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 for 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. Consequently, imprinted genes are expected to be disproportionately expressed in the endosperm in plants, the tissue where parental conflict operates (Haig 1997; Haig 2000). Thus, the proportion of maternally expressed genes (MEGs) and paternally expressed genes (PEGs) upregulated in the endosperm should exceed the genome-wide expectation derived from sampling all expressed genes. According to this hypothesis, many imprinted genes appear to be involved in resource accumulation in the endosperm and metabolic processes (Raissig et al., 2011). These theoretical expectations and previous empirical observations justify testing whether MEGs and PEGs are disproportionately upregulated in endosperm relative to random expectation.

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); Exactly, PEGs are predicted to evolve higher optimal expression levels than MEGs, because paternal alleles benefit from increased resource acquisition, whereas maternal alleles tend to limit growth (Haig 2000; Patten et al., 2014). Accordingly, we expect an higher expression per active gene

copy in PEGs than in MEGs and finally in non-imprinted endosperm genes; (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 test these predictions by combining expression, phylogenetic and population-genomic analyses to evaluate whether imprinted genes in *A. lyrata* show evolutionary

signatures consistent with the kinship theory. 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) correlations between evolutionary rates or diversity metrics used as exploratory indicators of evolutionary covariance expected in case of 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

General analytical framework

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). 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).

Throughout the Results section, we compare four categories of genes: maternally expressed genes (MEGs) and paternally expressed genes (PEGs) that are upregulated in the endosperm (referred “imprinted genes” if considered together), the 2,703 non-imprinted genes specifically upregulated in the endosperm (“endosperm genes”), and a set of 100 control genes. The control panel was randomly sampled from the set of 126 ubiquitously expressed genes and serves as a baseline for genome-wide patterns of selection unrelated to parental conflict. The control panel was chosen following the expression-based sampling strategy used in previous imprinting studies, rather than by matching genomic characteristics across groups. Differences in gene size, CDS proportion, and GC content among panels were therefore quantified and considered when interpreting downstream evolutionary analyses.

For each statistic, genes with insufficient informative sites were excluded. Then, significance was evaluated using 10,000 bootstrap resampling matching the size of focal group. framework because this method is unsensitive to difference of samples sizes. Rather than repeating the method in each subsection, we applied a single unified framework in which observed means are compared to the 5th–95th percentiles of bootstrap distributions. This approach was used consistently across analyses of expression, codon usage, phylogenetic metrics, and population polymorphism. Although the figures share a similar visual layout for consistency, each statistic captures a distinct evolutionary dimension allowing direct comparison of patterns across gene categories while retaining analytical complementarity.

We first compared gene size, CDS content (%) and GC content (%) across groups, as these features may influence polymorphism patterns and generate bioinformatic artefacts. Control genes and MEGs were significantly smaller than PEGs and endosperm genes, and CDS content was significantly higher in endosperm and imprinted genes relative to control genes. GC content was also lower in imprinted genes than in control genes (Table S2). These

differences were quantified and considered when interpreting all evolutionary analyses. Nevertheless, differences in gene size, CDS proportion and GC content were not corrected for directly, and we retained the control panel to be congruent with methods used in previous studies in the autogamous species *C. rubella* and *A. thaliana* (Hatorangan et al., 2016; Tuteja et al., 2019). In practice, all evolutionary comparisons evaluating predictions of the kinship theory were performed relative to endosperm-upregulated non-imprinted genes, which represent the expression-matched and biologically relevant background for imprinting analyses. The control panel was used only to describe genome-wide patterns of variation.

To evaluate whether these genomic differences could create artificial signals of selection, we performed several robustness checks. First, for population-genomic analyses, we verified that the SNP-density filter did not preferentially exclude short, GC-poor or low-diversity genes. Second, by comparing all statistics against empirical null distributions derived from endosperm-upregulated genes, we ensured that each test was evaluated against a biologically appropriate baseline sharing similar expression patterns, thereby reducing the influence of residual genomic confounders unrelated to imprinting or parental conflict. Finally, we compared models using the Akaike Information Criterion (AIC). The full model included gene group, gene length, CDS proportion and GC content as explanatory variables. We then evaluated a series of reduced models in which one parameter was removed at a time and retained the model with the lowest AIC value. This procedure was iterated until the best-supported model was identified. This approach allowed to determine (1) whether gene group remained a significant predictor in the best model and (2) whether genomic features accounted for the observed trends in the analysed statistics.

In previous study (Tuteja et al., 2019), validation of kinship theory to explain imprinting was performed following another enrichment-based logic, which evaluated not the global behaviour of imprinted genes but the proportion of them exhibiting extreme signatures compatible with parental conflict. Thus, we tested whether directional selection occurred more frequently in imprinted genes than in the empirical distribution of endosperm genes. Candidates for directional selection were defined as the 5% most extreme values for branch length, Tajima's D, or π_{NS}/π_S in allogamous populations. Under the null hypothesis of no enrichment, imprinted genes should fall into this class with probability 0.05. We therefore tested whether the observed proportion exceeded 5% using independent binomial tests.

Coevolution was assessed using three complementary measures (Pearson's correlation coefficients for π, branch lengths, and RF distances, for each pair of PEGs, MEGs, and endosperm genes), and only correlations exceeding thresholds defined from pair of endosperm genes (1st or 99th percentiles) were retained. Finally, we tested if the number of

correlations found between PEGs and MEGs and the number of imprinted genes involved were significantly higher than expected from 1,000 random samplings of 29 endosperm genes.

Imprinted genes are preferentially expressed in the endosperm

Under kinship theory, both MEGs and PEGs are expected to be disproportionately represented among endosperm-upregulated genes, because imprinting evolves specifically in tissues exposed to parental conflict. We therefore first tested whether the proportion of upregulated MEGs and PEGs exceeded expectations derived from random sampling of the transcriptome. The proportions of MEGs (0.56) and PEGs (0.30) upregulated in the endosperm were both significantly higher than expected from resampling of all genes (p values < 1e-04), and MEGs were more frequently upregulated than PEGs (based on resampling of imprinted genes; p value =0.03).

Because PEGs are predicted to increase offspring resource acquisition, while MEGs tend to restrict it, theory predicts higher optimal expression levels for PEGs, followed by MEGs, compared with non-imprinted endosperm genes. We therefore tested whether expression per expressed allele differed among groups after correcting for gene dosage. After adjusting expression for gene copy number (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 < 1e-03; Fig 1).

Using AIC, we found that the best model to explain this expression pattern (AIC = -4633) was this that include the group of genes, but also the gene size (coefficient = -5.4e-05), the CDS proportion (coefficient = -2.34e-03) and the GC proportion (coefficient = 2.66e-02), whereas the worst model (AIC = -2454) was this that exclude only the group of genes (AIC = -2454).

We next assessed whether this increased expression was associated with codon or amino acid usage bias. To avoid repeating the same test structure, results for codon usage are presented concisely: codon and amino acid usage differed significantly between imprinted and endosperm genes, although only two biases were shared by PEGs and MEGs (under-usage of GTC and over-usage of tryptophan; Table S3). However, PHE scores (an index of codon impact on expression, derived from overexpressed genes in *A. thaliana*; Sahoo et al., 2019) did not differ significantly across biased and unbiased codons (all p values > 0.09), indicating that codon usage differences were not associated with expression variation.

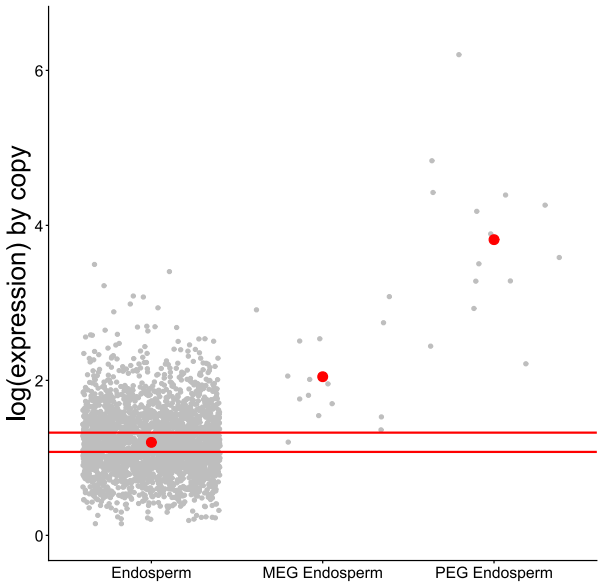


Figure 1: Imprinted genes are associated to higher expression level by gene copies than in endosperm genes. *The grey dots represent the mean expression level obtained for one gene. The red dots represent the mean expression level for each group of gene. The red lines represent 5% and 95% of the means 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

