Limited directional selection but coevolutionary signals among imprinted genes in *A. lyrata.*

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

[1] IHPE, Univ Perpignan Via Domitia, CNRS, IFREMER, Univ Montpellier, 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 thubfor Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany

[*] Corresponding author: audrey.le-veve@univ.perp.fr

Abstract

Genomic imprinting is a form of gene regulation leading to the unequal expression of maternal and paternal alleles. The main hypothesis invoked to explain the evolution of imprinted genes is the kinship theory, which posits a conflict between parental genomes over resource allocation in progeny. According to this theory, imprinted genes are expected to undergo a shift from balancing to directional selection, in coding or regulatory regions. However, most studies addressing selection in imprinted genes have focused on self-fertilizing species, where conflicts over resource allocation are predicted to be weak. Consequently, the impact of kinship theory on the evolution of imprinted genes remains largely untested. Furthermore, the predicted coevolution between certain imprinted genes under kinship theory has not been tested yet empirically.

Combined phylogenetic and population genomic approaches, we investigated signatures of selection on imprinted genes across the Brassicaceae family and in autogamous and allogamous populations of *Arabidopsis lyrata*, and searched for evidence of coevolution among imprinted genes. We found that endosperm-expressed genes exhibited signals of balancing selection across Brassicaceae and within allogamous populations, consistent with unresolved intralocus conflicts. Supporting the kinship theory, these signals varied with mating system. Moreover, phylogenetic analyses further indicated negative directional selection acting on imprinted genes. However, population-level signals were independent of mating system and showed limited concordance with kinship predictions, possibly due to stronger selection acting on expression than on coding sequences. Finally, we identified

coevolution between imprinted genes, although not at specific sites, suggesting that interactions beyond protein sequence drive this process.

Key words: Genomic imprinting, Kinship theory, Endosperm, Coevolution, Mating system.

Introduction

The kinship theory, or the parental conflict theory, was first proposed by Haig and Westoby (1989) in the context of endosperm development and elaborated in a review in 1991. It posits that in species where the maternal parent nourishes progeny sired by multiple males, maternal fitness is maximized by distributing resources equally among offspring, whereas paternal fitness is enhanced by securing more resources for their own offspring. This conflict is expected to drive intralocus conflict between growth-promoting and growth-limiting genes (Haig and Graham 1991; Moore and Haig 1991; Wilkins and Haig 2001), consequently leaving signatures of balancing selection, such as reduced divergence across species and increased polymorphism within populations (Day and Bonduriansky 2004; Kasimatis et al., 2017). This has been shown in *Arabidopsis thaliana,* whose crosses showed dosage effects on seed size, although these may reflect interactions between cytoplasmic inheritance, dosage imbalance, and imprinting (Scott et al., 1998). Notably, balancing selection in genes expressed in endosperm (embryo's nourishing tissue in the seed) has not been systematically studied to validate the kinship theory.

The kinship theory remains the leading explanation for the evolution of imprinted genes (Haig 1997; Spencer and Clark 2014; Patten et al., 2014). Imprinting is an epigenetic regulation in flowering plants and mammals in which allele expression depends on the parental origin (Haig 1997; Wilkins 2011). By creating hemizygous expression of parental alleles, the kinship theory argues that imprinting evolves to resolve parental conflicts over resource allocation. In plants, many imprinted genes appear to be involved in resource accumulation in the endosperm and metabolic processes (Raissig et al., 2011). Theoretical models using models of selection on one or two genes involved in resource allocation predicted that: (1) imprinting set distinct optimal expression levels for paternally expressed genes (PEGs) and maternally expressed genes (MEGs; Haig 1997); (2) PEGs and MEGs are under directional selection to increase and limit resource acquisition by offspring respectively (Haig 2000); and (3) imprinting at some growth-promoting locus may drive antagonistic coevolution with some growth-suppressing function (Wilkins and Haig 2001). However, these models focus on *IGF2* and *IGF2R* genes, which regulate foetal growth in mammals, and without investigating the expected signatures of selection for endosperm genes.

Empirically, phylogenetic signals of directional selection have been reported for *MEDEA* locus in the allogamous species *Arabidopsis lyrata*, but not in the autogamous sister species *A. thaliana* (Spillane et al., 2007; Miyake et al., 2009), and for PEGs at the population levels in the autogamous Brassicaceae species *Capsella rubella* and *A. thaliana* (Hatorangan et

al., 2016; Tuteja et al., 2019). However, MEGs did not show comparable signals, possibly due to technical challenges in detecting selection if selection primarily targets gene expression levels. Indeed, a central debate has emphasized that kinship theory may concern selection on the optimal expression levels of genes between maternal and paternal genomes, and not on alleles (Patten et al., 2014). Selection acting on gene expression may still leave detectable traces in amino acid or codon usage patterns (Patten et al., 2014).

The incongruent signals detected in imprinted genes may also be due to polymorphism dynamics shaped by the mating system. The kinship theory assumes that progenies are sired by multiple males and that imprinted genes are preferentially expressed in the endosperm (Haig and Westoby 1991; Haig 2000; Brandvain and Haig 2005). The transition to autogamy may alter selection acting on genes, potentially leading to more intense directional selection or relaxed selection (Soliman and Coughlan 2024). However, the empirical studies cited above did not assess endosperm-specific expression of the imprinted genes and only focused on autogamous species. Consequently, the evolutionary signatures of selection on imprinted genes and the contribution of the kinship theory to their evolution remain unresolved in plants.

Kinship theory also predicts a potential coevolution between antagonistic imprinted genes (Wilkins and Haig 2001; Willi 2013), yet no phylogenetic signal was detected for *IGF2* and *IGF2R* (McVean and Hurst 1997), and coevolution among endosperm-expressed genes remains unknown. Population genomic methods can detect coevolving sequences based on allele frequency variation; however, they may be biased by demographic or population structure effects. More integrative approaches based on phylogenetic comparisons can identify co-occurring changes in candidate coevolving genes (Dutheil and Galtier 2007).

Here, we compared imprinted genes in allogamous *A. lyrata* (Klosinska et al., 2016) with non-imprinted endosperm genes, using phylogenetic data from 23 Brassicaceae species and genomic data from nine allogamous populations and nine closely related autogamous populations of *A. lyrata* from North America (Ross-Ibarra et al., 2008). We tested for : (1) variations in expression patterns and/or codon usage; (2) reduced divergence and polymorphism, as expected under directional selection, and (3) coevolution. Our results indicate that endosperm-expressed genes display patterns consistent with intralocus conflict and that transition to autogamy promoted directional selection on these genes. Imprinted genes show evidence of ancient negative directional selection, codon usage bias, and substantial coevolution. However, population-level signals are less consistent with kinship theory, possibly due to stronger selection acting on expression than on coding sequences or by a recent relaxed selection on these genes in North American populations.

Results

Imprinted genes are preferentially expressed in the endosperm

Parental conflict is expected to occur primarily in the endosperm. Using transcriptomic data from 19,425 genes across five tissues (leaf, root, pollen, pistil, and endosperm), we identified 2,732 specifically upregulated genes in the endosperm of *A. lyrata* among 31,073 genes analysed. Among these upregulated genes, 15 were MEGs and 14 were PEGs from the 27 MEGs and 46 PEGs previously identified in *A. lyrata* (Klosinska et al., 2016). The proportions of endosperm-upregulated MEGs (0.56) and PEGs (0.30) were both significantly higher than expected from 10,000 random resamplings of all genes (p values < 1^{e-04}). Moreover, MEGs were significantly more frequently upregulated in the endosperm than PEGs, based on 10,000 resamplings restricted to imprinted genes (p value =0.03).

For subsequent analyses, PEGs and MEGs upregulated in the endosperm are hereafter referred to as PEGs and MEGs (or “imprinted genes” if PEGs and MEGs together), whereas imprinted genes not specifically upregulated in the endosperm were excluded. The 2,703 non-imprinted genes upregulated in the endosperm were used as a reference set (hereafter “endosperm genes”), and 100 randomly chosen non-imprinted ubiquitous genes were used as a control set.

Between *A. lyrata* and *A. thaliana*, 40% of MEGs (6/15) and 50% of PEGs (7/14) identified in our dataset were conserved as imprinted (Table S1).

To test whether imprinting influences gene expression levels, we compared normalized expression values between imprinted and endosperm genes, adjusted for the number of expressed copies (three expressed copies for non-imprinted endosperm genes, two for MEGs, and one for PEGs). Both PEGs and MEGs displayed significantly higher expression levels relative to endosperm genes, with 3.18- and 1.71-fold increases, respectively (bootstrap tests; p-values < 1^{e-03}; Fig 1).

