The effect of parental conflict in imprinting genes in *A. lyrata.*

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

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

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

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

Summary

1) Genomic imprinting is a form of gene regulation resulting in the unequal expression of the two parental alleles. A common assumption is that the parental conflict over resource allocation to the offspring promotes positive directional on imprinted genes, but most studies addressing this question have focused on autogamous species, where parental conflict is expected to be low. Moreover, parental conflict is expected to promote coevolution between imprinted genes, but this has never been tested so far.
2) We investigated and compared footprints of selection on imprinted genes in allogamous and autogamous populations of *Arabidopsis lyrata*. We then searched for signs of coevolution based on phylogenetic across the Brassicaceae family approaches.
3) The imprinted genes upregulated on endosperm present a general increase in the evolutionary rate in the *Brassicaceae* family associated with an ancient signal of negative selection. In allogamous populations, this signal of negative selection is weaker and associated only on PEG. However, in autogamous populations, the selection on imprinted genes appeared to be relaxed. Finally, we found signals of coevolution between some imprinted genes.
4) The parental conflict in endosperm of allogamous populations promotes negative selection on imprinted genes and coevolution between some genes.

Key words: genomic imprinting, parental conflict, positive Darwinian selection, endosperm, coevolution, mating type, Arabidopsis lyrata.

6487 words, 4 figures, 1 supplementary figure, 13 supplementary tables

Introduction

Genomic imprinting is a form of epigenetic gene regulation in flowering plants and mammals in which alleles of genes are expressed in a parent-of-origin dependent manner (Haig,1997; Wilkins, 2011). In plants, imprinting is initiated by distinct epigenetic landscapes between male and female gametophytes, stably inherited after fertilisation (Batista & Köhler, 2020). The genomic imprinting has motivated a variety of theories attempting to explain its existence (Haig & Weston, 1989; Haig,1997; Spencer & Clark, 2014). The most frequently invoked theory is the parental conflict hypothesis (Haig, 2000). If the maternal parent directly provides growing progeny and has offspring by multiple males, the reproductive fitness of the mother is optimal when providing equal resources among all her offspring, while the father's fitness improves with the preferential allocation of maternal resources to his own offspring. The parental conflict theory posits that imprinting is selected to resolve this conflict over the allocation of resources to the embryo between the maternal and paternal genomes.

In plants, parental conflict is expected at the level of the endosperm, the embryo's nourishing tissue in the seed. Moreover, growth-limiting genes are expected to be under maternal control, while growth-promoting genes should be under paternal control (Haig & Weston, 1989; Halg & Graham,1991; Moore & Haig, 1991). Consistent with a conflict over resource allocation, many imprinted genes appear to be involved in resource accumulation and metabolism (Raissig et al., 2011). Moreover, an increase in maternal genome dosage (cross between a tetraploid mother and a diploid father) produces small seeds, whereas a reciprocal cross (increased paternal dosage) produces larger seeds in *Arabidopsis thaliana* (Scott et al., 1998). However, the effects on seed size probably result from an interaction between cytoplasmic effects, dosage effects and genomic imprinting.

Such conflict should lead to directional selection on coding sequences of imprinted genes expressed from the maternally and paternally inherited genomes (Wilkins & Haig, 2001; Mills & Moore, 2004; Haig, 2014). According to this assumption, signatures of positive selection have been detected at the imprinted *MEDEA* gene in the allogamous plant *A. lyrata* but not in the autogamous sister species *A. thaliana* (Spillane and al., 2007; Miyake and al., 2009). Evidence of positive selection has also been reported for some Paternally Expressed Genes (PEG) of the autogamous species *Capsella rubella* and *A. thaliana* but not on Maternally Expressed Genes (MEG; Hatorangan et al., 2016; Tuteja et al., 2019). These observations raised the question of the generality of the implication of parental conflict on this directional selection detected, because parental conflict is not expected in autogamous species, and on the general selection expected on imprinting genes. Thus, the signature of selection

expected on imprinted genes and the implication of the parental conflict remains unclear in plants.

Another consequence of parental conflict expected at the molecular level is the coevolution of the reciprocally imprinted genes having antagonistic effects on offspring growth. This coevolution could be explained by direct interactions between the genes (Wilkins & Haig, 2001; Mills & Moore, 2004; Haig, 2014) or indirectly, by interactions with other genes implicated in parental conflict (Willi, 2013). Population genomic methods may be used to detect coevolving coding sequences. These methods used the variations on allelic frequencies to estimate coevolution. The main advantage of these methods is the possibility of studying all polymorphic genes, whatever the distribution of this gene across species. This point is particularly important for imprinted genes that exhibited low conservation between *A. thaliana* and *C. rubella* (Hatorangan et al., 2016). However, these methods require polymorphism across samples and could be biased by different factors such as genetic distance, demographic events or genetic structuration. More holistic approaches are based on reconstructing the evolutionary history of candidate coevolving genes, assuming that their evolutionary history should be common if they coevolve and also likely diverge from the evolutionary history of the rest of the genome (Dutheil & Galtier, 2007). Most of these methods consist of comparisons of the phylogenies, sometimes completed by functional data about the genes, to finally detect co-occurring changes in a phylogeny. However, to our knowledge, the signals of coevolution between imprinted genes, based on population genomics as on phylogenetics, were never investigated so far.