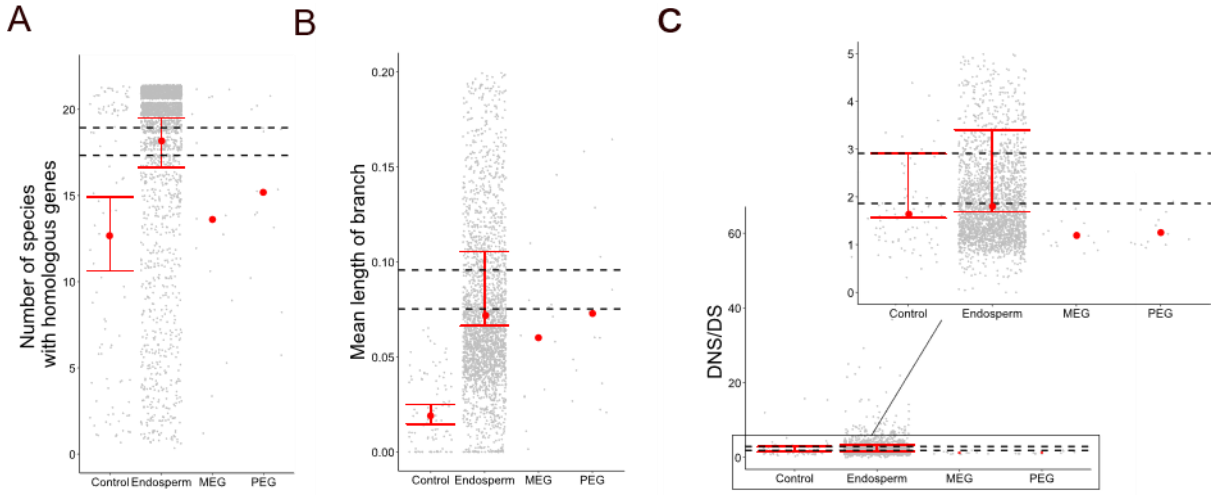
To explore phylogenetic patterns, we analysed genes homologs in *A. lyrata* and 21 Brassicaceae species (Table S4). As a proxy for gene turnover, we compared the number of species with homologous sequences across groups. Thus, endosperm genes were conserved across significantly more species than other categories, and PEGs more than control genes (Fig 2; Table S5). The best model to explain this pattern (AIC = 8696) was this that include the group of genes, the gene size (coefficient = 9.74-04) and the GC proportion (coefficient = 0.19), whereas the worst model (AIC = 9104) was this that exclude the genes size.

Gene-tree topologies of MEGs and PEGs did not differ significantly in imprinted genes from the species tree (based on Robinson–Foulds RF distances; bootstrap tests; p value = 0.60 for PEGs; p value = 0.34 for MEGs), so branch lengths (proxy for evolutionary rates; Fig. S1) were estimated directly from the species-tree topology to avoid repeating similar topological tests for genes having homologs in at least three species (including *A. lyrata*). Thus, endosperm and imprinted genes show significantly longer branches than in control genes (Fig 2; bootstrap; Table S5), even if MEGs exhibited significantly shorter branch lengths than endosperm genes. The best model to explain this pattern (AIC = -15231) was this that include the group of genes, the gene size (coefficient = 9.73e-06), the CDS proportion (coefficient = -8.10e-04) and the GC proportion (coefficient = 9.63e-04), whereas the worst model (AIC = -14976) was this that exclude the genes size.

Finally, mean D_{NS}/D_S between *A. lyrata* and the other Brassicaceae species decreased significantly in imprinted genes (Table S5), indicating a signal of purifying (negative) selection.

252 The best model to explain this pattern (AIC = 13382) was this that include the group of genes
253 and the GC proportion (coefficient = 0.18), whereas the worst model (AIC = 13401) was this
254 that exclude the group of genes.

255 Together, these phylogenetic patterns indicate long-term evolutionary constraints in
256 endosperm genes, a signal that can arise under ancient balancing selection or strong
257 functional conservation. Importantly, these inferences reflect deep evolutionary timescales
258 and therefore cannot be directly compared to population-level signatures of balancing
259 selection, which we analyse in the next section.



260

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

276 *Imprinted genes show elevated* π_{NS}/π_S *independent of mating system*

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 across 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 calculations these statistic required sufficient polymorphism in all populations, genes with fewer than 0.5% polymorphic sites were excluded from Tajima's D, and genes with $\pi_S = 0$ from π_{NS}/π_S analyses, resulting in a significantly higher retention rates for endosperm genes than for control genes for Tajima's D and π_{NS}/π_S (binomial tests; Table S7). However, more importantly, these filters retained similar proportions of genes across endosperm and imprinted genes, and across imprinted and control genes (Table S7). Thus, the observed contrasts in Tajima's D or π_{NS}/π_S for the imprinted genes cannot be attributed to unequal gene retention nor to systematic biases against low-diversity loci, supporting that the signals reflect genuine biological differences rather than filtering artefacts.

In allogamous populations, the Tajima's D was significantly higher in endosperm genes than in control genes and PEGs, while the π and the π_{NS}/π_S showed no significant differences. Imprinted genes showed an overall increase in π_{NS}/π_S, while MEGs displayed a significantly lower π relative to control genes (Fig 3; Table S8). The best model to explain this pattern of Tajima's D (AIC = 9322) was this that include only the gene size (coefficient = -1.59e-04), whereas the worst model (AIC = 9337) was this that exclude only the gene size. For the π, the best model (AIC = -36369) was this that include the gene size (coefficient = -9.89e-08) and the GC percentage (coefficient = -5.59e-05) whereas the worst model (AIC = 9337) was this that exclude only the gene size. Finally, the best model to explain this pattern of T π_{NS}/π_S (AIC = -721) was this that include only the proportion of cds (coefficient = 2.21e-03), whereas the worst model (AIC = 9337) was this that include all the parameters.

In autogamous populations, Tajima's D in endosperm genes no longer differed from controls (Table S8). At the opposite, MEGs showed a significantly higher Tajima's D than endosperm genes. Moreover, MEGs and endosperm genes exhibited decreased π and π_{NS}/π_S compared to control genes (Table S8). These trends are explained by mating-system that only significantly affects endosperm genes (Kruskal-Wallis tests; Table S9).

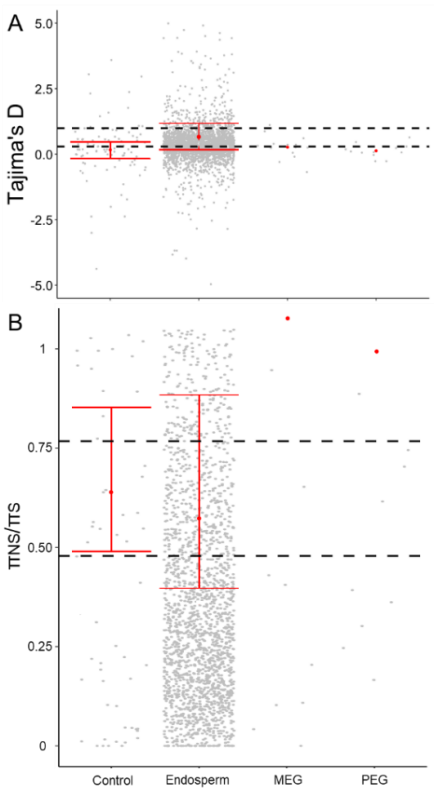


Figure 3: Imprinted genes present an unusual decrease of Tajima's D associated to an increase of π_{NS}/π_S in allogamous populations compared to endosperm genes. *A: Variations of the Tajima's D by group of genes. Only the genes with a mean between -5 and 5 are indicated. B: Variations of the π_{NS}/π_S 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 1,000 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.

To assess whether imprinted genes show an enrichment in signatures of directional selection, we tested whether they occur more frequently than expected in the extreme tail of the empirical distribution of endosperm genes. Candidates for directional selection were defined as the 5% most extreme values for branch length, Tajima's D, or π_{NS}/π_S in allogamous populations. However, none of these proportions were significantly higher than expected (p = 0.66, 1 and 1 respectively).

We identified only four MEGs potentially under strong directional selection (Table S1). Two MEGs were associated with a reduced 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), involved in female gametophyte, seed and early endosperm development (Bemer et al., 2010; Zhang et al., 2018). One MEG identifing by a reduced Tajima's D encodes a F-box domain–containing protein regulated by non-CG DNA methylation in *A. thaliana* (Garro et al., 2025) and associated with the regulation

343 of stress-responsive genes (Papdi et al., 2008). Finally, one MEG encoding a Dof-type zinc
344 finger protein involved in seed coat formation (Zou et al., 2012) shows a reduced π_{NS}/π_S.

345 *Clear trend of coevolution among imprinted genes in Brassicaceae*

346 Imprinting could promote coevolution between genes favouring paternal or maternal fitness in
347 an arms race scenario. To test this hypothesis, we investigated potential coevolutionary
348 signals between PEGs, MEGs and endosperm genes, based on (i) nucleotide diversity (π)
349 within populations, (ii) branch length in phylogenetic trees, and (iii) Robinson–Foulds (RF)
350 distances between tree topologies. Because the coevolution hypothesis is more closely
351 related to noncoding regions, we focused on π rather than π_{NS}/π_S.