We next assessed whether this increased expression was associated with codon or amino acid usage bias. Comparisons between imprinted and endosperm genes revealed several significant deviations in codon and amino acid usage (Table S2). Only two biases were shared between PEGs and MEGs: the under-usage of GTC (valine) and the over-usage of tryptophan. To test whether codon over-usage was associated with expression variation, we compared PHE scores (an index of codon impact on expression, derived from overexpressed genes in *A. thaliana*; Sahoo et al., 2019) of biased versus unbiased codons using 1,000 permutations. Contrary to our hypothesis, PHE scores did not differ significantly

between groups of codons (MEGs: p value = 0.09; PEGs: p value = 0.37). Similarly, no significant differences were observed when comparing codons that were underrepresented in imprinted genes (MEGs: p value = 0.44; PEGs: p value = 0.36).

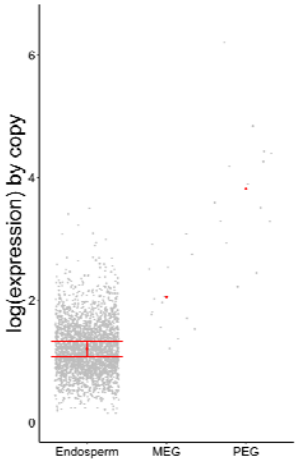

Figure 1: Imprinted genes are associated to higher expression level by gene copies than in endosperm genes. *The grey dots represent the mean expression level obtained for one gene. The red dots represent the mean expression level for each group of gene. The red lines represent 5% and 95% of the means obtained for endosperm genes after 1,000 resampling of 29 genes (number of imprinted genes considered in study). If the observed means of PEG and MEG are outside of these red lines, they are considered significantly different from the mean in endosperm genes, following the bootstrap approach.*

The MEGs and PEGs exhibit phylogenetic signals of negative selection in Brassicaceae

We first compared gene size, coding sequence (CDS) content (%), and GC content (%) of imprinted genes in *A. lyrata* with those of endosperm and control genes, using bootstrap tests, as these features may influence the levels of polymorphism. Control genes and MEGs were significantly smaller than PEGs and endosperm genes (Table S3). In addition, CDS content was significantly higher in endosperm and imprinted genes compared to control genes, while GC content was significantly lower in imprinted genes compared to control genes. These differences were taken into account when interpreting the subsequent results.

To explore phylogenetic patterns, we analysed imprinted genes in *A. lyrata* and 21 additional Brassicaceae species (Table S4) relative to endosperm and control genes. As a proxy for gene turnover, we compared the number of species with homologous sequences across groups. This number was significantly higher for endosperm genes than for other categories, and significantly higher for PEGs than for control genes (Fig 2; Table S5).

We then assessed phylogenetic congruence using Robinson–Foulds (RF) distances. We compared the topology of each gene tree with that of the species tree derived from endosperm genes present in our Brassicaceae. RF distances did not differ significantly among imprinted and endosperm gene groups (bootstrap; p value = 0.60 for PEGs; p value = 0.34 for MEGs). Therefore, we used the species tree topology to estimate branch lengths

(proxy for evolutionary rates; Fig. S1) for genes having homologs in at least three species (including *A. lyrata*).

The length of all branches found in the imprinted genes varied significantly among groups of genes (Fig 2; bootstrap; Table S5). Both endosperm and imprinted genes displayed significantly longer branch lengths than control genes. However, MEGs exhibited significantly shorter branch lengths than endosperm genes.

Finally, we assessed mean D_{NS}/D_S ratios between *A. lyrata* and the other Brassicaceae species. A D_{NS}/D_S increase indicates positive directional selection whereas a D_{NS}/D_S decrease indicates purifying (negative) selection. Imprinted genes exhibited significantly lower D_{NS}/D_S ratios than both endosperm and control genes (Table S5).

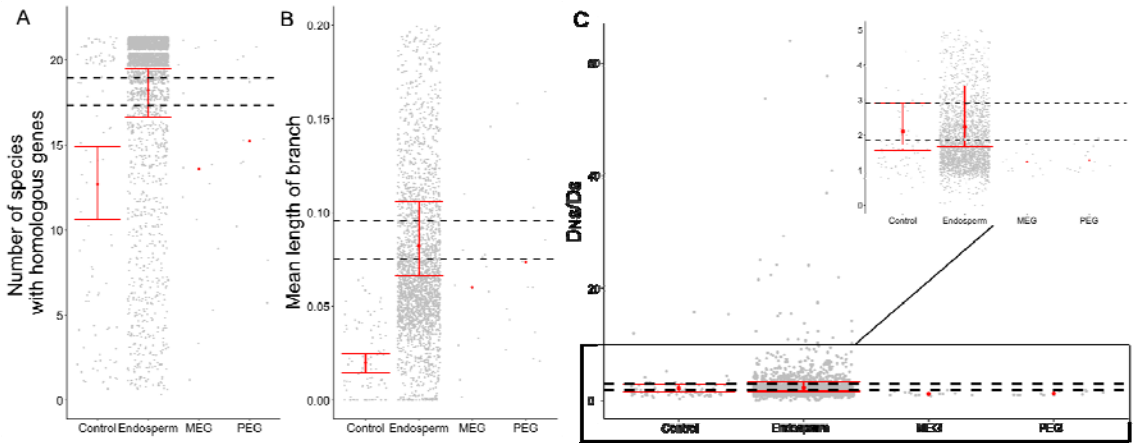

Figure 2: Imprinted genes present phylogenetic signal of negative selection compared to endosperm genes across the Brassicaceae. *A: Variations of the number of species with homologous genes by group of genes. B: Variations of the mean branch length by gene based on the phylogenetic tree obtained by HKY85 model based on the 1055 endosperm genes found in all species. C: Variations of the D_{NS}/D_S by group of genes with a focus on values lower than 5 (top right). Two extremes values (>200) were removed. The grey dots represent the mean value obtained for one gene. The red dots represent the mean in each group of gene. The dashed lines represent 5 and 95% of the means obtained for endosperm genes after 1000 resamples of 100 genes (size of group of control genes). If the real mean of control genes was not between these dashed lines, the value was considered as significantly different than in endosperm genes, following bootstrap approach. The red lines represent 5 and 95% of the means obtained for endosperm and control genes after 1000 resamples of 29 genes (number of imprinted genes considered in study). If the real means of*

PEG and MEG were not between these red lines, the values were considered as significantly different than in endosperm or control genes, following bootstrap approach.

Increase of π_{NS}/π_S not related to the mating system detected in imprinted genes

To evaluate the effect of mating system and imprinting on polymorphism, we compared Tajima's D, π and the π_{NS}/π_S for imprinted, endosperm and control genes in nine allogamous and nine autogamous populations (Table S6). A previous study reported an effect of mating system on seed size mediated by pollen in these populations (Willi, 2013), suggesting decreased intensity of parental conflict in autogamous populations.

Under directional selection, reductions of Tajima's D, π and π_{NS}/π_S are expected, whereas balancing selection is expected to increase these statistics. Because the calculation of Tajima's D and π_{NS}/π_S requires sufficient polymorphism, genes with fewer than 0.5% polymorphic sites in all populations were excluded from Tajima's D analyses, and genes with $\pi_S = 0$ in all populations were excluded from π_{NS}/π_S analyses. For both statistics, the proportion of endosperm genes retained the analyses was significantly higher than for control genes (binomial tests; Table S7).

In allogamous populations, the Tajima's D was significantly higher in endosperm genes compared to control genes and PEGs (Fig 3; Table S8). However, the π and the π_{NS}/π_S did not differ significantly from control genes. Imprinted genes showed an overall increase in π_{NS}/π_S compared to both endosperm and control genes, while MEGs displayed a significantly lower π relative to control genes.

In autogamous populations, the increase of Tajima's D in endosperm genes relative to control genes and PEGs observed in allogamous populations was no longer detected (Table S8). Tajima's D decreased significantly in endosperm genes in autogamous populations, whereas no effect of mating system was observed for the other gene categories (Kruskal-Wallis tests; Table S9). MEGs showed a significantly higher Tajima's D than endosperm genes, although this difference was not explained by mating system. MEGs in autogamous populations also exhibited a reduced π compared to control genes, and endosperm genes overall had significantly lower polymorphism than control genes (Table S8). These trends were further associated with significant decreases of π_{NS}/π_S for MEGs and endosperm genes compared to control genes (Table S8), with a significant reduction of π_{NS}/π_S in allogamous populations compared to autogamous populations only for endosperm genes (Table S9).