To determine whether genomic imprinting in allogamous species is associated with directional selection and parental conflict, we analysed the selective pressures acting on imprinted genes of *A. lyrata* (Klosinska et al., 2016). *A. lyrata* is a hermaphroditic plant species, close to the model species *A. thaliana*, which has cross-pollinating and selfing populations in North America (Ross-Ibarra et al., 2008). This species is, therefore, perfect for investigating and studying the effect of genomic imprinting and the mating type on polymorphism on imprinted genes. Using genomic data from nine allogamous and nine autogamous populations of *A. lyrata*, we searched: 1) the selective pressure acting on imprinted genes in allogamous populations and 2) the implication of mating type on this selective pressure. Moreover, based on phylogenetic variations of the imprinted genes across 23 species of *Brassicaceae*, we investigated 3) evidence of coevolution between imprinting genes and the other genes upregulated in the endosperm. Our findings suggest that parental conflict promotes negative selection on some imprinted genes but diversification on non-imprinted genes, specifically in allogamous populations.

Materials and Methods

We expected that parental conflict promotes directional selection on the imprinted genes compared to the other endosperm genes, specifically on allogamous populations. To detect signals of selection on imprinted genes and the implication of the parental conflict on the selection, we developed an approach based on the comparison between the phylogenetic pattern and the variation of polymorphism of the MEG and the PEG in *A. lyrata* (Klosinska et al., 2016) with the 2703 other genes upregulated in endosperm.

Expression data

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

Genes clusters

We used the list of the 49 PEG and 35 MEG found in *A. lyrata* (Klosinska et al., 2016). We filtered the genes with conserved coordinates in the reference genome of *A. lyrata* (Hu et al., 2011) V1.0.23 to finally study 27 MEG and 46 PEG. We distinguished the two types of imprinted genes in our study. We researched the best hit of these genes in the reference genome of *A. thaliana* by blast V2.10 (Altschul et al., 1990; see below for detailed protocol).

To discriminate the signature of selection related to endosperm, considered under parental conflict, to the signature associated with other possible selection process, we distinguished the imprinted genes upregulated specifically in the endosperm ($14PEG_e/15MEG_e$) with the other imprinted genes not specifically upregulated in the endosperm ($32PEG_o/12MEG_o$) using differential expression analysis. The significance of the number of PEG and MEG upregulated in endosperm between all the PEG and MEG were tested by permutation tests based on 10,000 resamples of the genes with transcriptomic data. For all the following analyses, we compared the 2703 other genes upregulated in the endosperm (*e.g.* endosperm genes) with

these clusters of imprinted genes and a set of 100 randomly chosen ubiquitous genes (*e.g.* control genes).

Based on the sequences on the reference genome of *A. lyrata* (Hu et al., 2011), we tested if the size of the gene, the GC content and the proportion of CDS content on the gene were significantly different across the endosperm genes and the imprinted genes. The significance of the variations across the cluster of genes were tested by the Mann–Whitney–Wilcoxon test, as implemented in R version 4.3.2.

Read mapping and variant calling

We worked on natural accessions from *A. lyrata* of North America represented by 18 samples of 25 individuals from nine allogamous and nine autogamous populations (table S2), using available pool-sequencing data (Willi et al., 2018).

Raw reads from Willi et al (2018) datasets were mapped onto the complete *A. lyrata* reference genome V1.0.23 (Hu et al., 2011) using Bowtie2 v2.4.1 (Langmead and Salzberg, 2012). File formats were then converted to BAM using samtools v1.3.1 (Li et al., 2009) and duplicated reads were removed with the MarkDuplicates program of picard-tools v1.119 (http://broadinstitute.github.io/picard). Biallelic SNPs in these regions were called using the Genome Analysis Toolkit v. 3.8 (GATK, DePristo et al., 2011) with the option GVCF and a quality score threshold of 60 using vcftool v0.1.15 (Danecek et al., 2011). We excluded sites with less than 15 reads aligned. Sites covered by at least 15 reads but containing no SNP were considered as monomorphic. The final number of sites in each sample set is summarised in table S2.

Then, the 0-fold and 4-fold degenerate sites were identified using the script NewAnnotateRef.py (Williamson et al., 2014). Because all mutations at 0-fold degenerate sites alter the sequence of the encoded protein, we assumed that these mutations are non-synonymous (NS). In contrast, we assumed that the mutations at the 4-fold degenerate sites are synonymous (S). For the sake of simplicity, we discarded mutations at 2- or 3-fold degenerate sites.

Estimation of population's polymorphism

We decomposed the estimation of the intra-population's polymorphism into two elementary statistics: the Tajima's D and the ratio of average pairwise heterozygosity between the non-synonymous and synonymous sites π_{NS}/π_S. We estimated the Tajima's D and the π obtained for all positions of each gene using vcftools with --TajimaD and –site-pi options respectively (Danecek et al., 2011). We removed the genes without polymorphic sites.

Moreover, for the position of the *A. lyrata* genome covered but not polymorphic, the π was set to 0. Based on the annotation of each site, we estimated the mean π_{NS} , π_S and π_{NS}/π_S for each gene in each population. We removed the genes with a π_s to 0.

For the allogamous and autogamous populations independently, we compared the value obtained for the endosperm genes with the control genes (effect of parental conflict) and with the imprinted genes (effect of imprinting) by Krustal-Wallis test.

Distribution of the genes across Brassicaceae

We extracted the sequences of the imprinted genes and of the endosperm genes upregulated in the *A. lyrata* reference genome (Hu et al., 2011) with bedtools V2.30 (Quinlan et al., 2010). Then, we researched the best hit of these sequences with an e-value lower than 1^{e-5} on 21 other Brassicaceae by blast V2.10 (Altschul et al., 1990). All the reference genomes were extracted from the Phytozome database (Goodstein et al., 2012; table S3). If more of one hit were found for a gene in one specie, we determined the best hit as the hit with: the higher length alignment, the higher % of identity, then the higher coverage on the gene of *A. lyrata* and the higher reciprocal coverage on the gene of the other species. Then we extracted the sequence of the best hit for each species of each gene.