352 First, using thresholds defined from endosperm genes (−0.66 and 3.20 for 1 and 99% of
353 distribution), 13 significant Pearson's correlation coefficient correlations involving 6 PEGs and
354 9 MEGs were detected. Both the number of PEGs-MEGs correlations and the number of
355 imprinted genes involved exceeded expectations (p-values= 4.5e-02 and <1e-03 respectively;
356 Table S10). Considering endosperm genes significantly correlated with these imprinted genes,
357 as well as indirect links between imprinted and endosperm genes, the resulting network
358 contained 14 MEGs, 13 PEGs, and 1,643 endosperm genes (=1,670 genes; Fig 4.A; Table
359 S10). This network did not show enrichment for specific GO terms compared to random
360 networks of equal size. Notably, one PEG encoding a chloroplast-localized protein required
361 for embryonic development (Bryant et al., 2010) was correlated with more genes than 99% of
362 endosperm genes (Table S1 and S10).

363 Following the same methodology, the thresholds based on branches length(−0.64 and 0.89)
364 identified 16 significant correlations among 6 PEGs and 9 MEGs. Both the number of PEGs-
365 MEGs correlations and the number of imprinted genes involved exceeded expectations (p-
366 values = 8e-03 and 3e-03 respectively; Table S11). None of these correlations was in common
367 with the previous 13 significant correlations of polymorphism. Considering endosperm genes
368 significantly correlated with these imprinted genes, the resulting network contained 14 MEGs,
369 14 PEGs, and 1,027 endosperm genes (=1,055 genes; Fig 4.B; Table S11). This network was
370 enriched for the GO categories "*negative regulation of catalytic activity"* and "*syntaxin binding"*
371 (Table S12). One MEG in this network, encoding the AGAMOUS-like protein previously
372 mentioned, was correlated with more genes than 99% of endosperm genes (Table S11).

373 Finally, we investigated gene pairs that showed high similarity together (RF = 0) but higher
374 deviation from the species topology than 95% of endosperm genes (RF > 0.71 with species
375 topology; Fig. S1). Based on this approach, the topology of only one PEG encoding an F-box

376 protein was associated with six endosperm genes (Fig. 4.C; Table S13). This small network
377 was enriched for the GO term *DNA-binding transcription factor activity* (Table S12). Moreover,
378 the PEG exhibited a significantly higher number of associations than 99% of non-imprinted
379 endosperm genes (Table S1; Table S13).

380 Across all analyses, 22 imprinted genes pairs (29 genes) were identified. To test for
381 coevolution between specific mutations within these pairs, we applied the clustering approach
382 implemented in the CoMap package (Dutheil and Galtier, 2007).

383 Specifically, we searched for groups of 2 to 10 coevolving sites within each candidate pair of
384 imprinted genes. To limit false positives, the number of such groups detected in candidate
385 pairs was compared to the number observed in non-candidate pairs of imprinted genes that
386 were not associated in the previous networks and were present in at least three species.
387 CoMap did not detect an excess of coevolving residues in candidate pairs relative to non-
388 candidate imprinted pairs (Table S14). These directly correlated pairs of imprinted genes were
389 therefore not associated with elevated levels of site-level coevolution.

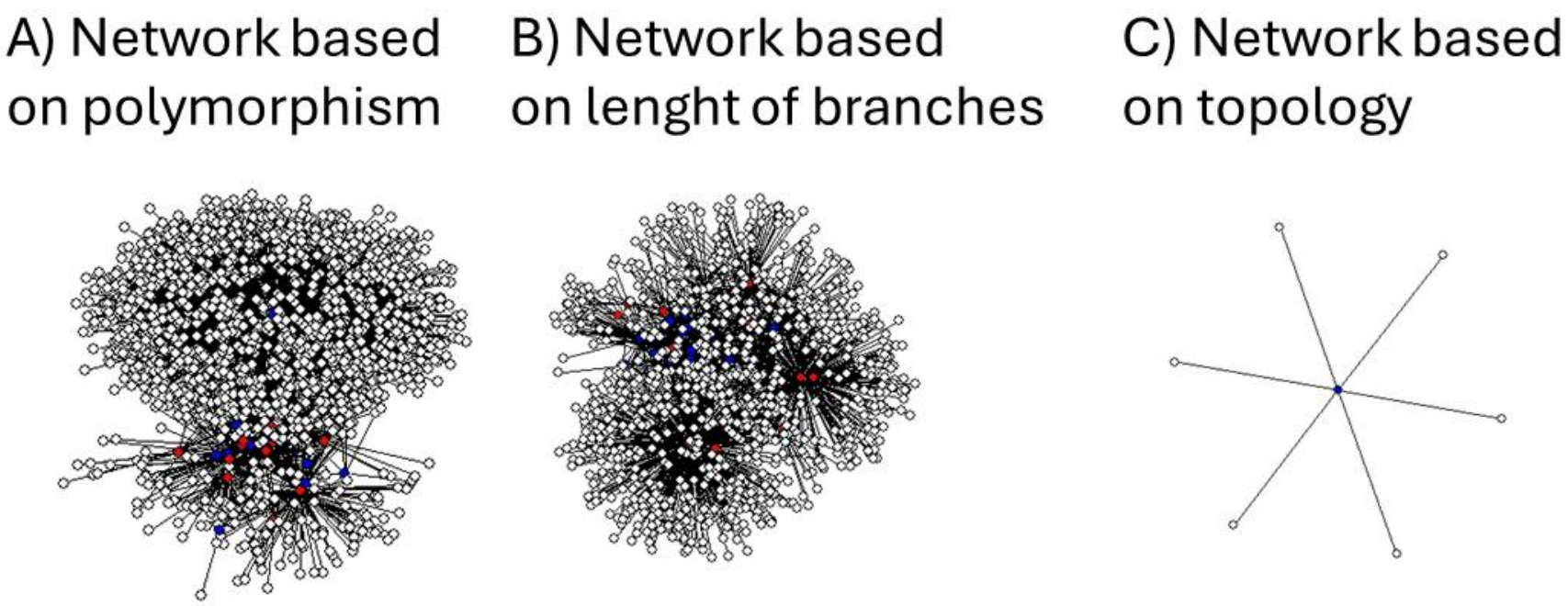


Figure 4: Networks of the imprinted and endosperm genes based on A) polymorphism in population , B) length of branches and C) phylogenetic topologies across the Brassicaceae. *A) Network based on the Pearson's correlations of π found on genes in the populations. B) Network based on the Pearson's correlations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

396 # Discussion

397 The kinship theory remains the prominent selective framework for explaining the evolution of
398 genomic imprinting (Patten et al., 2014). According to this theory, imprinted genes are should
399 (I) be involved in resource allocation within systems with multiple mating (Haig 2000), (II)
400 exhibit signatures of directional selection relative to non-imprinted endosperm genes (Day and
401 Bonduriansky 2004; Kasimatis et al., 2017; Table 1), and could present some signals of
402 coevolution among genes (Wilkins and Haig 2001; Willi 2013). However, a selection acting
403 primarily on gene expression may attenuated these genomic signatures, and their sensitivity
404 to mating-system variation remains unclear in this case.

405 Here, we tested these predictions by comparing imprinted and non-imprinted endosperm
406 genes in *A. lyrata*, an allogamous species that shift to autogamy in some population from North
407 America, and assessing the impact of mating system shifts. While several figures follow a
408 unified graphical structure to facilitate comparison across gene categories, the underlying
409 analyses encompass expression-level differences, population genetic statistics, phylogenetic
410 metrics, and coevolutionary patterns, providing a broader and complementary evaluation of
411 kinship-theory predictions than previous work (Miyake et al., 2009; Hatorangan et al., 2016;
412 Tuteja et al., 2019). Although we detected patterns consistent with predictions, the signals
413 differed across statistics and were not consistently replicated among analyses. This suggests
414 that selection associated with parental conflict may act heterogeneously across loci or
415 primarily through regulatory rather than coding changes, and underscores the importance of
416 integrating multiple lines of evidence when assessing evolutionary hypotheses of imprinting.

417 *Signals of ancient balancing selection in endosperm genes*

418 In the absence of genomic imprinting, genes involved in resource allocation conflict are
419 expected to evolve under balancing selection (Day and Bonduriansky 2004; Sylvestre et al.,
420 2023). Consistently, non-imprinted endosperm genes were more widely conserved across
421 Brassicaceae, suggesting lower turnover, and displayed longer phylogenetic branches
422 (Edwards, 2009; Kasimatis et al., 2017; Llaurens et al., 2017). These assumptions are
423 comforted by the presence of the group of genes in best models predicted by AIC for these
424 statistics. In addition, these genes showed elevated Tajima's D values in allogamous
425 populations. However, the absence of a parallel increase in polymorphism, and the absence
426 of this parameter in best model predicted by AIC, suggests that this signal may not solely
427 reflect balancing selection.