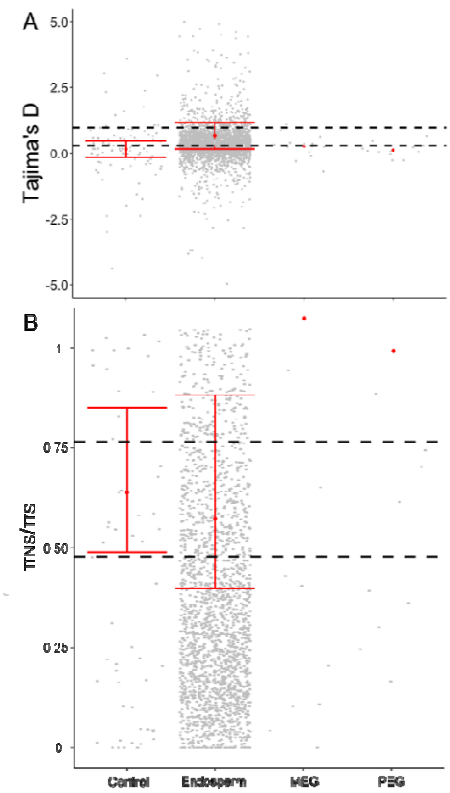

Figure 3: Imprinted genes present an unusual decrease of Tajima's D associated to an increase of π_{NS}/π_S in allogamous populations compared to endosperm genes. *A: Variations of the Tajima's D by group of genes. Only the genes with a mean between -5 and 5 are indicated. B: Variations of the π_{NS}/π_S by group of genes. Only the genes with a mean between 0 and 1.1 are indicated. The grey dots represent the mean value obtained for one gene. The red dots represent the mean in each group of gene. The dashed lines represent 5 and 95% of the means obtained for endosperm genes after 1000 resamples of 100 genes (size of group of control genes). If the real mean of control genes was not between these dashed lines, the value was considered as significantly different than in endosperm genes, following bootstrap approach. The red lines represent 5 and 95% of the means obtained for endosperm and control genes after 1000 resamples of 29 genes (number of imprinted genes considered in study). If the real means of PEG and MEG were not between these red lines, the values were considered as significantly different than in endosperm or control genes, following bootstrap approach.*

A limited number of MEGs exhibited strong signals of directional selection.

We identified imprinted genes potentially under strong directional selection by comparing their mean branch length in their phylogenetic trees, mean Tajima's D and mean π_{NS}/π_S in allogamous populations to the lower 5% of the corresponding values in endosperm genes (Table S1).

Two MEGs were associated with a reduced mean branches length (Table S1). One lacked homologous gene in *A. thaliana*. The second encoded an AGAMOUS-like protein expressed during proliferative endosperm development (Day et al., 2008), which is involved in female gametophyte, seed and early endosperm development (Bemer et al., 2010; Zhang et al., 2018).

One MEG was identified with a lower mean Tajima's D in allogamous populations. Its *A. thaliana* homolog encodes an F-box domain–containing protein regulated by non-CG DNA methylation (Garro et al., 2025) and associated with the regulation of stress-responsive

genes (Papdi et al., 2008). In addition, one MEG encoding a Dof-type zinc finger protein involved in seed coat formation (Zou et al., 2012) showed a reduced π_{NS}/π_S.

Finally, we tested whether the proportion of imprinted genes identified as candidates for directional selection exceeded 5% of the endosperm genes by performing independent binomial tests for each statistic, but none of these proportions were significantly higher than 5% (p values binomial tests = 0.66, 1 and 1 respectively).

Clear trend of coevolution among imprinted genes in Brassicaceae

An important prediction of the kinship theory is that imprinting could promote coevolution between genes favoring paternal or maternal fitness in an arms race scenario. To test this hypothesis, we investigated potential coevolutionary signals between PEGs, MEGs, and endosperm genes. Specifically, we assessed correlations based on (i) nucleotide diversity (π) within populations, (ii) branch length in phylogenetic trees, and (iii) Robinson–Foulds (RF) distances between tree topologies. Because the coevolution hypothesis is more closely related to noncoding regions, we focused on π rather than π_{NS}/π_S.

First, correlations of π were estimated for each pair of PEGs, MEGs, and endosperm genes using Pearson's correlation coefficient. The top 1% of negative ($\pi = -0.66$) and positive ($\pi = 3.20$) correlations among endosperm genes were then used as thresholds to identify significant correlations involving imprinted genes. Based on these thresholds, we identified 12 significant correlations among 6 PEGs and 9 MEGs. Both the number of correlations between PEGs and MEGs (13) and the number of imprinted genes involved (15) were significantly higher than expected from 1,000 random samplings of 29 endosperm genes (p-values= 4.5^{e-02} and $<1^{e-03}$ respectively; Table S10).

When considering endosperm genes significantly correlated with these imprinted genes, as well as indirect links between imprinted and endosperm genes, the resulting network contained 14 MEGs, 13 PEGs, and 1,643 endosperm genes (Fig 4.A; Table S10). This network did not show enrichment for specific GO terms compared to random networks of equal size (1,670 genes). Notably, one PEG encoding a chloroplast-localized protein required for embryonic development (Bryant et al., 2010) was correlated with more genes than 99% of endosperm genes (Table S1 and S10).

Secondly, following the same methodology, thresholds to identify significant correlations between branches length involving imprinted genes were estimated at −0.64 and 0.89. Based on them, we identified 16 significant correlations among 6 PEGs and 9 MEGs. Both the number of correlations between PEGs and MEGs (16) and the number of imprinted

genes involved (15) were significantly higher than expected from 1,000 random samplings of 29 endosperm genes (p-values = 8^{e-03} and 3^{e-03} respectively; Table S11).

When considering endosperm genes significantly correlated with these imprinted genes, the resulting network contained 14 MEGs, 14 PEGs, and 1,027 endosperm genes (Fig 4.B; Table S11). The complete network was significantly enriched for the GO categories "*negative regulation of catalytic activity"* and "*syntaxin binding"* compared to random samples of endosperm and imprinted genes of equal size (1,055 genes; Table S12). Within this network, one MEG encoding the AGAMOUS-like protein previously mentioned was correlated with more genes than 99% of endosperm genes (Table S1; Table S11).

Finally, we investigated gene pairs that showed correlated deviations from the species topology. Genes whose genealogies closely matched the global topology of endosperm genes were excluded, retaining only those with RF distances above the 95th percentile of the distribution obtained for endosperm genes (RF = 0.71; Fig. S1). Among these, we identified pairs including imprinted genes with null RF distances as significantly correlated in topology. Based on this approach, the topology of one PEG encoding an F-box protein, was significantly correlated with six endosperm genes (Fig. 4.C; Table S13). This small network of seven genes was significantly enriched for the GO term *DNA-binding transcription factor activity* compared to random samples of equivalent size (Table S12). Moreover, the PEG exhibited a significantly higher number of associations than 99% of non-imprinted endosperm genes (Table S1; Table S13).

Thus, among the possible pairs of imprinted genes, 29 (including 22 PEG/MEG pairs) were directly associated in the networks based on π or branch length. To test for coevolution between specific mutations within these pairs, we applied the clustering approach implemented in the CoMap package (Dutheil and Galtier, 2007). Specifically, we searched for groups of 2 to 10 coevolving sites within each candidate pair of imprinted genes. To limit false positives, the number of such groups detected in candidate pairs was compared to the number observed in non-candidate pairs of imprinted genes that were not associated in the previous networks and were present in at least three species. Across all group sizes considered, no significant increase in the number of coevolving sites was detected in candidate pairs relative to non-candidate pairs (Table S14). These directly correlated pairs of imprinted genes were therefore not associated with elevated levels of site-level coevolution.