We compared the number of species with homologous genes for the imprinted and endosperm genes by Krustal-Wallis test.

Divergence of the imprinted genes across Brassicaceae

The rate of nonsynonymous mutations per nonsynonymous site (D_{ns}) and the rate of synonymous mutations per synonymous site (D_s) is assumed to follow the neutral evolutionary process and the ratio, such that D_{ns}/D_s, is therefore approximate to the selective pressure on the protein product of a gene. An increase of the D_{ns}/D_s signifies positive selection at the imprinted genes, while a decrease indicates purifying selection. We estimated the mean divergence and the mean evolutionary rates D_{ns}/D_s for each gene found in at least 3 species between the reference genome of *A. lyrata* and the 21 other Brassicaceae (tables S1 and S4). The genes with D_s of 0 were removed of the analysis of the D_{ns}/D_s.

To test for deviation from neutral expectations, we further compared the fixation index on imprinted genes with endosperm genes based on the ratio of the between-species divergence (D_{ns}/D_s) to the within-species polymorphism (π_{NS}/π_S) on the population of allogamous as used in Tuteja et al (2019).

Topology of the genes across Brassicaceae

For each gene, we computed the gene and all the best hits found in the different species in fasta files to do alignment with Muscle V3.8.155 (Edgar, 2004). Then, we build a phylogenetic tree following HKY85 model for each gene with phyml V3.3 (Guindon et al., 2010). Moreover, we concatenated the alignment of the 1055 genes upregulated in the endosperm found in the 21 others Brassicaceae in a fasta file to build a phylogenetic tree of the endosperm without branch length using phyml V3.3 (Guindon et al., 2010). This tree was used as a proxy of the global topology of the endosperm genes. Then, to determine if the imprinted genes present a different topology than the other endosperm genes, we compared the Robinson-Foulds distance of this tree with the tree of each gene obtained previously using the R package TreeDist. Then we compared the similarity of the imprinted genes and genes upregulated in the endosperm by Krustal-Wallis tests.

Variation of the branch length across Brassicaceae

For each gene, we estimated the branch length for each species with the global topology of the endosperm genes with phyml V3.3 (Guindon et al., 2010). We extracted the branch length with the R package 'ape' version 5.7-1. The species without homologous genes found were not considered. Then we compared the branch length of the imprinted and endosperm genes by Krustal-Wallis tests.

Genes in coevolution

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

To detect coevolution based on polymorphism in population, we used Pearson linear model between the mean π of each pair of imprinted and endosperm genes studied. The p-values were corrected following Bonferroni correction. Only pair of genes with correlations with a p-value lower than 0.05 were considered in the networks of coevolution.

To detect coevolution based on evolutionary rates across species, we compared the length of branches for each species of each imprinted gene with all the other imprinted genes and endosperm genes. The p-values were corrected following Bonferroni correction. Only pair of genes with correlations with a p-value lower than 0.05 were considered in the networks of coevolution.

Finally, to detect coevolution based on topology with other genes, we defined the distribution of the Robinson-Foulds distance expected between two endosperm genes by bootstrap with 10,000 reiterations. Then, we determined the threshold value for high similarity as the value obtained in 99% of the distribution. We removed the gene pairs with at least one gene with a Robinson-Foulds distance of one obtained by comparison with the global endosperm topology. All pairs of genes including at least one imprinted gene with a Robinson-Foulds distance higher than this threshold were considered in the last coevolution network.

Identification of sites in coevolution

We used a model-based approach for detecting coevolving positions of each pair of imprinted genes with a direct coevolution signal within at least one network previously mentioned. To use this method, we concatenated the sequence of pairs of genes for each common species with homologous genes.

Coevolving positions within each pair of genes were predicted using the clustering method implemented in the CoMap package (version 1.5.2; Dutheil & Galtier, 2007). The built-in algorithm was then used to correct for multiple testing and estimate the false discovery rate (FDR). In this study, 1,000 simulations per tested pair were performed, and the groups of sites with an adjusted p-value < 1% using simple correlation were considered coevolving.

For each pair, we extracted the coevolving groups of substitutions and classified them by size (between 2 and 10 coevolving sites).

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

Results

Expression patterns and characteristics of imprinted genes

To disentangle the impact of parental conflict and tissue specificity on selective processes, the genes specifically upregulated in the endosperm of *A. lyrata* were identified based on transcriptomic data obtained for 19,425 genes from 5 tissues (leaf, root, pollen, pistil, endosperm). A total of 2,732 genes were upregulated in the endosperm, including fiveteen MEG and fourteen PEG between the twenty-seven and forty-six found respectively in *A. lyrata* (Klosinska et al., 2016). These proportions of genes upregulated specifically in the endosperm between the MEG and PEG were significantly higher than those predicted by 10,000 random resampling of all the genes studied (p value < 1^{e-04}). For the following analysis, PEG and MEG upregulated in the endosperm will be referred to as PEG_e and MEG_e. The other PEGs and MEGs have been named PEG_o and MEG_o. In these genes, we expected less effect of parental conflict because these genes could be under other selection processes in the other tissues. A set of 100 randomly chosen not imprinted ubiquitous genes will be considered as control for genes not under potential parental conflict.

We compared the genome size, the size of the coding regions and non-coding regions, the CDS content and the GC content of the imprinted genes in *A. lyrata* with the endosperm genes because these statistics could impact the statistics to detect selection used after (table S5). We detected a significant decrease in the gene size associated with a general decrease in the size of the coding region and in the GC content in the MEG_e. We took this observation into account for the discussion because a decrease in GC content is commonly associated with a decrease in the number of segregated synonymous sites (Eyre-Walker & Hurst, 2001).