428

429 ***Table 1: Summary of the expectations on endosperm and imprinted genes under kinship theory and results obtained.*** *The expectations*
430 *validated by the results are signified in bold.* : expectations on endosperm and imprinted genes tested by bootstrap relatively to control and*
431 *endosperm genes respectively. ** : validated only in MEGs.*

Hypotheses tested	References	Statistic analysed	Expectation on endosperm genes	Expectation on imprinted genes
Imprinted genes are related to tissue implicated on resource's allocations	Haig, 1997	Proportion of endosperm genes	NA	**Increase**
Imprinted genes are selected to optimize gene expression	Patten et al., 2014	Expression by copy		**Upregulated with PEGs > MEGs***
Selection on expression promotes codon and amino acid bias in imprinted genes		Usage of each codon and amino acid		**Increase for some*, decrease for others***
Reducing of intralocus conflict in imprinted genes promotes lower conservation of genes across species and higher turnover	Haig 2000; Spencer 2000	Number of species with homologous genes	**Increase***	**Decrease***
Reducing of intralocus conflict in imprinted genes promotes directional selection		Evolutionary rate by length of branches	**Increase***	Increase if positive*, **decrease if negative*** **
	Spencer 2000	Sens of selection by D_{NS}/D_S	**No variation***	Increase if positive*, **decrease if negative***
Reducing of intralocus conflict in imprinting genes promotes reduced polymorphism in allogamous population by directional selection, more intensively in non-synonymous sites	Spencer 2000; Day and Bonduriansky 2004	Tajima's D	**Increase***	Decrease*
		π	Increase*	Decrease*
	Haig 2000	π_{NS}/π_S	**No variation***	Decrease*
Reducing of intralocus conflict in autogamous populations promotes directional selection in non-imprinting genes		Tajima's D, π_{NS}/π_S	**Decrease**	**No variation if already fixed**, increase if relaxed selection
Possible antagonist coevolution between genes (arms race)	Moore and Haig, 1991	π, length of branches, topology by RF distance	NA	**Excess of PEG–MEG correlations***

432

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

439 It is important to emphasize that mainly the phylogenetic analyses, but only Tajima's D for the
440 population-genetic ones, suggest signals compatible with balancing selection in endosperm
441 or imprinted genes. Phylogenetic patterns reflect long-term evolutionary processes, potentially
442 including ancient or persistent balancing selection, whereas population-level statistics detect
443 more recent forms of balancing selection and are sensitive to demography, recombination,
444 and reduced polymorphism. The absence of consistent population-level evidence therefore
445 does not contradict the phylogenetic signal but rather indicates that any balancing selection
446 acting on these loci is either ancient, weak, or too localized to be detected with gene-wide
447 summary statistics.

448 *Codon usage bias and modulation of expression in imprinted genes*

449 Kinship theory predicts that imprinted genes should be preferentially expressed in the
450 endosperm (Haig 2000; Patten et al., 2014). Consistent with this prediction, imprinted genes
451 were enriched among endosperm-expressed genes. Interestingly, the MEGs are more related
452 to endosperm than PEGs, as observed in maize (Eaux et al., 2013). Yet, more than half were
453 also upregulated in other tissues suggesting that distinct selective pressures unrelated to
454 parental conflict may act on them.

455 An important debate suggests that detecting directional selection in imprinted genes is difficult
456 because selection may primarily target gene expression (Patten et al., 2014). Consistent with
457 this idea, both PEGs and MEGs showed increased expression per gene copy, and the group
458 of genes is the more essential parameter in models to predict expression pattern by AIC.
459 Interestingly, the PEGs are associated to an increase of expression level, despite that gene
460 size reduced globally the efficiency of expression in some study(Grishkevich and Yanai, 2014),
461 including in our study.

462 Moreover, imprinted genes displayed codon usage biases relative to endosperm genes,
463 however, we were not able to determine if these biases enhance or inhibit the expression of
464 imprinted genes (Sahoo et al., 2019). More studies with direct analyses of effect on expression
465 of the different codon usage are required to verify this hypothesis in future.

466 *Phylogenetic signals of ancient purifying selection in imprinted genes.*

467 The kinship theory is expected to promote directional selection on imprinted genes (Wilkins
468 and Haig 2001; Mills and Moore 2004; Day and Bonduriansky 2004; Kasimatis et al., 2017).
469 Consistently, our phylogenetic analyses showed that imprinted genes exhibit lower
470 conservation of homologs across Brassicaceae species, likely reflecting a higher turnover
471 driven by accelerated evolutionary rates. Furthermore, imprinted genes displayed reduced
472 D_{NS}/D_S, indicative of strong negative directional selection, and the group of genes is the more
473 essential parameter in models to predict this pattern by AIC.

474 This signal of purifying selection was also evident in MEGs when branch lengths were
475 compared to those of endosperm genes. We note that two MEGs, including AGAMOUS-like
476 protein implicated in one coevolutionary network, exhibited branch lengths suggesting a
477 particularly strong negative selection. Our observations were not in line with the positive
478 selection detected on PEGs on the autogamous species *A. thaliana* and *C. rubella*
479 (Hatorangan et al., 2016; Tuteja et al., 2019), raising the question of the effect of the mating
480 system on the selection of imprinted genes in these species.

481 Nevertheless, caution is warranted in attributing the observed purifying directional selection
482 solely to imprinting. In fact, the detected purifying selection could be explained by alternative
483 explanations (Patten et al., 2014), including (1) an increase of the efficiency of selection by
484 hemizygous expression (Haig and Westoby 1989), (2) selection against transposable element
485 (Padilla-García et al., 2025) or (3) local adaptation (Kremer et al., 2025). However, none of
486 the current models rigorously predict the effects of these mechanisms on polymorphism and
487 divergence, making it difficult to discriminate among alternative explanations.

488 *Implications for reproductive isolation and adaptation*

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

497 Previous studies empirically demonstrated that some imprinted genes are involved in these
498 barriers, including in the autogamous species *A. thaliana* (Wolff et al., 2015), between the

499 allogamous *Solanum peruvianum* et *S. chilense* (Florez-Rueda et al., 2016), between the
500 allogamous *C. grandiflora* and the autogamous *C. rubella* and *C. orientalis* (Lafon-Placette et
501 al., 2018), between the allogamous *Mimulus guttatus*, *M. luteus* (Kinser et al., 2021), *M.*
502 *caespitosa* and *M. tilingi* (Sandstedt and Sweigart 2022), and between the autogamous wheat
503 species *Triticum durum* and *T. aestivum* (Yang et al., 2023).

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

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

510 Consistent with the expectations under the kinship theory, PEGs were associated with
511 reduced Tajima's D (Haig 2014). However, this signal was not observed for π_{NS}/π_S. More
512 interestingly, π_{NS}/π_S values were higher in both PEGs and MEGs compared to endosperm
513 and control genes, which could indicate a relaxation of selection on coding regions (Galtier
514 and Rousselle 2020). This apparent relaxation may not reflect parental conflict but instead
515 could result from the recent colonization of North America by *A. lyrata* from Europe (Ross-
516 Ibarra et al., 2008; Otto et al., 2015; Willi et al., 2018) that could have reduced effective
517 population size (Galtier and Rousselle 2020) and the efficiency of purifying selection (Laenen
518 et al., 2018), especially on transposable element (Padilla-García et al., 2025). Testing this
519 hypothesis would require future comparisons of selection signals in endosperm and imprinted
520 genes across allogamous populations with varying levels of polymorphism. Nevertheless, we
521 note that two MEGs exhibited reduced Tajima's D or π_{NS}/π_S, consistent with ongoing
522 directional selection.

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

524 Conflict over resource allocation is expected to be reduced in autogamous lineages relative to
525 allogamous ones (Haig 2000; Brandvain and Haig 2005). However, most studies of genetic
526 and phylogenetic signatures of selection associated with imprinting have focused on
527 autogamous species, leaving the effect of shifting mating system on polymorphism of
528 imprinted genes unresolved. In the autogamous species *A. thaliana*, the imprinted *MEDEA*
529 locus shows increased polymorphism but not in its allogamous sister species *A. lyrata*
530 (Spillane et al., 2007; Miyake et al., 2009). However, PEGs were associated to positive
531 selection in the autogamous species *A. thaliana* and *C. rubella* (Hatorangan et al., 2016;

532 Tuteja et al., 2019). In our study, the selection signals in imprinted genes, contrary to
533 endosperm genes, were insensitive to the mating system. The absence of variation in Tajima's
534 D could be explained if directional selection had already fixed the optimal alleles prior to the
535 transition from allogamy to autogamy, which could also explain the signal of positive selection
536 detected in PEGs in *A. thaliana* and *C. rubella* (Hatorangan et al., 2016; Tuteja et al., 2019).
537 However, this explanation does not account for the high π_{NS}/π_S observed in allogamous and
538 autogamous populations. Theoretical studies are necessary to estimate the cumulated effect
539 of selection related to kinship theory, imprinting, demographic history and mating system.