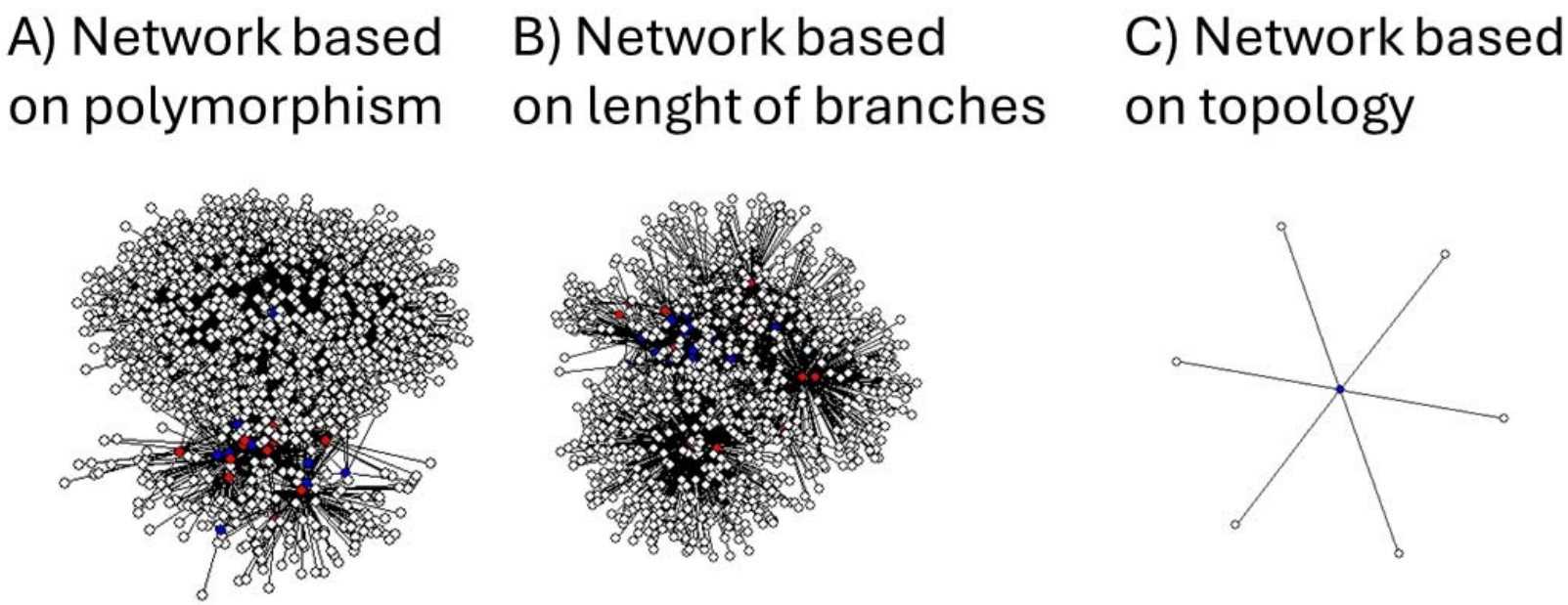

Figure 4: Networks of the imprinted and endosperm genes based on A) polymorphism in population , B) length of branches and C) phylogenetic topologies across the Brassicaceae. *A) Network based on the Pearson's correlations of π found on genes in the populations. B) Network based on the Pearson's correlations 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 dots: PEG. Red dots: MEG. White circles: non-imprinted endosperm genes.*

Discussion

The kinship theory remains the most prominent selective framework for explaining the evolution of genomic imprinting (Patten et al., 2014). According to this theory, imprinted genes are expected to be involved in resource allocation within systems with multiple mating (Haig 2000). Such genes are thought to mitigate parental conflict, thereby producing phylogenetic signatures of directional selection when compared with non-imprinted endosperm genes (Day and Bonduriansky 2004; Kasimatis et al., 2017). However, if selection primarily acts on gene expression rather than coding sequence, these signatures may be attenuated, and their sensitivity to mating system shifts is not well established. Additionally, kinship theory predicts potential coevolution among certain imprinted genes (Wilkins and Haig 2001; Willi 2013). We tested these predictions, previously tested only in autogamous species (Miyake et al., 2009; Hatorangan et al., 2016; Tuteja et al., 2019), by comparing imprinted and non-imprinted endosperm genes in *A. lyrata*, an allogamous species that shift to autogamy in some population from North America, integrating both phylogenetic and population genomic approaches, and assessing the impact of mating system shifts. A synthesis of all hypotheses tested is provided in Table 1.

Signals of ancient balancing selection in endosperm genes

In the absence of genomic imprinting, genes involved in resource allocation conflict are expected to evolve under balancing selection (Day and Bonduriansky 2004; Sylvestre et al., 2023). Consistently, non-imprinted endosperm genes were more widely conserved across Brassicaceae, suggesting lower turnover, and displayed longer phylogenetic branches (Edwards, 2009; Kasimatis et al., 2017; Llaurens et al., 2017). Interestingly, these genes are also associated to an increase of size, a characteristic associated to higher phylogenetic genes conservation (Lopes et al,. 2021). However, this characteristic, as well as the increase of CDS content found in endosperm genes, are more related to reduced branch length and mutation rates (Abdel-Haleem, 2007; Massey, 2013; Koonin, 2015), suggesting that the signal of balancing selection may be underestimated.

In addition, these genes showed elevated Tajima's D values in allogamous populations. However, this increase should be interpreted cautiously. First, Tajima's D is a relatively weak statistic for detecting selection in pooled sequencing data (Korneliussen et al., 2013), and although we minimized dataset-specific biases by comparing with equally affected control genes, technical noise cannot be fully excluded. Second, the absence of a parallel increase in polymorphism suggests that this signal may not solely reflect balancing selection.

Conflict over resource allocation is predicted only in multiple mating systems (Haig 2000). This conflict is resolved in autogamous systems, potentially allowing directional selection on endosperm genes. In line with this, Tajima's D and π_{NS}/π_S decreased in autogamous populations, indicating that relaxation of conflict facilitates directional selection and that the effect of mating system is detectable even in populations that have recently and partially transitioned to autogamy (Ross-Ibarra et al., 2008; Foxe et al., 2010; Hough et al., 2013; Willi et al., 2018).

Codon usage bias and modulation of expression in imprinted genes

Kinship theory predicts that imprinted genes should be preferentially expressed in the endosperm (Haig 2000; Patten et al., 2014). Consistent with this prediction, imprinted genes were enriched among endosperm-expressed genes. Interestingly, the MEGs are more related to endosperm than PEGs, as observed in maize (Eaux et al., 2013). Yet, more than half were also upregulated in other tissues suggesting that distinct selective pressures unrelated to parental conflict may act on them. Some of these imprinted genes may have been inadvertently excluded from the analysis due to the method used to identify differentially expressed genes by tissue, which relied on RNA-seq data pooled from both allogamous and autogamous populations. Future studies using more targeted and controlled experimental designs will be necessary to accurately determine the proportion of imprinted genes that are specifically expressed in the endosperm.

An important debate suggests that detecting directional selection in imprinted genes is difficult because selection may primarily target gene expression (Patten et al., 2014). Consistent with this idea, both PEGs and MEGs showed increased expression per gene copy, even though the increased gene size in PEGs could be due to a less efficient expression (Grishkevich and Yanai, 2014). Moreover, imprinted genes displayed codon usage biases relative to endosperm genes. This could be explained by the increase of GC content in these genes (Sharp and Li, 1987). Even if an increase of GC content was associated to an increase of expression (Bernardi, 2000), the codon usage scores did not enable us to determine whether these biases enhance or inhibit the expression of imprinted genes (Sahoo et al., 2019).

Table 1: Summary of the expectations on phylogenetics characteristics and polymorphism on non-imprinted and imprinted endosperm genes under kinship theory. *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.*

	Endosperm compared to control	**Imprinted compared to endosperm**	**References**
Expression levels	NA	Upregulated	Patten et al., 2014
Codon and amino acid bias	NA	Yes if selection on expression	Patten et al., 2014
Number of homologous genes across species	Increase (less turn over with intralocus conflict)	Decrease (more turnover by high evolutionary rate)	Haig 2000; Spencer 2000
Length of branches	Increase	Increase if positive, decrease if negative	Haig 2000; Spencer 2000
D_{NS}/D_S	No variation	Increase if positive, decrease if negative	Spencer 2000
Tajima's D in allogamous population	Increase (intralocus conflict)	Decrease (directional selection)	Spencer 2000; Day and Bonduriansky 2004
π in allogamous population	Increase (intralocus conflict)	Decrease (directional selection)	Spencer 2000; Day and Bonduriansky 2004
$\pi NS/\pi S$ in allogamous population	No variation	Decrease	Spencer 2000; Day and Bonduriansky 2004
Variation of $\pi NS/\pi S$ and Tajima's D with mating system	Yes (decrease with autogamy)	No (already fixed) or increase (Relaxed selection)	Haig 2000; Day and Bonduriansky 2004
Coevolution between genes	NA	Possible	Moore and Haig 1991

Phylogenetic signals of ancient purifying selection in imprinted genes.

The kinship theory is expected to promote directional selection on imprinted genes (Wilkins and Haig 2001; Mills and Moore 2004; Day and Bonduriansky 2004; Kasimatis et al., 2017). Consistently, the imprinted genes are associated with an increase of CDS content, usually found in genes under strong purifying selection (Abdel-Haleem, 2007; Koonin, 2015). Moreover, our phylogenetic analyses showed that imprinted genes exhibit lower conservation of homologs across Brassicaceae species, despite the higher size of PEGs that is expected to reduced turnover (Massey, 2013; Lopes et al,. 2021), likely reflecting a higher turnover driven by accelerated evolutionary rates. Furthermore, imprinted genes displayed reduced D_{NS}/D_S, indicative of strong negative directional selection.

This signal of purifying selection was also evident in MEGs when branch lengths were compared to those of endosperm genes. We note that two MEGs, including AGAMOUS-like protein implicated in one coevolutionary network, exhibited branch lengths suggesting a particularly strong negative selection. 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), raising the question of the effect of the mating system on the selection of imprinted genes.

Nevertheless, caution is warranted in attributing the observed purifying directional selection solely to imprinting. First, the set of imprinted genes analysed was derived from a study on two geographically and genetically distinct allogamous subspecies (European *A. lyrata petraea* and North American *A. lyrata lyrata*; Klosinska et al., 2016), assuming conservation of imprinting, because no within-lineage data on imprinting in *A. lyrata* exist. Yet, imprinting is poorly conserved within *A. thaliana* (Gehring et al., 2011; Wolff et al., 2011; Hsieh et al., 2011; Pignatta et al., 2014), between *A. lyrata* and *A. thaliana* in our study and more generally across species (Eaux et al., 2013; Hatorangan et al., 2016; Roth et al., 2018; Flores-Vergara et al., 2020). It is therefore possible that the "imprinted genes" studied are actually not imprinted in our populations due to important variations on transposable element promoting imprinting (Padilla-García et al., 2025). Future population-specific surveys of imprinting will be critical to resolve this issue.