No similar phylogenetic characteristic across imprinted genes in Brassicaceae

We investigated the characteristics of the imprinted genes in all the Brassicaceae compared to *A. lyrata*. First, we tested if the imprinted genes showed a higher turnover rate across the Brassicaceae by comparing the number of species with homologous genes found for imprinted and endosperm genes (Fig. 1a; table S6). We found a significant decrease of this number in all imprinted genes compared to endosperm genes, except for the MEG_o.

Then, we tested if the imprinted genes present a different phylogenetic topology than the topology based on the endosperm genes found in all the Brassicaceae by comparison of the Robinson-Foulds distance. This distance was lower for the PEG_e genes, suggesting that

these genes followed another topology than the general topology of the endosperm (table S7).

We tested if the imprinted genes were associated with a variation of evolutionary rates by comparison of the length of branch estimated with the general topology of the endosperm (Fig. 1b). We studied only the genes with homologous genes found in at least three species (including *A. lyrata*; table S4). If we considered all the branches, the mean length of branch found in the imprinted genes, except the PEG_o, was significantly lower than in the endosperm genes (Fig. 1c; table S7). However, if we considered only the branch of *A. lyrata*, the mean length of branches found in the imprinted genes was not significantly different than in the endosperm genes (table S7).

Finally, we estimated the variations of the D_{ns}/D_s between *A. lyrata* and the other Brassicaceae species. Before analysis, the genes with homologous genes found in less than three other Brassicaceae and without divergent synonymous sites were removed (tables S1 and S4). For this ratio, we did not detect a significant difference between endosperm and control genes (Fig. 1d; table S6). However, the MEG_e, the PEG_e and the PEG_o genes presented a significant decrease in this ratio compared to endosperm genes (Fig. 1d; table S6).

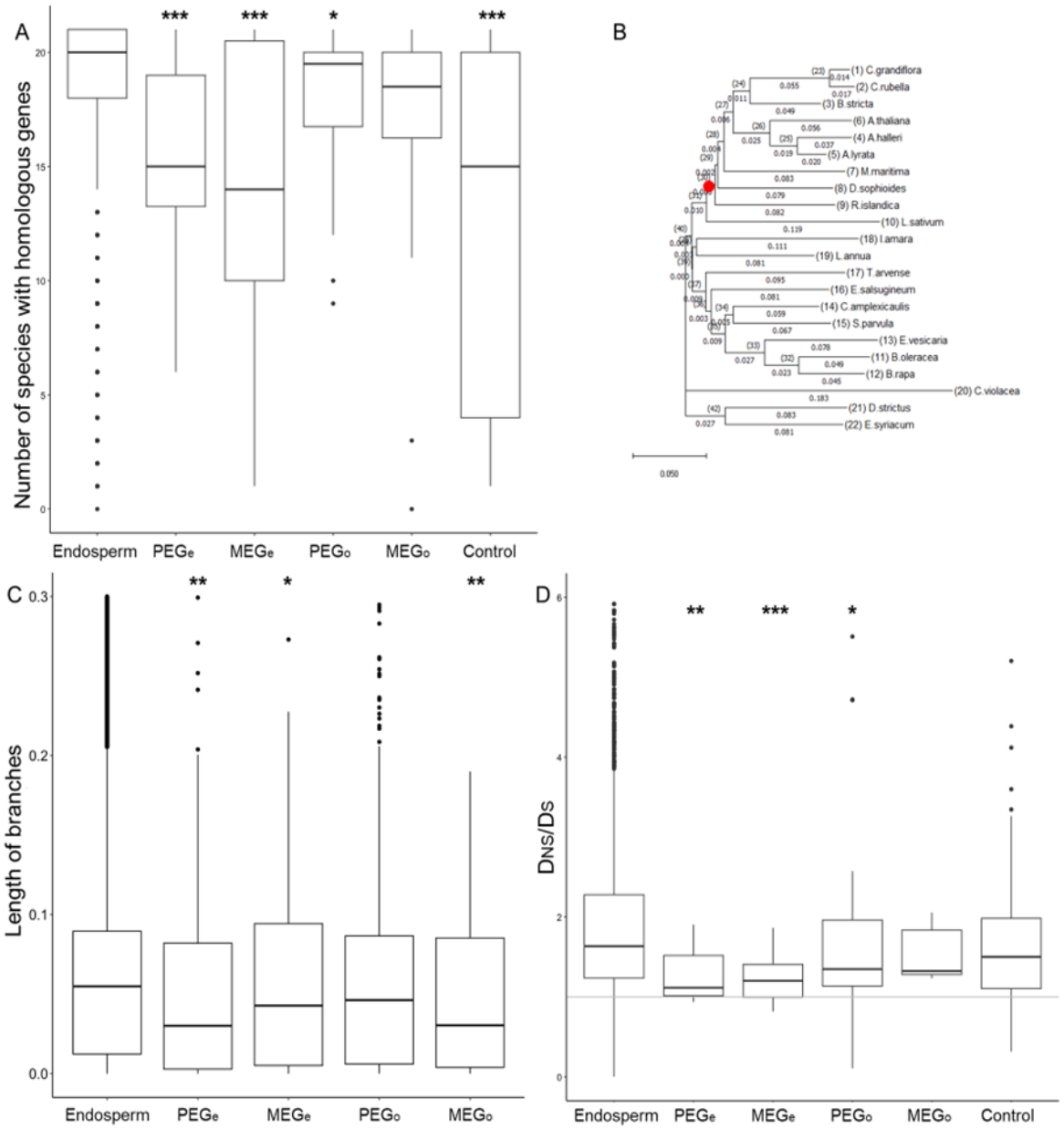

Figure 1: Phylogenetic characteristic of the imprinted genes compared to endosperm genes across the Brassicaceae. *A: boxplot of the variations of the number of species with homologous genes by group of genes. B: phylogenetic tree obtained by HKY85 model based on the 1055 endosperm genes found in all 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 the number of substitutions per site. C: boxplots of the variations of the median length branch by gene based on all the species. D: boxplot of the variations of the D_{NS}/D_S by gene. The light grey lines represent the threshold value for neutrality. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. Only the significant p-values with endosperm genes were specified.*

No common signal of selection was detected on imprinted genes.

