 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

13) 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

14) 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

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

16) 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

17) 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

18) Estavillo, G.M., Crisp, P.A., Pornsiriwong, W., Wirtz, M., Collinge, D., Carrie, C., Giraud, E., Whelan, J., David, P., Javot, H., Brearley, C., Hell, R., Marin, E., Pogson, B.J., 2011. Evidence for a SAL1-PAP Chloroplast Retrograde Pathway That Functions in Drought and High Light Signaling in Arabidopsis. The Plant Cell 23, 3992–4012. https://doi.org/10.1105/tpc.111.091033

19) Eyre-Walker, A., Hurst, L.D., 2001. The evolution of isochores. Nat Rev Genet 2, 549–555. https://doi.org/10.1038/35080577

20) Feng, C., Cai, X.-W., Su, Y.-N., Li, L., Chen, S., He, X.-J., 2021. Arabidopsis RPD3-like histone deacetylases form multiple complexes involved in stress response. Journal of Genetics and Genomics 48, 369–383. https://doi.org/10.1016/j.jgg.2021.04.004

21) 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

22) Garcia, M.E., Lynch, T., Peeters, J., Snowden, C., Finkelstein, R., 2008. A small plant-specific protein family of ABI five binding proteins (AFPs) regulates stress response in germinating Arabidopsis seeds and seedlings. Plant Mol Biol 67, 643–658. https://doi.org/10.1007/s11103-008-9344-2

23) 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

24) 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

25) 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

26) 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

27) 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

28) 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

29) 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

30) 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

31) 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

32) 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

33) 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

34) Huang, F., Yuan, W., Tian, S., Zheng, Q., He, Y., 2019. SIN3 LIKE genes mediate long-day induction of flowering but inhibit the floral transition in short days through histone deacetylation in Arabidopsis. The Plant Journal 100, 101–113. https://doi.org/10.1111/tpj.14430

35) 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

36) 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

37) 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

38) Jiang, J., Chai, X., Manavski, N., Williams-Carrier, R., He, B., Brachmann, A., Ji, D., Ouyang, M., Liu, Y., Barkan, A., Meurer, J., Zhang, L., Chi, W., 2019. An RNA Chaperone–Like Protein Plays Critical Roles in Chloroplast mRNA Stability and Translation in Arabidopsis and Maize. The Plant Cell 31, 1308–1327. https://doi.org/10.1105/tpc.18.00946

39) 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

40) Konishi, M., Okitsu, T., Yanagisawa, S., 2021. Nitrate-responsive NIN-like protein transcription factors perform unique and redundant roles in Arabidopsis. Journal of Experimental Botany 72, 5735–5750. https://doi.org/10.1093/jxb/erab246

41) 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

42) 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

43) Ma, L., Hatlen, A., Kelly, L.J., Becher, H., Wang, W., Kovarik, A., Leitch, I.J., Leitch, A.R., 2015. Angiosperms Are Unique among Land Plant Lineages in the Occurrence

of Key Genes in the RNA-Directed DNA Methylation (RdDM) Pathway. Genome Biology and Evolution 7, 2648–2662. https://doi.org/10.1093/gbe/evv171

44) Meyer, L.J., Gao, J., Xu, D., Thelen, J.J., 2012. Phosphoproteomic Analysis of Seed Maturation in Arabidopsis, Rapeseed, and Soybean. Plant Physiology 159, 517–528. https://doi.org/10.1104/pp.111.191700

45) 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

46) 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

47) 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

48) 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

49) 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

50) 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

51) 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

52) 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

53) 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

54) 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

55) 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

56) 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

57) 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

58) 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

59) Takeda, T., Shirai, K., Kim, Y.-W., Higuchi-Takeuchi, M., Shimizu, M., Kondo, T., Ushijima, T., Matsushita, T., Shinozaki, K., Hanada, K., 2023. A de novo gene originating from the mitochondria controls floral transition in Arabidopsis thaliana. Plant Mol Biol 111, 189–203. https://doi.org/10.1007/s11103-022-01320-6

60) Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., Handakumbura, P.P., Xiong, G., Wang, C., Corwin, J., Tsoukalas, A., Zhang, L., Ware, D., Pauly, M., Kliebenstein, D.J., Dehesh, K., Tagkopoulos, I., Breton, G., Pruneda-Paz, J.L., Ahnert, S.E., Kay, S.A., Hazen, S.P., Brady, S.M., 2015. An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature 517, 571–575. https://doi.org/10.1038/nature14099

61) 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

62) Wang, Y., Zhang, W.-Z., Song, L.-F., Zou, J.-J., Su, Z., Wu, W.-H., 2008. Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. Plant Physiology 148, 1201–1211. https://doi.org/10.1104/pp.108.126375