Secondly, the detected purifying selection could be explained by alternative explanations (Patten et al., 2014), including (1) an increase of the efficiency of selection by hemizygous expression (Haig and Westoby 1989), (2) selection against transposable element (Padilla-García et al., 2025) or (3) local adaptation (Kremer et al., 2025). However, none of the

current models rigorously predict the effects of these mechanisms on polymorphism and divergence, making it difficult to discriminate among alternative explanations.

Implications for reproductive isolation and adaptation

The evolutionary consequences of purifying selection on imprinted genes remains poorly understood. Directional selection on imprinted genes could (1) influence adaptive potential by shaping the spectrum of selectable alleles within populations (Cvijović et al., 2018), (2) contribute to speciation processes by generating genetic incompatibilities between lineages (Seehausen et al., 2014), (3) facilitate the transition to apomixy in plants (Hojsgaard and Schartl 2021) and (4) promote the establishment of reproductive isolation (Maheshwari and Barbash 2011; Cvijović et al., 2018; İltaş et al., 2021), more particularly through endosperm-based post-zygotic barriers (Haig 2000; Spencer 2000; Kinoshita 2007; Köhler et al., 2010).

Previous studies empirically demonstrated that some imprinted genes are involved in these barriers, including in the autogamous species *A. thaliana* (Wolff et al., 2015), between the allogamous *Solanum peruvianum* et *S. chilense* (Florez-Rueda et al., 2016), between the allogamous *C. grandiflora* and the autogamous *C. rubella* and *C. orientalis* (Lafon-Placette et al., 2018), between the allogamous *Mimulus guttatus*, *M. luteus* (Kinser et al., 2021), *M. caespitosa* and *M. tilingi* (Sandstedt and Sweigart 2022), and between the autogamous wheat species *Triticum durum* and *T. aestivum* (Yang et al., 2023).

Imprinted genes may also act as central hubs within complex regulatory networks that contribute to hybrid vigor and reproductive barriers (Botet and Keurentjes 2020). However, despite these observations, the broader evolutionary implications of these network-level effects remain largely unresolved, particularly regarding how they interact with demographic history, mating system variation, and other selective pressures.

Imprinted genes are associated with singular signals of selection in North America

Consistent with the expectations under the kinship theory, PEGs were associated with reduced Tajima's D (Haig 2014). However, this signal was not observed for π_{NS}/π_S. More interestingly, π_{NS}/π_S values were higher in both PEGs and MEGs compared to endosperm and control genes, which could indicate a relaxation of selection on coding regions (Galtier and Rousselle 2020). This apparent relaxation may not reflect parental conflict but instead could result from the recent colonization of North America by *A. lyrata* from Europe (Ross-Ibarra et al., 2008; Otto et al., 2015; Willi et al., 2018) that could have reduced effective population size (Galtier and Rousselle 2020) and the efficiency of purifying selection (Laenen et al., 2018), especially on transposable element (Padilla-García et al., 2025).

Testing this hypothesis would require future comparisons of selection signals in endosperm and imprinted genes across allogamous populations with varying levels of polymorphism. Nevertheless, we note that two MEGs exhibited reduced Tajima's D or π_{NS}/π_S, consistent with ongoing directional selection.

Imprinted genes are not sensitive to shifting mating system in A. lyrata

Conflict over resource allocation is expected to be reduced in autogamous lineages relative to allogamous ones (Haig 2000; Brandvain and Haig 2005). However, most studies of genetic and phylogenetic signatures of selection associated with imprinting have focused on autogamous species, leaving the effect of shifting mating system on polymorphism of imprinted genes unresolved. In the autogamous species *A. thaliana*, the imprinted *MEDEA* locus shows increased polymorphism but not in its allogamous sister species *A. lyrata* (Spillane et al., 2007; Miyake et al., 2009). However, PEGs were associated to positive selection in the autogamous species *A. thaliana* and *C. rubella* (Hatorangan et al., 2016; Tuteja et al., 2019). In our study, the selection signals in imprinted genes, contrary to endosperm genes, were insensitive to the mating system. The absence of variation in Tajima's D could be explained if directional selection had already fixed the optimal alleles prior to the transition from allogamy to autogamy, which could also explain the signal of positive selection detected in PEGs in *A. thaliana* and *C. rubella* (Hatorangan et al., 2016; Tuteja et al., 2019). However, this explanation does not account for the high π_{NS}/π_S observed in allogamous and autogamous populations. Theoretical studies are necessary to estimate the cumulated effect of selection related to kinship theory, imprinting, demographic history and mating system.

Imprinted genes present signal of ancient and recent coevolution

Another potential, but not mandatory, predicted outcome of the kinship theory is the coevolution of some imprinted genes with antagonistic effects on offspring growth (Moore and Haig, 1991; Wilkins and Haig 2001; Mills and Moore, 2004). To investigate this, we applied both population genomic and phylogenetic approaches (Dutheil and Galtier 2007). We found that most PEGs and MEGs were connected within coevolutionary networks, with a significantly higher number of direct associations than expected by chance. These networks encompass essential functions for endosperm development.

For instance, one network based on branch-length correlations was enriched in enzymatic activities related to cell wall formation and modification, a key process during endosperm cellularization in Arabidopsis (Hehenberger et al., 2012), which in turn influences seed size

(Pignatta et al., 2018; Zhang et al., 2020), a likely target of parental conflict. These correlated evolutionary rates could reflect equivalent selective constraints (Galtier and Rousselle 2020) or just a common chromosomic context (e.g. transposable element) that modify mutation rate (Lynch 2010). Interestingly, this network appeared to be strongly dependent on a MEG, an AGAMOUS-like protein characterized by reduced branch lengths, consistent with expectations that coevolution may drive coordinated evolution among genes primarily through rare compensatory mutations (Fraser et al., 2002)

The polymorphism-based network involved genes related to syntaxin-binding proteins, crucial for membrane activity, molecular transport, and protein storage, whereas the topology-based network was enriched in DNA-binding transcription factors, including those regulating seed development, tissue differentiation, and storage metabolism (Noguero et al., 2013).

These signals of coevolution could promote the maintenance of the gene functions (Clark et al., 2012), limit their evolutionary independency (Yang and Bielawski 2000) and be another source of reproductive barrier (Presgraves 2010). Interestingly, the coevolution signal did not involve increased site-by-site coevolution, suggesting that protein–protein molecular interactions, often drivers of coevolution (Green et al., 2021), played a limited role here. Instead, the observed patterns may reflect selection acting on gene dosage rather than protein products (Clark et al., 2012).

Conclusions

Overall, our analyses revealed that endosperm-expressed and imprinted genes in *A. lyrata* exhibit evolutionary signatures partly consistent with kinship theory, including evidence for ancient balancing selection in endosperm genes, expression-level regulation in imprinted genes, ancient purifying selection, and coevolutionary interactions. However, signals at the population level were less congruent with theoretical expectations, likely reflecting demographic history, mating system shifts, and the poor conservation of imprinting across populations. Future work should combine empirical surveys of imprinting with theoretical models explicitly incorporating demography and mating system variation to refine predictions about the evolutionary dynamics of imprinted genes.

Materials and Methods

We hypothesized, based on kinship theory, that imprinted genes should display distinct genomic signatures of selection compared to non-imprinted endosperm genes (Table 1). To test this prediction, we designed an approach combining phylogenetic analyses and polymorphism data. Specifically, we focused on a set of 27 maternally expressed genes (MEGs) and 46 paternally expressed genes (PEGs) previously characterized in *A. lyrata* (Klosinska et al., 2016) and unambiguously annotated in the reference genome (Hu et al., 2011; https://doi.org/10.5281/zenodo.17287897). These imprinted genes were systematically compared against two control datasets: (i) 2,703 non-imprinted genes upregulated in the endosperm, and (ii) a reference panel of 100 randomly selected ubiquitous genes, expressed across multiple tissues. This design allowed us to disentangle signatures of selection specifically associated with imprinting from those affecting endosperm genes more generally or the whole genome.