The parental conflict on imprinted genes is expected to promote directional selection on these genes in allogamous populations compared to non-imprinted genes. To test this hypothesis, we compared Tajima's D and the π_{NS}/π_S ratio on the imprinted and endosperm genes in the allogamous and autogamous populations.

In the case of directional selection on imprinted genes, we expect a reduction of the Tajima's D compared to endosperm genes, whereas we expect an increase of the Tajima's D in case of balancing selection. The genes without polymorphic sites in all the populations were removed before analysis (table S4). The endosperm genes presented a significant increase of the Tajima's D compared to control genes in the populations of allogamous but not in the autogamous populations (Fig. 2a; table S8). Compared to endosperm genes, the PEG_e genes presented a significant decrease of the Tajima's D in the allogamous populations, whereas we detected a significant increase on the MEG_e in the allogamous populations.

In the case of directional selection on imprinted genes, we expect also a reduction of the π_{NS}/π_S compared with the other endosperm genes. The genes with a π_S to zero in all the populations were removed before analysis (tables S1 and S4). The endosperm genes presented a significant decrease in this ratio compared to control genes in the autogamous population (table S9). Compared to endosperm genes, this ratio increased significantly for the MEG_o and the PEG_e genes in the allogamous and autogamous populations (Fig. 2b; table S8), whereas π_{NS}/π_S increased for the PEG_o genes only in the autogamous populations.