540 *Imprinted genes present signal of ancient and recent coevolution*

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

548 For instance, one network based on branch-length correlations was enriched in enzymatic
549 activities related to cell wall formation and modification, a key process during endosperm
550 cellularization in Arabidopsis (Hehenberger et al., 2012), which in turn influences seed size
551 (Pignatta et al., 2018; Zhang et al., 2020), a likely target of parental conflict. These correlated
552 evolutionary rates could reflect equivalent selective constraints (Galtier and Rousselle 2020)
553 or just a common chromosomic context (e.g. transposable element) that modify mutation rate
554 (Lynch 2010). Interestingly, this network appeared to be strongly dependent on a MEG, an
555 AGAMOUS-like protein, characterized also by reduced branch lengths, consistent with
556 expectations that coevolution may drive coordinated evolution among genes primarily through
557 rare compensatory mutations (Fraser et al., 2002)

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

562 These signals of coevolution could promote the maintenance of the gene functions (Clark et
563 al., 2012), limit their evolutionary independency (Yang and Bielawski 2000) and be another
564 source of reproductive barrier (Presgraves 2010). Interestingly, the coevolution signal did not

565 involve increased site-by-site coevolution, suggesting that protein–protein molecular
566 interactions, often drivers of coevolution (Green et al., 2021), played a limited role here.
567 Instead, the observed patterns may reflect selection acting on gene dosage rather than protein
568 products (Clark et al., 2012).

569 However, as correlations between evolutionary statistics can arise from mutation-rate
570 variation, recombination environment or shared functional constraints, we interpret them as
571 exploratory signals only. The more stringent CoMap analysis (Dutheil & Galtier 2007) did not
572 reveal enrichment for residue-level co-evolution, indicating that correlation-based networks
573 should be treated cautiously and do not by themselves confirm co-evolution.

574 *Main limitations in our study*

575 Although our study combines multiple phylogenetic and population genomic statistics, the
576 signals detected were not consistently congruent across approaches. This inconsistency may
577 reflect evolutionary processes not accounted for, (such as demographic history or
578 chromosomal context) as well as several methodological limitations, outlined below.

579 First of all, estimating Tajima's D from pooled sequencing data is challenging because SFS-
580 based statistics are sensitive to coverage variation and sampling noise, which increase
581 estimation variance (Futschik & Schlötterer 2010; Ferretti et al. 2013; Lynch et al. 2014).
582 Although we minimized dataset-specific biases by comparing genes processed under identical
583 conditions, technical noise cannot be fully excluded. In addition, short genes, such as many
584 MEGs or control genes, yield fewer segregating sites, reducing the reliability of Tajima's D
585 (Tajima 1989).

586 A second limitation concerns the imprinting dataset. The set of imprinted genes analysed was
587 derived from a study on a geographically and genetically distinct allogamous subspecies
588 European *A. lyrata petraea* (Klosinska et al., 2016), assuming conservation of imprinting,
589 because no within-lineage data on imprinting in *A. lyrata* exist. However, imprinting shows
590 limited conservation within *A. thaliana* (Gehring et al., 2011; Wolff et al., 2011; Hsieh et al.,
591 2011; Pignatta et al., 2014), between *A. thaliana* and *A. lyrata* in our data, and generally across
592 species (Eaux et al., 2013; Hatorangan et al., 2016; Roth et al., 2018; Flores-Vergara et al.,
593 2020). It is therefore possible that several genes classified as imprinted in our analyses are
594 not imprinted in the studied populations , especially given known population-level variation in
595 TE-driven imprinting (Padilla-García et al., 2025). Future population-specific characterizations
596 of imprinting will be critical to resolve this issue..

597 Moreover, some imprinted genes identified by Klosinska et al. (2016) may have been
598 inadvertently excluded from our dataset due to our method for detecting tissue-specific
599 expression, which relied on RNA-seq data pooled from both allogamous and autogamous
600 populations. More controlled and population-specific experimental designs will be required to
601 recover the complete set of endosperm-expressed imprinted genes relevant to parental
602 conflict.

603 A last important limitation is about the panels of genes compared (control, endosperm and
604 imprinting genes). To be congruent with previous studies (Spillane et al., 2007; Miyake et al.,
605 2009; Hatorangan et al., 2016; Tuteja et al., 2019), we also choose genes in panels based on
606 expression patterns. However, the three gene panels differ substantially in gene size, CDS
607 proportion, and GC content, factors known to influence both polymorphism and phylogenetic
608 metrics, but seldom considered in previous studies (Spillane et al., 2007; Miyake et al., 2009;
609 Hatorangan et al., 2016; Tuteja et al., 2019).

610 Endosperm genes and PEGs are substantially longer, whereas MEGs and control genes are
611 shorter. The size of the gene is an important parameter that could modify an important amount
612 of statistic. For example, larger genes tend to evolve more slowly due to increased functional
613 constraint and pleiotropy (Massey 2013; Lopes et al., 2021), display higher conservation and
614 reduced turnover (Abdel-Haleem 2007; Koonin 2015), and can inflate the number of species
615 with detectable homologues. In this sense, we found a positive correlation between the
616 number of species with homologous genes and the gene size. This may explain partially the
617 higher homology counts observed for PEGs and endosperm genes relative to control genes.
618 Conversely, we found that the gene size is negatively correlated with polymorphism and
619 Tajima's D. However, this tendency contradict that longer genes as endosperm genes
620 presented higher Tajima's D or that the shorter genes as the MEGs were associated to lower
621 polymorphism, suggesting that, maybe, the effects of parental conflict on polymorphism in
622 these genes are underestimated, more specifically for MEG because shorter genes may yield
623 less stable estimates of Tajima's D (Tajima 1989). We found also a negative correlation with
624 the expression level, that could explain partially why the expression of MEG is more efficient,
625 even if the group of gene stays the more important parameter predicted by AIC, but could not
626 explain why the expression of PEG is higher than endosperm genes.

627 The effect of gene size on branch length is more debated: while longer genes trivially
628 encompass more substitutions (Graur & Li 2000), several studies show that long, constrained
629 genes tend to exhibit shorter branches due to stronger purifying selection (Kimura 1983;
630 Abdel-Haleem 2007; Massey 2013; Koonin 2015). These opposing forces complicate
631 interpretation of branch-length differences between MEGs, PEGs and endosperm genes. In

our study, we found a positive correlation between the length of the branches and the gene size, maybe explained by the amount of endosperm genes with long branches. This correlation suggests that the longer branches observed in PEGs and endosperm genes are partially explained by the size of the genes, but could not explain the longer branches observed in MEGs.

Moreover, both imprinted and endosperm genes show elevated CDS content relative to controls. Thus, this parameter could not explain the increase of efficiency observed in expression of imprinted genes, despite the negative correlation detected. In the same way, this parameter could not explain the reduced π_{NS}/π_S observed in imprinted genes compared to endosperm genes, despite the positive correlation detected. Interestingly, high CDS proportion is associated with stronger purifying selection (Hughes et al., 2003), which can reduced the length of the branches, as observed in our study. However, the increase of length of branches detected in endosperm and imprinted genes is the reverse pattern predicted, suggesting that the effect on length of branches in these genes is maybe underestimated.

Finally, imprinted genes show reduced GC content relative to endosperm and control genes. GC-poor genes tend to have reduced recombination rates, lower polymorphism (Wright et al., 2003), and higher AT-biased substitution rates, which can artificially lengthen branches (Duret & Galtier 2009), and explain the longer branches observed in PEG and endosperm genes. However, contrary to these expectations, we found a negative collation between the length of the branches and the GC content, maybe explain by the important amount of endosperm genes. Interestingly, we found also a negative correlation between the polymorphism and the GC proportion in our study, that suggest that the reduced polymorphism observed in MEGs is maybe underestimated. Low GC also reduces the likelihood of finding conserved homologues across species (Urrutia & Hurst 2003), as found in our study, but is not able to explain why the PEGs and the endosperm genes are associated to higher number of species with homologous genes despite lower GC content than in control genes. Low GC is expected to also inflate D_{NS}/D_S by increasing the probability of non-synonymous changes (Kryazhimskiy & Plotkin 2008). However, we detected the reverse correlation in our study, that could explain why imprinted genes are associated to a reduced D_{NS}/D_S.

The lower GC of imprinted genes is also unexpected given their elevated expression (Sharp & Li 1987; Bernardi 2000), suggesting that expression differences may be underestimated. For these reasons, although the three panels differ in genomic features, endosperm-upregulated genes remain the only biologically appropriate baseline for testing imprinting-related hypotheses. A feature-matched control panel would reduce variation but would not provide a more relevant comparison for detecting signatures associated with parental conflict.

Together, these factors indicate that differences in gene size, CDS proportion, and GC content may partly account for the heterogeneity of evolutionary signals observed across statistics. Some signatures consistent with parental conflict may therefore be underestimated or confounded, while others may reflect intrinsic genomic properties rather than imprinting-related selective pressures. Future work will require gene panels matched not only for expression, but also for fine-scale genomic features.