Expression data

RNA-seq expression data from multiple tissues, including leaf, root, pistil, and pollen, were obtained from two *A. lyrata* populations representing 27,820 genes from both European and North American lineages, as described in İltaş et al. (2024). To maximize gene representation and increase the statistical power, the data from both mating systems was pooled. To complement these datasets, we also incorporated endosperm RNA-seq data derived from five allogamous *A. lyrata* individuals, previously generated by Klosinska et al. (2016). The endosperm data were reprocessed and integrated into our analytical pipeline following the instructions and R script available in GitHub (https://github.com/thelovelab/DESeq2). To identifying endosperm-specific genes, we retained only those that showed significant upregulation (log2FC > 1, adjusted P < 0.05) in all pairwise comparisons against other tissues. The RNA-seq datasets for leaf, root, pistil, and pollen are available in the NCBI BioProject database under accession PRJNA1067428, while the endosperm RNA-seq dataset is accessible under accession GSE76076.

Genes clusters

We used the list of the 49 PEGs and 35 MEGs found in *A. lyrata* (Klosinska et al., 2016). We kept the genes having conserved coordinates in the reference genome of *A. lyrata* V1.0.23 (Hu et al., 2011; https://doi.org/10.5281/zenodo.17287897), resulting in a final list of study 27 MEGs and 46 PEGs.

To disentangle signatures of selection associated specifically with the endosperm, where conflicts over resource allocation are expected, from those potentially arising through other selective processes, we separated imprinted genes upregulated in the endosperm (14 PEGs and 15 MEGs) from imprinted genes not showing endosperm-specific upregulation (32 PEGs and 12 MEGs), as determined by differential expression analysis. To evaluate whether imprinted genes were significantly enriched among those upregulated in the endosperm, we performed 10,000 resamples of 49 and 35 genes between the 27,820 genes with available transcriptomic data for PEGs and MEGs respectively to determine how many upregulated in endosperm could be expected. In addition, we tested for significant differences in the number of upregulated genes between PEGs and MEGs comparing the number of PEGs upregulated in endosperm observed with the number found in a second series of 10,000 random resamples of 35 genes (size of MEGs group) between the 27,820 genes with available transcriptomic data. For all subsequent analyses, we focused exclusively on endosperm-upregulated imprinted genes and compared them with two reference groups: the set of 2,703 non-imprinted endosperm-upregulated genes (hereafter, “endosperm genes”), the sets of 15 MEG and 14 PEG upregulated in endosperm (hereafter, “MEG” and “PEG”) and a set of 100 randomly selected ubiquitously expressed genes (hereafter, “control genes”).

We then conducted bootstrap analyses to compare the normalized expression (read counts per gene copy) of PEGs and MEGs upregulated in the endosperm with expression levels obtained from 10,000 resamples of 29 non-imprinted endosperm genes in European allogamous populations. To account for gene dosage, expression values were normalized by the expected number of expressed copies: three for non-imprinted endosperm genes, two for MEGs, and one for PEGs.

Based on the sequences from the reference genome of *A. lyrata* (Hu et al., 2011; https://doi.org/10.5281/zenodo.17287897), we conducted bootstrap analyses to compare the gene size, the GC and the CDS content (%) of PEGs and MEGs with values obtained from 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we also conducted bootstrap analyses to compare the values on control genes with values obtained from 10,000 resamples of 100 endosperm genes.

The R code for the bootstrap tests is available in GitHub (https://github.com/leveveaudrey/bootstrap).

Estimation of codon and amino acid bias

Based on the sequences from the reference genome of *A. lyrata* (Hu et al., 2011; https://doi.org/10.5281/zenodo.17287897) and a custom python code (https://github.com/leveveaudrey/codon-bias), we estimated the proportion of each amino acid within the gene and the proportion of each possible codon for each amino acid. Then, we conducted bootstrap analyses to compare these proportions in PEGs and MEGs with values obtained from 10,000 resamples of 29 endosperm genes. The R code for the bootstrap test is available in GitHub (https://github.com/leveveaudrey/bootstrap). Thus, we defined the codons "overused" as codons more used in imprinted genes than in endosperm genes, the codons "underused" as codons less used in imprinted genes than in endosperm genes and the "unbiased" codons as codons equivalently used in imprinted genes than in endosperm genes.

To test if the overused or underused codons were related to gene expression, we compared the variations of PHE score induced by codons with the score of unbiased codons by permutations test with 1,000 replications for MEGs and PEGs independently. The PHE score was estimated based on overexpressed genes in *A. thaliana* and reflect the importance of a codon on genes expression (Sahoo et al,. 2019). We hypothesized that codons overused in imprinted genes were associated to increases of PHE score because associated to increases of gene expression whereas the underused codon were associated to reduce PHE scores. The code used for permutation test was available in GitHub (https://github.com/leveveaudrey/permutation-test-)

Distribution of the genes across Brassicaceae

We assessed whether the imprinted genes presented specific signals of selection compared to the endosperm and control genes in the phylogeny of 22 Brassicaceae. All the complete sequence of reference genomes (fasta files) were extracted from the Phytozome database (Goodstein et al., 2012; Table S4). 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 retained the best hit of these sequences on 22 other Brassicaceae using BLAST V2.10 (Altschul et al., 1990) with e-value < $1^{e-}5$. If more of one hit was 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 in *A. lyrata* and the higher reciprocal coverage on the gene in the other species. Then we extracted the sequence of the best hit for each species of each gene. These steps were integrated into a pipeline of analysis available in GitHub (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee).

To test for significant differences in the number of species with homologous genes across the group of genes, we conducted bootstrap analyses to compare the value on PEGs and MEGs with values obtained from 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we conducted bootstrap analyses to compare the values on control genes with values obtained from 10,000 resamples of 100 endosperm genes. The R code for the bootstrap test is available in GitHub (https://github.com/leveveaudrey/bootstrap).

Divergence of the imprinted genes across Brassicaceae

The rates of nonsynonymous mutations (D_{NS}) and synonymous mutations (D_S) are 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 indicates positive selection, while a decrease indicates negative selection. To test whether imprinted genes display accelerated evolutionary rates compared to endosperm genes, using the annotation of the 0-fold and 4-fold degenerate sites in *A. lyrata*, we estimated for each gene the mean evolutionary rates D_{NS}/D_S between the reference genome of *A. lyrata* and all the species with an homologue found. Only the genes found in at least 3 species and with a D_S higher than 0 were considered. These steps were integrated into a pipeline of analysis available in GitHub (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee).

To test for significant differences in the D_{NS}/D_S across the group of genes, we conducted bootstrap analyses to compare the value on PEGs and MEGs with values obtained from 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we conducted bootstrap analyses to compare the values on control genes with values obtained from 10,000 resamples of 100 endosperm genes. The R code for the bootstrap test is available in GitHub (https://github.com/leveveaudrey/bootstrap).

Topology of the genes across Brassicaceae

For each gene, we aligned the gene and all its best hits found across the different species in with Muscle V3.8.155 (Edgar 2004). Then, we built a phylogenetic tree following the HKY85 model for each gene with phyml V3.3 (Guindon et al., 2010). In parallel, we concatenated the alignment of the 1,055 endosperm genes found in the 22 others Brassicaceae to build a phylogenetic tree of the endosperm, without branch length, using phyml V3.3 (Guindon et al., 2010). This tree was used as a proxy of the topology of the endosperm. These steps were integrated into a pipeline of analysis available in GitHub

(https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee). 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 between the RF distances across the group of genes, we conducted bootstrap analyses to compare the value on PEGs and MEGs with values obtained from 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we conducted bootstrap analyses to compare the values on control genes with values obtained from 10,000 resamples of 100 endosperm genes. The R code for the bootstrap test is available in GitHub (https://github.com/leveveaudrey/bootstrap).

Variation of the length of branches across Brassicaceae

We investigated whether the imprinted genes presented variation of the branch length 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 1,055 endosperm genes using phyml V3.3 (Guindon et al., 2010). These steps were integrated into a pipeline of analysis available in GitHub (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee). We extracted the branches length for 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 in the branch length across the group of genes, we conducted bootstrap analyses to compare the mean in PEGs and MEGs with means obtained from 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we conducted bootstrap analyses to compare the means on control genes with values obtained from 10,000 resamples of 100 endosperm genes. The R code for the bootstrap test is available in GitHub (https://github.com/leveveaudrey/bootstrap).

Read mapping and variant calling

We used natural accessions of *A. lyrata* from North America from nine allogamous and nine autogamous populations of 25 individuals each (Table S6), using available pool-sequencing data (Willi et al., 2018).

The raw reads from Willi et al (2018) datasets were mapped onto the complete *A. lyrata* reference genome V1.0.23 (Hu et al., 2011; https://doi.org/10.5281/zenodo.17287897) 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,

690 DePristo et al., 2011) with the option GVCF and a quality score threshold of 60 using
691 vcftools v0.1.15 (Danecek et al., 2011). We excluded sites with less than 15 reads aligned.
692 Sites covered by at least 15 reads but containing no SNP were considered monomorphic.
693 The final number of sites in each sample set is summarised in Table S6. The command lines
694 used for all these steps are available in Github (https://github.com/leveveaudrey/alignment-
695 and-snp-calling).