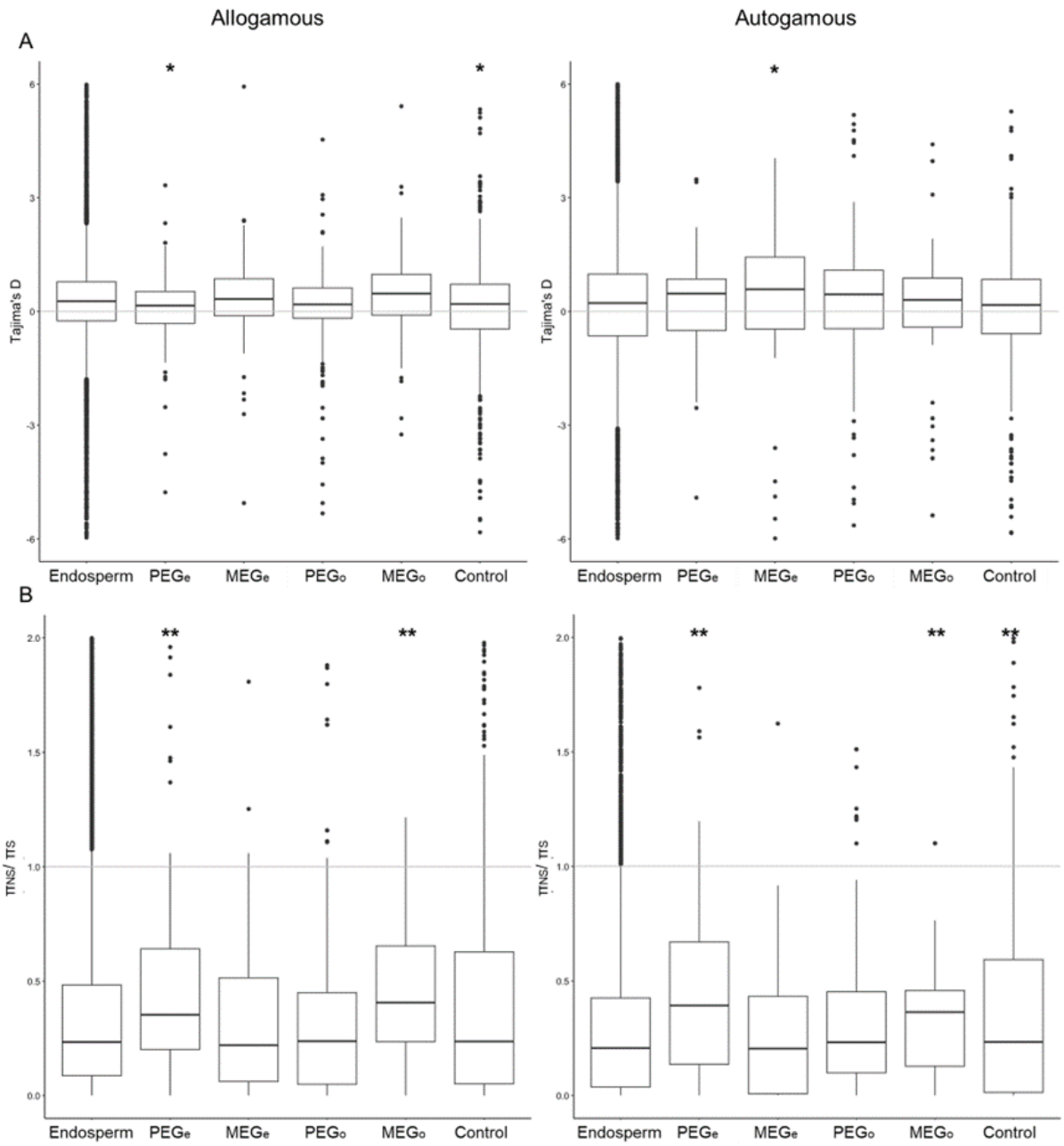

Figure 2: Variation of the polymorphism in the allogamous and autogamous populations in the imprinted genes compared to endosperm and control genes. *A: variation of the median Tajima's D by gene. B: variation of the median π_{NS}/π_S by gene. The grey bars represent the median value obtained for each group of genes. The intervalles represent the 90% of the distribution for each group of genes.The light grey lines represent the threshold value for neutrality. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. Only the significant p-values with endosperm genes were specified.*

We attempted to detect signals of selection by a variation of the fixation index $(D_{ns}/D_s)/(\pi_{NS}/\pi_S)$, as in Tuteja et al. (2019). Exactly, we expected an increase in the fixation

index (FI) in case of positive selection. The genes without polymorphic sites in all the populations were removed before analysis (tables S1 and S4). Compared to control genes, the FI on endosperm genes was significantly higher in autogamous populations, but not on allogamous. Compared to endosperm genes, the FI was significantly lower for the PEG_e and the MEG_o in the allogamous and autogamous populations (Fig. 3; tables S8 and S9).

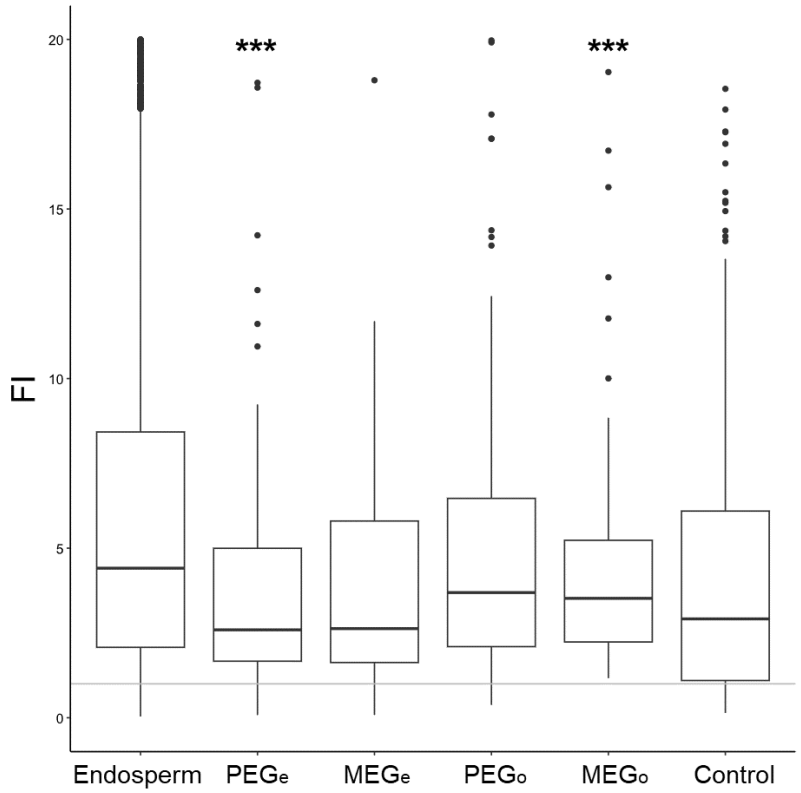

Figure 3: Variation of the divergence and of the fixation index in the imprinted genes compared to endosperm and control genes. *Variation of the median FI by gene based on the allogamous populations. The grey bars represent the median value obtained for each group of genes. The intervals represent 90% of the distribution for each group of genes. The light grey lines represent the threshold value for neutrality. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. Only the significant p-values with endosperm genes were specified.*

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

We supposed that only some imprinted genes in allogamous populations were associated with strong directional selection. Thus, we researched the imprinted genes with extreme decrease of Tajima's D and/or π_{NS}/π_S by comparing the medians obtained in these genes with the median values found in the endosperm genes (table S1). Exactly, we researched genes with lower values that 5% of the distributions in endosperm genes. We found one PEG_o and one PEG_e, respectively the homologous genes *AT5G54070* and *AT2G26070* in *A. thaliana*, with extremely low Tajima's D. These genes encode for a heat stress transcription factor and a protein of the Golgi's membrane involved in the negative regulation of the response to ethylene. We also found three PEG_o, one PEG_e and two MEG_e presented at an extremely low π_{NS}/π_S (table S1), respectively the homologous genes of *AT1G70560*,

AT4G17270, *AT5G26210, AT3G20750, AT1G18100* and *AT4G21080* in *A. thaliana*. They encode respectively for a tryptophan aminotransferase involved in ethylene response, a nuclear transport factor, a protein serine/threonine kinase activator activity and three transcription factors.

No clear trend of coevolution between imprinted genes in Brassicaceae

The last important expectation about imprinting genes is that parental conflict promotes coevolution between genes favouring paternal or maternal fitness in an arms race scenario. To test this assumption, different coevolution signals between PEG, MEG and the other endosperm genes were investigated. More precisely, we assessed whether some pair of genes presented correlations based on 1) the π in the populations, 2) the length of branches or 3) the Robinson-Foulds distance of topologies of the phylogenetic trees.