Although several genomic features differ among gene panels, we performed targeted robustness checks to ensure that competing non-selective scenarios did not spuriously generate the patterns we report. In particular, the SNP-density threshold retained comparable fractions of MEGs, PEGs, endosperm genes and controls, ruling out retention bias as a driver of polymorphism contrasts. Moreover, all evolutionary inferences were tested against empirical null distributions from endosperm-upregulated genes, which share expression patterns and chromatin context with imprinted genes, limiting confounding effects of gene architecture or local mutation rate. These checks do not eliminate all alternative explanations, but they demonstrate that the signals observed are unlikely to arise solely from genomic artefacts or filtering procedures.

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.

We used several GitHub repositories, each corresponding to a distinct analytical module. This structure reflects the fact that the datasets and workflows differ substantially across analyses and that several scripts were developed independently and used in other publications. For clarity, all repositories are indexed in a Zenodo record (https://zenodo.org/records/17954040), and each includes documentation, test data, and commented scripts. This multi-repository organization was reviewed and approved for reproducibility and FAIR compliance by the PCI Ecology Data Editor (Checklist PCI Ecology 867).

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.

727 *Genes clusters*

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

732 To disentangle signatures of selection associated specifically with the endosperm, where
733 conflicts over resource allocation are expected, from those potentially arising through other
734 selective processes, we separated imprinted genes upregulated in the endosperm (14 PEGs
735 and 15 MEGs) from imprinted genes not showing endosperm-specific upregulation (32 PEGs
736 and 12 MEGs), as determined by differential expression analysis.

737 To evaluate whether the observed proportions of MEGs and PEGs upregulated in endosperm
738 deviated from theoretical expectations under kinship theory, we compared them with null
739 distributions generated by 10,000 random sampling of 49 and 35 genes (size of PEGs and
740 MEGs groups) between the 27,820 genes expressed genes.

741 For all subsequent analyses, we focused exclusively on endosperm-upregulated imprinted
742 genes and compared them with two reference groups: the set of 2,703 non-imprinted
743 endosperm-upregulated genes (hereafter, “endosperm genes”), the sets of 15 MEG and 14
744 PEG upregulated in endosperm (hereafter, “MEG” and “PEG”) and a set of 100 randomly
745 selected ubiquitously expressed genes (hereafter, “control genes”). To remain consistent with
746 previous studies of imprinting and parental conflict (Spillane et al., 2007; Miyake et al., 2009;
747 Hatorangan et al., 2016; Tuteja et al., 2019), the control panel was defined on the basis of
748 expression patterns only. This set is intended to represent a genome-wide baseline unrelated
749 to endosperm-specific functions or parental conflict, rather than a genomic feature–matched
750 control. Consequently, no constraints were imposed on gene size, GC content, CDS
751 proportion, or chromosomal position. However, for each statistics studied after, we compared
752 the full model included gene group, gene length, CDS proportion and GC content as
753 explanatory variables using the Akaike Information Criterion (AIC). We then evaluated a series
754 of reduced models in which one parameter was removed at a time and retained the model
755 with the lowest AIC value. This procedure was iterated until the best-supported model was
756 identified using the function StepAIC of the MASS library (version 7.3-65) in R. This approach
757 allowed to determine (1) whether gene group remained a significant predictor in the best model
758 and (2) whether genomic features accounted for the observed trends in the analysed statistics.

759 To test whether expression levels conformed to predictions of higher optimal expression in
760 PEGs than in MEGs, we normalized expression by gene copy number and compared
761 observed values with bootstrap distributions from non-imprinted endosperm genes (10,000
762 resamples of 29 genes). To account for gene dosage, expression values were normalized by
763 the expected number of expressed copies: three for non-imprinted endosperm genes, two for
764 MEGs, and one for PEGs.

765 Based on the sequences from the reference genome of *A. lyrata* (Hu et al., 2011;
766 https://doi.org/10.5281/zenodo.17287897), we conducted bootstrap analyses to compare the
767 gene size, the GC and the CDS content (%) of PEGs and MEGs with values obtained from
768 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we also
769 conducted bootstrap analyses to compare the values on control genes with values obtained
770 from 10,000 resamples of 100 endosperm genes.

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

773 *Estimation of codon and amino acid bias*

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

784 To test if the overused or underused codons were related to gene expression, we compared
785 the variations of PHE score induced by codons with the score of unbiased codons by
786 permutations test with 1,000 replications for MEGs and PEGs independently. The PHE score
787 was estimated based on overexpressed genes in *A. thaliana* and reflect the importance of a
788 codon on genes expression (Sahoo et al,. 2019). We hypothesized that codons overused in
789 imprinted genes were associated to increases of PHE score because associated to increases
790 of gene expression whereas the underused codon were associated to reduce PHE scores.

791 The code used for permutation test was available in GitHub
792 (https://github.com/leveveaudrey/permutation-test-)

793 *Distribution of the genes across Brassicaceae*

794 We assessed whether the imprinted genes presented specific signals of selection compared
795 to the endosperm and control genes in the phylogeny of 22 Brassicaceae. All the complete
796 sequence of reference genomes (fasta files) were extracted from the Phytozome database
797 (Goodstein et al., 2012; Table S4). We extracted the sequences of the genes in the *A. lyrata*
798 reference genome (Hu et al., 2011) using bedtools V2.30 (Quinlan et al., 2010). Then, we
799 retained the best hit of these sequences on 22 other Brassicaceae using BLAST V2.10
800 (Altschul et al., 1990) with e-value < 1e-5. If more of one hit was found for a gene in one
801 specie, we filtered the best hit as the hit with: the higher length alignment, the higher % of
802 identity, then the higher coverage on the gene in *A. lyrata* and the higher reciprocal coverage
803 on the gene in the other species. Then we extracted the sequence of the best hit for each
804 species of each gene. These steps were integrated into a pipeline of analysis available in
805 GitHub (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee).

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

812 *Divergence of the imprinted genes across Brassicaceae*

813 The rates of nonsynonymous mutations (D_{NS}) and synonymous mutations (D_S) are assumed
814 to follow the neutral evolutionary process and the ratio D_{NS}/D_S is therefore approximate to the
815 selective pressure on the protein product of a gene. An increase of the D_{NS}/D_S in genes
816 indicates positive selection, while a decrease indicates negative selection. To test whether
817 imprinted genes display accelerated evolutionary rates compared to endosperm genes, using
818 the annotation of the 0-fold and 4-fold degenerate sites in *A. lyrata*, we estimated for each
819 gene the mean evolutionary rates D_{NS}/D_S between the reference genome of *A. lyrata* and all
820 the species with an homologue found. Only the genes found in at least 3 species and with a
821 D_S higher than 0 were considered. These steps were integrated into a pipeline of analysis
822 available in GitHub (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee).

823 To test for significant differences in the D_{NS}/D_S across the group of genes, we conducted
824 bootstrap analyses to compare the value on PEGs and MEGs with values obtained from
825 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we
826 conducted bootstrap analyses to compare the values on control genes with values obtained
827 from 10,000 resamples of 100 endosperm genes. The R code for the bootstrap test is available
828 in GitHub (https://github.com/leveveaudrey/bootstrap).

829 *Topology of the genes across Brassicaceae*

830 For each gene, we aligned the gene and all its best hits found across the different species in
831 with Muscle V3.8.155 (Edgar 2004). Then, we built a phylogenetic tree following the HKY85
832 model for each gene with phyml V3.3 (Guindon et al., 2010). In parallel, we concatenated the
833 alignment of the 1,055 endosperm genes found in the 22 others Brassicaceae to build a
834 phylogenetic tree of the endosperm, without branch length, using phyml V3.3 (Guindon et al.,
835 2010). This tree was used as a proxy of the topology of the endosperm. These steps were
836 integrated into a pipeline of analysis available in GitHub
837 (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee). Then, we estimated
838 the Robinson-Foulds distance between the topology of this tree with the topology of the tree
839 of each gene obtained previously using the R package TreeDist. Then, to test for significant
840 differences between the RF distances across the group of genes, we conducted bootstrap
841 analyses to compare the value on PEGs and MEGs with values obtained from 10,000
842 resamples of 29 endosperm genes and control genes independently. Similarly, we conducted
843 bootstrap analyses to compare the values on control genes with values obtained from 10,000
844 resamples of 100 endosperm genes. The R code for the bootstrap test is available in GitHub
845 (https://github.com/leveveaudrey/bootstrap).

846 *Variation of the length of branches across Brassicaceae*

847 We investigated whether the imprinted genes presented variation of the branch length
848 compared to the endosperm and control genes. For each gene, we estimated the length of
849 branches for each species with the topology of the phylogenetic tree obtained with the 1,055
850 endosperm genes using phyml V3.3 (Guindon et al., 2010). These steps were integrated into
851 a pipeline of analysis available in GitHub
852 (https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee). We extracted the
853 branches length for each gene with the R package 'ape' version 5.7-1. If the gene was not
854 present in a species, the value was replaced by "NA". To test for significant differences in the
855 branch length across the group of genes, we conducted bootstrap analyses to compare the
856 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, DePristo et al., 2011) with the option GVCF and a quality score threshold of 60 using vcftools v0.1.15 (Danecek et al., 2011), to retain only highly reliable SNPs. This stringent cutoff is commonly used in high-coverage datasets to minimise false positives, which is particularly important for downstream analyses of selection and co-evolution that are sensitive to SNP-calling errors (Schlötterer et al., 2014; GATK Best Practices). We excluded sites with less than 15 reads aligned. Sites covered by at least 15 reads but containing no SNP were considered monomorphic. The final number of sites in each sample set is summarised in Table S6. The command lines used for all these steps are available in Github (https://github.com/leveveaudrey/alignment-and-snp-calling).