696 *Effect of imprinting on global population polymorphism*

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

702 For each gene of each population, we estimated the π, the π_{NS}/π_S for all positions and the
703 Tajima's D of each gene using a custom python code
704 (https://github.com/leveveaudrey/POOLSEQ-analysis-of-polymorphism). If a position in the
705 *A. lyrata* genome was covered but not polymorphic, the π was set to 0. Then, the 0-fold and
706 4-fold degenerated sites were identified and extracted from the reference genome and the
707 genome annotation using the script NewAnnotateRef.py (Williamson et al., 2014; updated
708 version available in https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee).
709 Because all mutations at 0-fold degenerated sites alter the sequence of the encoded protein,
710 we assumed that these mutations are non-synonymous (NS). In contrast, mutations at the 4-
711 fold degenerate sites never alter the encoded amino acid, thus we assumed that these
712 mutations are synonymous (S). For the sake of simplicity, we discarded mutations at 2- or 3-
713 fold degenerate sites. Based on the π estimated for each site, we estimated the mean π_{NS}
714 and π_S, for each gene in each population, in order to estimate the π_{NS}/π_S. We removed the
715 genes with a π_S to 0. For the Tajima's D, we removed the genes having less than 0.5%
716 polymorphic sites.

717 To test for significant differences across the group of genes, we conducted bootstrap
718 analyses to compare the mean π, π_{NS}/π_S and Tajima's D of PEGs and MEGs with means
719 obtained from 10,000 resamples of 29 endosperm genes and control genes independently.
720 Similarly, we conducted bootstrap analyses to compare the values on control genes with
721 values obtained from 10,000 resamples of 100 endosperm genes. The R code for the
722 bootstrap test is available in GitHub (https://github.com/leveveaudrey/bootstrap). We
723 performed the test on allogamous and autogamous populations independently. Finally, for

each statistic and group independently, we tested for the mating system effect by a performing Kruskal-Wallis test as implemented in R version 4.3.2.

Genes in coevolution

To detect coevolving coding sequences, we used three different methods based on the following assumptions: coevolution lead to 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 correlations matrices, we extracted the correlations between pairs of endosperm genes to define the thresholds of extreme values in the distributions. Then, all pairs of genes that included imprinted genes with a correlation 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 the networks of coevolution.

To detect coevolution based on topology with other genes, we used first the distribution of RF distances between endosperm genes and the global endosperm topology to estimate the value obtained in 95% of the distribution. Then, we excluded all the imprinted and endosperm genes having lower RF distance than this value to remove genes whose topology was explained by the global topology. Finally, we compared the topology of each remaining imprinted genes 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 1,000 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 across 1,000 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 to detect coevolving sites for each pair of genes showing 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 sites within each pair of genes were predicted using the clustering method implemented in the CoMap package (version 1.5.2; https://jydu.github.io/comap/; Dutheil and Galtier 2007). The built-in algorithm was then used to correct for multiple testing and to 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 comparison pair, we extracted the coevolving groups of substitutions and classified them by size (between 2 and 10).

To test for significant excess of correlations, we compared the results obtained for the pair of imprinted genes found in networks previously mentioned with the results obtained for equivalent number of pair comparisons of imprinted genes not 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 C.L-P). 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. A.L thanks personally Stella Huynh for our beneficial suggestions and commentaries.

Author contributions

O.I developed and performed the transcriptional analysis. J-Y.D and A.L developed and performed the coevolution analysis. A.L analysed and interpreted the data. A.L wrote the manuscript. All authors edited the manuscript.

Declaration of interests

The authors declare no competing interests.

Bibliography

1) Abdel-Haleem, H., 2007. The Origins of Genome Architecture. J Hered 98, 633–634. https://doi.org/10.1093/jhered/esm073

2) 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

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) 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) Bernardi, G., 2000. Isochores and the evolutionary genomics of vertebrates. Gene 241, 3–17. https://doi.org/10.1016/s0378-1119(99)00485-0

6) Botet, R., and Joost J. B. Keurentjes., 2020. The Role of Transcriptional Regulation in Hybrid Vigor. Frontiers in Plant Science 11 (April). https://doi.org/10.3389/fpls.2020.00410

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) Clark, N.L., Alani, E., Aquadro C.F., 2012. Evolutionary Rate Covariation Reveals Shared Functionality and Coexpression of Genes. Genome Research 22 (4): 714–20. https://doi.org/10.1101/gr.132647.111

10) Cvijović, I., Good, B.H., Desai, M.M., 2018. The Effect of Strong Purifying Selection on Genetic Diversity. Genetics 209 (4): 1235–78. https://doi.org/10.1534/genetics.118.301058

11) 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

12) 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

13) 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

14) 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

15) 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

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

17) Eaux, A.J., Bilinski, P., Eichten, S.R., Vaughn, M.W., Ross-Ibarra, J., Gehring, M., Reprinte, N.M., 2013. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proceedings of the National Academy of Sciences 110, 19639–19644. https://doi.org/10.1073/pnas.1309182110

18) 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

19) 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

20) Florez-Rueda, A.M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U., Städler, T., 2016. Genomic Imprinting in the Endosperm Is Systematically Perturbed in Abortive Hybrid Tomato Seeds. Mol Biol Evol 33, 2935–2946. https://doi.org/10.1093/molbev/msw175

21) Flores-Vergara, M.A., Oneal, E., Costa, M., Villarino, G., Roberts, C., De Luis Balaguer, M.A., Coimbra, S., Willis, J., Franks, R.G., 2020. Developmental Analysis of Mimulus Seed Transcriptomes Reveals Functional Gene Expression Clusters and Four Imprinted, Endosperm-Expressed Genes. Front. Plant Sci. 11. https://doi.org/10.3389/fpls.2020.00132

22) 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

23) Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W., 2002. Evolutionary Rate in the Protein Interaction Network. Science 296, 750–752. https://doi.org/10.1126/science.1068696

24) Galtier, N., Rousselle, M., 2020. How Much Does Ne Vary Among Species? Genetics 216, 559–572. https://doi.org/10.1534/genetics.120.303622

25) Garro, M., Greco, E., Vannay, G.J., Leonova, A., Bruno, L., Capella, M., 2025. Non-CG DNA methylation represses SDC expression to modulate hypocotyl elongation during thermormorphogenesis in Arabidopsis. J Exp Bot 76, 2517–2534. https://doi.org/10.1093/jxb/eraf105

26) Gaudinier, A., Rodriguez-Medina, J., Zhang, L., Olson, A., Liseron-Monfils, C., Bågman, A.-M., Foret, J., Abbitt, S., Tang, M., Li, B., Runcie, D.E., Kliebenstein, D.J., Shen, B., Frank, M.J., Ware, D., Brady, S.M., 2018. Transcriptional regulation of nitrogen-associated metabolism and growth. Nature 563, 259–264. https://doi.org/10.1038/s41586-018-0656-3

27) 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

28) 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

29) 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

30) Grishkevich, V., Yanai, I., 2014. Gene length and expression level shape genomic novelties. Genome Res 24, 1497–1503. https://doi.org/10.1101/gr.169722.113

31) 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

32) 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

33) 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. https://doi.org/10.1098/rstb.1991.0057

34) 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

35) 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

36) 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

37) 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

38) 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

39) 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

40) Hojsgaard, D., Schartl, M., 2021. Skipping sex: A nonrecombinant genomic assemblage of complementary reproductive modules. BioEssays 43, 2000111. https://doi.org/10.1002/bies.202000111

41) 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

42) 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

43) 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

44) İltaş, Ö., Svitok, M., Cornille, A., Schmickl, R., Lafon Placette, C., 2021. Early evolution of reproductive isolation: A case of weak inbreeder/strong outbreeder leads to an intraspecific hybridization barrier in Arabidopsis lyrata. Evol 75, 1466–1476. https://doi.org/10.1111/evo.14240

45) 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

46) Jiang, H., Guo, D., Ye, J., Gao, Y., Liu, H., Wang, Y., Xue, M., Yan, Q., Chen, J., Duan, L., Li, G., Li, X., Xie, L., 2021. Genome-wide analysis of genomic imprinting in the endosperm and allelic variation in flax. The Plant Journal 107, 1697–1710. https://doi.org/10.1111/tpj.15411

47) 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

48) Kinoshita, T., 2007. Reproductive barrier and genomic imprinting in the endosperm of flowering plants. Genes & Genetic Systems 82, 177–186. https://doi.org/10.1266/ggs.82.177

49) Kinser, T.J., Smith, R.D., Lawrence, A.H., Cooley, A.M., Vallejo-Marín, M., Conradi Smith, G.D., Puzey, J.R., 2021. Endosperm-based incompatibilities in hybrid monkeyflowers. Plant Cell 33, 2235–2257. https://doi.org/10.1093/plcell/koab117

50) Köhler, C., Scheid, O.M., Erilova, A., 2010. The impact of the triploid block on the origin and evolution of polyploid plants. Trends in Genetics 26, 142–148. https://doi.org/10.1016/j.tig.2009.12.006