The correlations of the π between each pair of imprinted and endosperm genes were estimated by Pearson correlation. We found 9 imprinted genes implicated in 101 significant correlations (p-values Bonferroni <=0.05). Between them, 83% of the correlations implicated 5 PEG_o and MEG_o, correlated together and with 76 endosperm genes (Fig. 4a; table S10). No specific function was associated with this network by gene ontology.

The length branch between each pair of imprinted and endosperm genes determined 24 imprinted genes implicated in 324 significant correlations (table S11, p-values Bonferroni <=0.05). Between them, explained 41% of the correlations implicated two PEG_o directly positively correlated, one other PEG_o indirectly correlated by three endosperm genes and 87 other endosperm genes (Fig. 4b; table S11). This network of genes and was associated with the catalytic activity by gene ontology (FDR=2.18^{e-02}; 49/93 genes). Moreover, 37% of the correlations were associated with two MEG_e and one MEG_o associated by indirect correlations with thirteen endosperm genes and 107 other endosperm genes. This network was associated with the endomembrane system (FDR=3.8^{e-03}; 25/123 genes) and the golgi apparatus (FDR=2.77^{e-02}; 15/123 genes) by gene ontology. The other networks found were not associated with a specific function based in gene ontology (Fig. 4b; table S11).

Finally, based on topologies, we researched the pair of genes with a Robinson-Foulds distance higher than the 99% distribution found between all the pairs of endosperm genes (distance higher or equal than 0.979). Moreover, to remove pair of genes with a signal of coevolution explained by the similarity with the global topology, we removed the pair of genes that implicated at least one gene with a Robinson-Foulds distance of 1 with the global topology of the endosperm genes (Fig. 1b). We found only one network compound by 982

significant pairs of genes with similar topologies, including six PEG_e, eigthteen PEG_e, ten MEG_e, five MEG_o and 490 endosperm genes (Fig. 4c; table S12). We observed that eigthteen of them directly related. The other PEG and MEG were indirectly related and required six endosperm genes (Fig. 4c; table S12). This network of genes was associated by gene ontology with intracellular protein transport (FDR=1.76^{e-02}; 20 genes) and the macromolecule localisation (FDR=2.14^{e-02}; 38 genes). In this network, we observed that one MEG_e and one MEG_o, *AL1G54190* and *AL1G46480*, presented a significant number of associations compared to endosperm genes. Whereas 99% of the endosperm genes were associated by less than 255 other genes by distance of topology, these genes were associated with 378 and 260 other genes, respectively (table S12). The first gene is the homologous gene of *AT1G47760* in *A. thaliana* (table S1) that encodes for a DNA-binding transcription factor related to the RNA polymerase II activity. However, we did not find a homologous gene in *A. thaliana* for the second gene (table S1).

With the three analyses cumulated, forty-five PEG/MEG (63%) were directly or indirectly associated (*e.g.* by a non-imprinted endosperm gene), but we did not find a common coevolution signal (*e.g.* common pair of genes) based on π, the length of branches and the topology.

To detect coevolution between specific mutations within the twenty-seven pairs of PEG/MEG directly associated in the networks, we used the clustering method of the pipeline CoMap package (Dutheil & Galtier, 2007). Exactly, we researched groups of compounds of two to ten sites in coevolution and distributed across the two genes of each pair of. PEG/MEG candidates. To avoid false association, we compared the number of groups of each size found in the candidates with the number of groups of sites found with equivalent size for twenty-seven pairs of PEG/MEG never associated in the previous networks randomly chosen as control (Fig. S1a). We found significant increases in the number of groups detected in candidates compared to control for the groups with at least eight sites in coevolution (table S13; Fig. S1a). The groups of coevolving sites composed of less than eight sites were not considered.

For the 4 pairs of PEG/MEG directly associated by π (Fig. 4a), we found ten groups of at least eight correlated sites (table S13; Fig. S1b for an example of one group of nine coevolving sites between the MEG_o *AL2G40710* and the PEG_o *AL6G44530*). For the two PEGs directly associated by length of branches (Fig. 4c), we found two groups of ten correlated sites (table S13; Fig. S1c for an example). Finally, for the twenty-two pairs of PEG/MEG directly associated by topologies (Fig. 4d), we found seventy-four groups of at least eight correlated sites for only twelve pairs of genes (55% of the candidates; table S13;

Fig. S1d for an example of one group of nine correlated sites between the MEG_e *AL1G54190* and the PEG_o *AL1G64840*).