Effect of imprinting 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.

For each gene of each population, we estimated the π, the π_{NS}/π_S for all positions and the Tajima's D of each gene using a custom python code (https://github.com/leveveaudrey/POOLSEQ-analysis-of-polymorphism). Briefly, we estimate the Tajima's D by gene following the classical equation described in Tajima (1989): $D = (\pi - (S / \Sigma(1/i))) / \sqrt{e1S + e2S\,(S-1)}$, with e1 = (n+1) /3 * (n+1) - (1/ Σ(1/i))/ Σ(1/i) and

891 $e2= (2 * (n * n + n + 3)) / (9 * n * (n-1)) - ((n+2) /(\Sigma(1/i) *n-1)) + (\Sigma(1/i^2) / (\Sigma(1/i)^2)) / (\Sigma(1/i)^2$
892 $+ \Sigma(1/i^2))$. S is for the number of segregating sites, i for value between 1 and n-1 haplotypes
893 and n is for the number of haplotypes (number of individuals multiplied by two for diploid).

894 If a position in the *A. lyrata* genome was covered but not polymorphic, the π was set to 0.
895 Then, the 0-fold and 4-fold degenerated sites were identified and extracted from the reference
896 genome and the genome annotation using the script NewAnnotateRef.py (Williamson et al.,
897 2014; updated version available in
898 https://github.com/leveveaudrey/phylogenetic_analyse_brassicacee). Because all mutations
899 at 0-fold degenerated sites alter the sequence of the encoded protein, we assumed that these
900 mutations are non-synonymous (NS). In contrast, mutations at the 4-fold degenerate sites
901 never alter the encoded amino acid, thus we assumed that these mutations are synonymous
902 (S). For the sake of simplicity, we discarded mutations at 2- or 3-fold degenerate sites. Based
903 on the π estimated for each site, we estimated the mean π_{NS} and π_S, for each gene in each
904 population, in order to estimate the π_{NS}/π_S. We removed the genes with a π_S to 0. For the
905 Tajima's D, we applied a SNP-density threshold (≥0.5% polymorphic sites) to ensure reliable
906 estimates. This density-based criterion avoids gene-length biases inherent to minimum-SNP
907 thresholds, as genes vary widely in size across categories. We verified that this filtering did
908 not disproportionately exclude any gene category: the proportions of retained genes were
909 comparable across endosperm and imprinting genes and across imprinting and control genes
910 (Table S7).

911 To test for significant differences across the group of genes, we conducted bootstrap analyses
912 to compare the mean π, π_{NS}/π_S and Tajima's D of PEGs and MEGs with means obtained from
913 10,000 resamples of 29 endosperm genes and control genes independently. Similarly, we
914 conducted bootstrap analyses to compare the values on control genes with values obtained
915 from 10,000 resamples of 100 endosperm genes. The R code for the bootstrap test is available
916 in GitHub (https://github.com/leveveaudrey/bootstrap). We performed the test on allogamous
917 and autogamous populations independently. Finally, for each statistic and group
918 independently, we tested for the mating system effect by a performing Kruskal-Wallis test as
919 implemented in R version 4.3.2.

920 *Genes in coevolution*

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

925 Correlations between evolutionary rates or diversity metrics were used as exploratory
926 indicators of evolutionary covariance, following approaches such as evolutionary rate
927 covariation (ERC), which has been widely applied to detect co-functionality or shared selective
928 pressures (Clark et al., 2012). Because such correlations may reflect genomic or mutational
929 confounders, we used endosperm-upregulated genes as an empirical baseline for defining
930 extreme thresholds.

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

939 To detect coevolution based on topology with other genes, we used first the distribution of RF
940 distances between endosperm genes and the global endosperm topology to estimate the
941 value obtained in 95% of the distribution. Then, we excluded all the imprinted and endosperm
942 genes having lower RF distance than this value to remove genes whose topology was
943 explained by the global topology. Finally, we compared the topology of each remaining
944 imprinted genes and endosperm genes. All pairs of genes with a null RF distance were
945 considered as significantly similar and used to build the last coevolution network.

946 For each network of coevolution, we compared the number of pairs of imprinted genes directly
947 correlated with the number found between genes in 1,000 random samples of 29 endosperm
948 genes. Moreover, we tested for gene enrichment on each GO term in each network based on
949 the occurrences of each term across 1,000 random samples of equivalent size of all the
950 endosperm and imprinted genes. The complete list of GO term associated to each gene was
951 downloaded from Gene Ontology online platform for *A. thaliana* (Ashburner et al,. 2000).

952 *Identification of sites in coevolution*

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

958 Coevolving sites within each pair of genes were predicted using the clustering method
959 implemented in the CoMap package (version 1.5.2; https://jydu.github.io/comap/; Dutheil and
960 Galtier 2007). The built-in algorithm was then used to correct for multiple testing and to
961 estimate the false discovery rate (FDR). In this study, 1,000 simulations per tested pair were
962 performed, and the groups of sites with an adjusted p value < 1% were considered coevolving.
963 Simple correlation was used as a measure of coevolution. For each comparison pair, we
964 extracted the coevolving groups of substitutions and classified them by size (between 2 and
965 10).

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

973 **Data availability**

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

977 **Acknowledgements**

978 This work has been supported by the Czech Science Foundation (GACR 22-20240S granted
979 to C.L-P). Computational resources were supplied by the project 'e-Infrastruktura CZ' (e-
980 INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech
981 Republic. A.L thanks personally Stella Huynh for our beneficial suggestions and
982 commentaries.

983 **Author contributions**

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

987 **Declaration of interests**

988 The authors declare no competing interests.

989 **Bibliography**

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

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

995 3) Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P.,
996 Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L.,
997 Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M.,
998 Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25,
999 25–29. https://doi.org/10.1038/75556

1000 4) Bemer, M., Heijmans, K., Airoldi, C., Davies, B., Angenent, G.C., 2010. An Atlas of
1001 Type I MADS Box Gene Expression during Female Gametophyte and Seed
1002 Development in Arabidopsis. Plant Physiology 154, 287–300.
1003 https://doi.org/10.1104/pp.110.160770

1004 5) Bernardi, G., 2000. Isochores and the evolutionary genomics of vertebrates. Gene 241,
1005 3–17. https://doi.org/10.1016/s0378-1119(99)00485-0

1006 6) Botet, R., and Joost J. B. Keurentjes., 2020. The Role of Transcriptional Regulation in
1007 Hybrid Vigor. Frontiers in Plant Science 11 (April).
1008 https://doi.org/10.3389/fpls.2020.00410

1009 7) Brandvain, Y., Haig, D., 2005. Divergent mating systems and parental conflict as a
1010 barrier to hybridization in flowering plants. Am Nat 166, 330–338.
1011 https://doi.org/10.1086/432036

1012 8) Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., Meinke, D., 2010. Identification of
1013 Nuclear Genes Encoding Chloroplast-Localized Proteins Required for Embryo
1014 Development in Arabidopsis. Plant Physiology 155, 1678.
1015 https://doi.org/10.1104/pp.110.168120

1016 9) Clark, N.L., Alani, E., Aquadro C.F., 2012. Evolutionary Rate Covariation Reveals
1017 Shared Functionality and Coexpression of Genes. Genome Research 22 (4): 714–20.
1018 https://doi.org/10.1101/gr.132647.111

1019 10) Cvijović, I., Good, B.H., Desai, M.M., 2018. The Effect of Strong Purifying Selection on
1020 Genetic Diversity. Genetics 209 (4): 1235–78.
1021 https://doi.org/10.1534/genetics.118.301058

1022 11) Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A.,
1023 Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000
1024 Genomes Project Analysis Group, 2011. The variant call format and VCFtools.
1025 Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

1026 12) Day, T., Bonduriansky, R., 2004. Intralocus sexual conflict can drive the evolution of
1027 genomic imprinting. Genetics 167, 1537. https://doi.org/10.1534/genetics.103.026211

1028 13) Day, R.C., Herridge, R.P., Ambrose, B.A., Macknight, R.C., 2008. Transcriptome
1029 Analysis of Proliferating Arabidopsis Endosperm Reveals Biological Implications for
1030 the Control of Syncytial Division, Cytokinin Signaling, and Gene Expression
1031 Regulation. Plant Physiology 148, 1964–1984. https://doi.org/10.1104/pp.108.128108