51) Koonin, E.V., 2015. Why the Central Dogma: on the nature of the great biological exclusion principle. Biology Direct 10, 52. https://doi.org/10.1186/s13062-015-0084-3

52) Korneliussen, T.S., Moltke, I., Albrechtsen, A., Nielsen, R., 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics 14, 289. https://doi.org/10.1186/1471-2105-14-289

53) 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

54) Kremer, A., Chen, J., Lascoux, M., 2025. 'Chimes of resilience': what makes forest trees genetically resilient? New Phytologist 246, 1934–1951. https://doi.org/10.1111/nph.70108

55) Laenen, B., Tedder, A., Nowak, M.D., Toräng, P., Wunder, J., Wötzel, S., Steige, K.A., Kourmpetis, Y., Odong, T., Drouzas, A.D., Bink, M.C.A.M., Ågren, J., Coupland, G., Slotte, T., 2018. Demography and mating system shape the genome-wide impact of purifying selection in Arabis alpina. Proc Natl Acad Sci U S A 115, 816–821. https://doi.org/10.1073/pnas.1707492115

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

57) 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

58) 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

59) 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

60) Lopes, I., Altab, G., Raina, P., de Magalhães, J.P., 2021. Gene Size Matters: An Analysis of Gene Length in the Human Genome. Front Genet 12, 559998. https://doi.org/10.3389/fgene.2021.559998

61) Lynch, M., 2010. Evolution of the mutation rate. Trends Genet 26, 345–352. https://doi.org/10.1016/j.tig.2010.05.003

62) Maheshwari, S., Barbash, D.A., 2011. The Genetics of Hybrid Incompatibilities. Annu. Rev. Genet. 45, 331–355. https://doi.org/10.1146/annurev-genet-110410-132514

63) Massey, S.E., 2013. Proteome size as the major factor determining mutation rates. Proceedings of the National Academy of Sciences 110, E858–E859. https://doi.org/10.1073/pnas.1219306110

64) McVean, G.T., Hurst, L.D., 1997. Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. Proceedings of the Royal Society of London. Series B: Biological Sciences 264, 739–746. https://doi.org/10.1098/rspb.1997.0105

65) 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

66) 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

67) 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

68) 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

69) Onosato, H., Fujimoto, G., Higami, T., Sakamoto, T., Yamada, A., Suzuki, T., Ozawa, R., Matsunaga, S., Seki, M., Ueda, M., Sako, K., Galis, I., Arimura, G., 2022. Sustained defense response via volatile signaling and its epigenetic transcriptional regulation. Plant Physiol 189, 922–933. https://doi.org/10.1093/plphys/kiac077

70) 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

71) Padilla-García, N., Le Veve, A., Čermák, V., İltaş, Ö., Contreras-Garrido, A., Legrand, S., Aury, J.-M., Horvath, R., Lafon Placette, C., 2025. The Demographic History of Populations and Genomic Imprinting have Shaped the Transposon Patterns in Arabidopsis lyrata. Mol Biol Evol 42, msaf093. https://doi.org/10.1093/molbev/msaf093

72) Papdi, C., Ábrahám, E., Joseph, M.P., Popescu, C., Koncz, C., Szabados, L., 2008. Functional Identification of Arabidopsis Stress Regulatory Genes Using the Controlled cDNA Overexpression System. Plant Physiol 147, 528–542. https://doi.org/10.1104/pp.108.116897

73) 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

74) 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

75) 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

76) Presgraves, D.C., 2010. The molecular evolutionary basis of species formation. Nat Rev Genet 11, 175–180. https://doi.org/10.1038/nrg2718

77) 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

78) 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

79) 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

80) Roth, M., Florez-Rueda, A.M., Paris, M., Städler, T., 2018. Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. The Plant Journal 95, 1084–1101. https://doi.org/10.1111/tpj.14012

81) Sahoo, S., Das, S.S., Rakshit, R., 2019. Codon usage pattern and predicted gene expression in Arabidopsis thaliana. Gene, Articles initially published in Gene: X 1-4, 2019 721, 100012. https://doi.org/10.1016/j.gene.2019.100012

82) Sandstedt, G.D., Sweigart, A.L., 2022. Developmental evidence for parental conflict in driving Mimulus species barriers. New Phytol 236, 1545–1557. https://doi.org/10.1111/nph.18438

83) 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

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

85) Sharp, P.M., Li, W.H., 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15, 1281–1295. https://doi.org/10.1093/nar/15.3.1281

86) 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

87) Spencer, H. G. 2000. Population Genetics and Evolution of Genomic Imprinting. Annual Review of Genetics 34: 457–77. https://doi.org/10.1146/annurev.genet.34.1.457.

88) 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

89) 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

90) 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

91) 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

92) 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

93) 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

94) 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

95) 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

96) Wilkins, J.F., Haig, D., 2003. What Good Is Genomic Imprinting: The Function of Parent-Specific Gene Expression. Nature Reviews Genetics 4 (5): 359–68. https://doi.org/10.1038/nrg1062.

97) 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

98) 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

99) 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

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

101) 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

102) Yang, G., Feng, M., Yu, K., Cui, G., Zhou, Y., Sun, L., Gao, L., Zhang, Y., Peng, H., Yao, Y., Hu, Z., Rossi, V., De Smet, I., Ni, Z., Sun, Q., Xin, M., 2023. Paternally imprinted LATE-FLOWERING2 transcription factor contributes to paternal-excess interploidy hybridization barriers in wheat. Journal of Integrative Plant Biology 65, 2587–2603. https://doi.org/10.1111/jipb.13574

103) 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

104) 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

105) 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

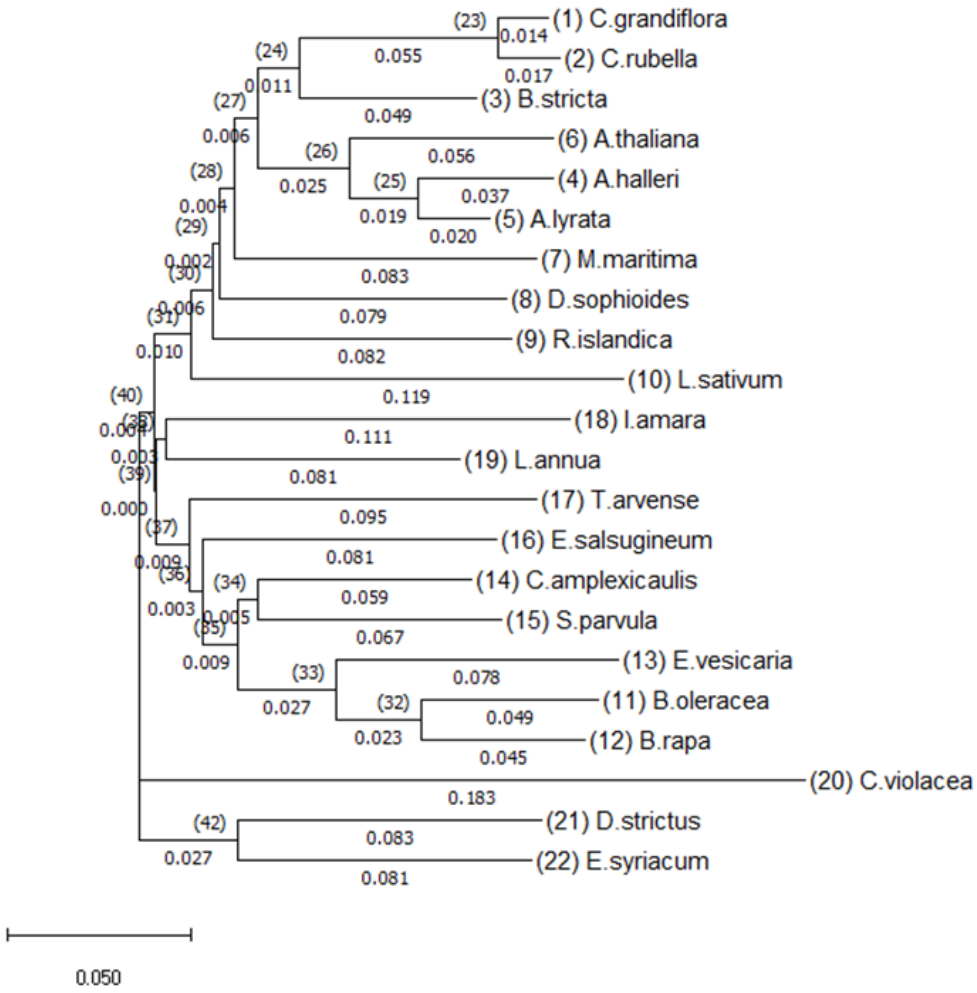

Figure S1: phylogenetic tree obtained by HKY85 model based on the 1055 endosperm genes found in all species after 1000 bootstrap. *The percentage of trees in which the associated haplotypes clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in mean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