63) 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

64) 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

65) 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

66) 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

67) 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

68) 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

69) 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

70) 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

71) 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

72) Zhu, J., Zhang, G., Chang, Y., Li, X., Yang, J., Huang, X., Yu, Q., Chen, H., Wu, T., Yang, Z., 2010. AtMYB103 is a crucial regulator of several pathways affecting Arabidopsis anther development. Sci. China Life Sci. 53, 1112–1122. https://doi.org/10.1007/s11427-010-4060-y

73) 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

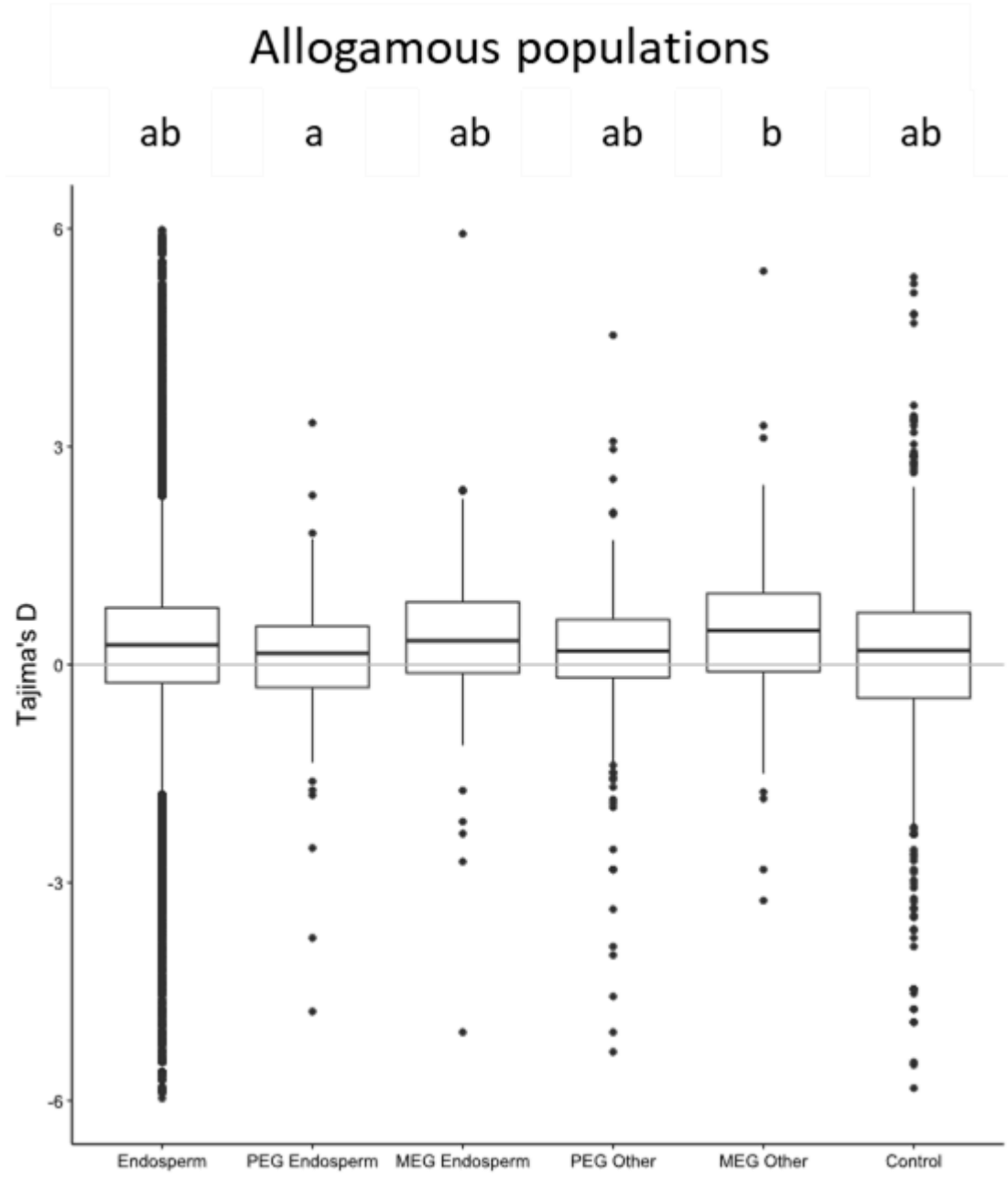

Figure S1: Variations of the Tajima's D in the allogamous populations in the imprinted genes compared to endosperm and control genes. *boxplot of the variations of the median Tajima's D by gene. The light grey line represents the threshold value for neutrality. The minuscule letters (a,b,c) signified the groups significantly different following Dunn's test (p-values < 0.05.).*