1032 14) Dekkers, B.J.W., Pearce, S., van Bolderen-Veldkamp, R.P., Marshall, A., Widera, P.,
1033 Gilbert, J., Drost, H.-G., Bassel, G.W., Müller, K., King, J.R., Wood, A.T.A., Grosse, I.,

1034 Quint, M., Krasnogor, N., Leubner-Metzger, G., Holdsworth, M.J., Bentsink, L., 2013.
1035 Transcriptional Dynamics of Two Seed Compartments with Opposing Roles in
1036 Arabidopsis Seed Germination. Plant Physiology 163, 205–215.
1037 https://doi.org/10.1104/pp.113.223511

1038 15) DePristo, M.A., Banks, E., Poplin, R.E., Garimella, K.V., Maguire, J.R., Hartl, C.,
1039 Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J.,
1040 Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly,
1041 M.J., 2011. A framework for variation discovery and genotyping using next-generation
1042 DNA sequencing data. Nat Genet 43, 491–498. https://doi.org/10.1038/ng.806

1043 16) Dutheil, J. ,Galtier, N., 2007. Coevolution within and between Genes.
1044 https://doi.org/10.1159/000107599

1045 17) Eaux, A.J., Bilinski, P., Eichten, S.R., Vaughn, M.W., Ross-Ibarra, J., Gehring, M.,
1046 Reprinte, N.M., 2013. Comprehensive analysis of imprinted genes in maize reveals
1047 allelic variation for imprinting and limited conservation with other species. Proceedings
1048 of the National Academy of Sciences 110, 19639–19644.
1049 https://doi.org/10.1073/pnas.1309182110

1050 18) Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and
1051 high throughput. Nucleic Acids Res 32, 1792–1797.
1052 https://doi.org/10.1093/nar/gkh340

1053 19) Edwards, S.V., 2009. Natural selection and phylogenetic analysis. Proceedings of the
1054 National Academy of Sciences 106, 8799–8800.
1055 https://doi.org/10.1073/pnas.0904103106

1056 20) Ferretti, L., Ramos-Onsins, S.E., Pérez-Enciso, M., 2013. Population genomics from
1057 pool sequencing. Mol Ecol 22, 5561–5576. https://doi.org/10.1111/mec.12522

1058 21) Florez-Rueda, A.M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U., Städler, T.,
1059 2016. Genomic Imprinting in the Endosperm Is Systematically Perturbed in Abortive
1060 Hybrid Tomato Seeds. Mol Biol Evol 33, 2935–2946.
1061 https://doi.org/10.1093/molbev/msw175

1062 22) Flores-Vergara, M.A., Oneal, E., Costa, M., Villarino, G., Roberts, C., De Luis
1063 Balaguer, M.A., Coimbra, S., Willis, J., Franks, R.G., 2020. Developmental Analysis of
1064 Mimulus Seed Transcriptomes Reveals Functional Gene Expression Clusters and
1065 Four Imprinted, Endosperm-Expressed Genes. Front. Plant Sci. 11.
1066 https://doi.org/10.3389/fpls.2020.00132

1067 23) Foxe, J.P., Stift, M., Tedder, A., Haudry, A., Wright, S.I., Mable, B.K., 2010.
1068 Reconstructing Origins of Loss of Self-Incompatibility and Selfing in North American
1069 Arabidopsis Lyrata: A Population Genetic Context. Evolution 64, 3495–3510.
1070 https://doi.org/10.1111/j.1558-5646.2010.01094.x

1071 24) Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., Feldman, M.W., 2002.
1072 Evolutionary Rate in the Protein Interaction Network. Science 296, 750–752.
1073 https://doi.org/10.1126/science.1068696

1074 25) Futschik, A., Schlötterer, C., 2010. The Next Generation of Molecular Markers From
1075 Massively Parallel Sequencing of Pooled DNA Samples. Genetics 186, 207–218.
1076 https://doi.org/10.1534/genetics.110.114397

1077 26) Galtier, N., Rousselle, M., 2020. How Much Does Ne Vary Among Species? Genetics
1078 216, 559–572. https://doi.org/10.1534/genetics.120.303622

27) 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

28) 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

29) 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

30) 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

31) 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

32) 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

33) 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

34) 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

35) 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

36) 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

37) 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

38) 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

39) 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

40) 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

41) 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

42) 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

43) 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

44) 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

45) 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

46) 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

47) 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

48) 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

49) 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

50) 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

51) 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

52) 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

53) 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

54) 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

55) 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

56) 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

57) 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

58) 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

59) 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

60) 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

61) 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

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

63) Lynch, M., Bost, D., Wilson, S., Maruki, T., Harrison, S., 2014. Population-Genetic Inference from Pooled-Sequencing Data. Genome Biol Evol 6, 1210–1218. https://doi.org/10.1093/gbe/evu085

64) 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

65) 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

66) 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

67) 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

68) 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

69) 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

70) 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

71) 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

72) 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

73) 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

74) 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

75) 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

76) 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

77) 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

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

79) 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

80) 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

81) 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

82) 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

83) 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

84) 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

85) 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

86) 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

87) 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

88) 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

89) 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.

90) 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

91) 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

92) 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

93) 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

94) 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

95) 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

96) 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

97) 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

98) 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.

99) 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

100) 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

101) 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

102) 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

103) 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

104) 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

105) 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

106) 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

107) 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

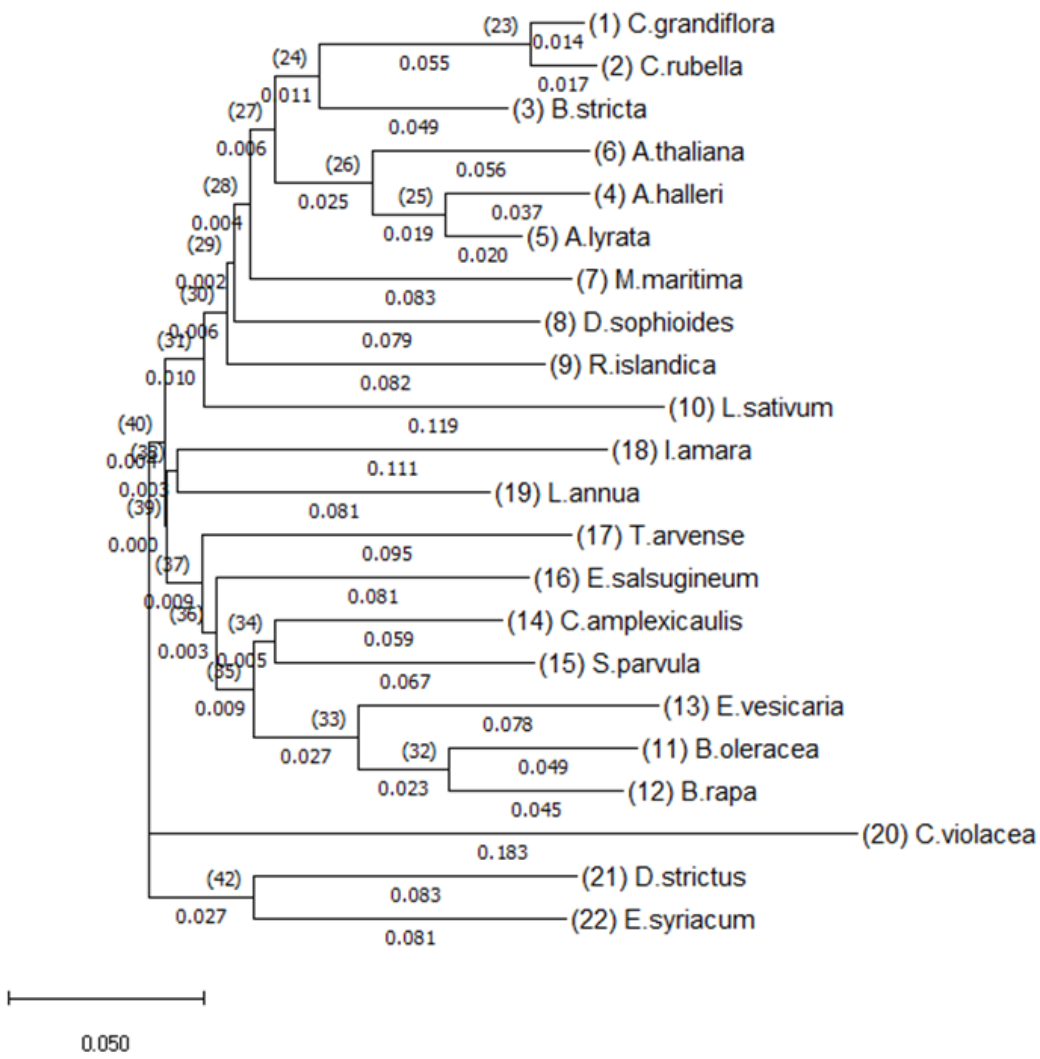


Figure S1: phylogenetic tree obtained by HKY85 model based on the 1055 endosperm genes found in all species after 1000 bootstrap. *The percentage of trees in which the associated haplotypes clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in mean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