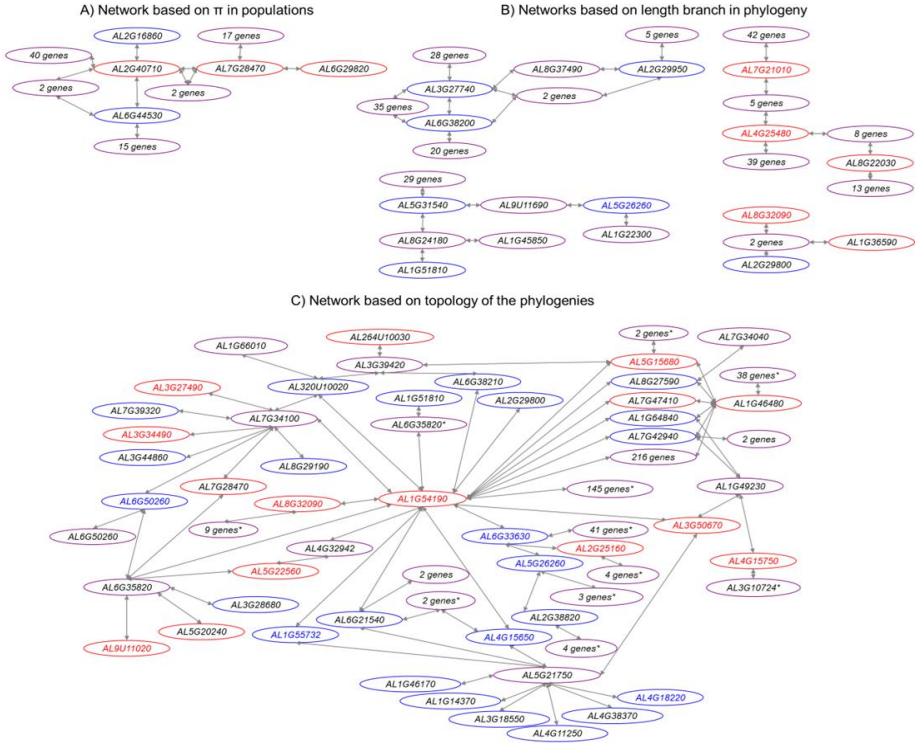

Figure 4: Networks of the imprinted and non-imprinted genes based on A) population polymorphism, B) length of branches and C) phylogenetic topologies. *A) Network based on the Pearson's correlations of π found on genes in the populations. B) Networks based on the Pearson's correlations of the length of branches for each gene estimated with the global topology. C) Networks based on the similarities of topologies on pairs of genes higher than 99% of the similarities found on the pair of non-imprinted endosperm genes and not related to the global topology. Only the networks implicating more than one PEG/MEG are represented. The grey lines represent the relations found. Blue circles written in blue: PEG$_e$. Blue circles written in black: PEG$_o$. Red circles written in red: MEG$_e$. Red circles written in black: MEG$_o$. Purple circles written in black: non-imprinted endosperm genes. *: Non-imprinted endosperm genes with other relations were found with the PEG/MEG genes in the networks but not symbolised to simplify.*

Discussion

Genomic imprinting is a form of epigenetic gene regulation in which alleles of genes are expressed in a parent-of-origin dependent manner (Haig, 1997; Wilkins, 2011). The most frequently invoked selective explanation for the evolution of imprinting is the parental conflict hypothesis (Haig, 2000). In species in which the maternal parent directly provides growing progeny and has offspring by multiple males, the mother's reproductive fitness is greater if the resources are equally distributed among all the offspring, while a father benefits from the concentration of maternal resources to his own offspring. This conflict is expected to lead imprinted genes to directional selection in allogamous populations and to promote the coevolution of imprinted genes having antagonistic effects on offspring growth (Wilkins & Haig, 2001; Mills & Moore, 2004). This coevolution could be explained by direct interactions between the antagonist imprinted genes (Wilkins & Haig, 2001; Mills & Moore, 2004) or indirectly by interactions with other genes implicated in parental conflict (Willi, 2013). Based on population genomic and phylogenetic approaches, we found that imprinted genes upregulated in endosperm exhibited signals of negative selection in *A. lyrata* in allogamous populations, however this signal disappeared in autogamous populations.

Phylogenetic signals of negative selection on imprinted genes.

The parental conflict theory predicts that many genes under imprinting should affect growth (Haig, 2000). Thus, we supposed that, if the imprinted genes are implicated in parental conflict, these genes must be preferentially expressed on endosperm tissue. We compared the expression of the imprinted genes in *A. lyrata* with the genes specifically upregulated in the endosperm. According to the parental conflict hypothesis, the imprinted genes were associated with a significant excess of genes specifically upregulated in the endosperm. However, over half of the imprinted genes were also upregulated in one or more other tissues. We supposed that this differential pattern of expression across imprinted genes could affect selection on these genes, because the genes not only upregulated on endosperm could be under different selection processes, not only related to parental conflict.

An important common point found in the imprinted genes in *C. rubella* and *A. thaliana* is the detection of rapid evolution compared to the rest of the genome (Hatorangan et al., 2016; Tuteja et al., 2019). Consistent with these studies, based on phylogenetic approaches, we found that imprinted genes upregulated in endosperm present a reduced length of branches, translating a fast evolutionary rate. Moreover, the reduction of the D_{NS}/D_S on these genes suggests a process of negative selection. Interestingly, the PEG of other tissus did not exhibit this acceleration of the evolutionary rate and the MEG of other tissus were not

associated with a reduction of D_{NS}/D_S. These observations tend to confirm that the PEG and MEG not specific to endosperm are under other processes of selection.

Contrary to the other imprinted genes, the MEG_e exhibited also a specific decrease in gene size and GC content compared to endosperm genes. Even if different hypotheses were reported to explain the variation of GC content across the genome (Eyre-Walker & Hurst, 2001), it is now widely accepted that the repair bias that favours GC over AT alleles during meiotic recombination is the major driver of the GC heterogeneity (Romiguier & Roux, 2017). If we followed this hypothesis, the MEG_e appeared to be non-recombinant genes. Moreover, the small genes were associated with recent genes in many studies (Wolf and al., 2009; Vishnoi and al., 2010; Grishkevich & Yanai, 2014), suggesting recent origins of the MEG_e. This hypothesis could explain that the imprinted genes were less distributed across Brassicaceae species that others endosperm genes, even if this result could also translate an important turnover rate for these genes across species.

No specific signal of selection on imprinted genes in allogamous populations.

Parental conflict is expected to promote diversification of genes implicated on allocation resources in allogamous populations because the alleles will be differentially selected in the two parents. However, in the case of imprinting, conflict should lead the imprinted genes to directional selection and promote the fixation or the loss of the advantageous or disadvantageous allele for the parent that expresses the genes (Wilkins & Haig, 2001; Mills & Moore, 2004). According to our expectation, we found an increase of the Tajima's D on endosperm genes compared to control genes, suggesting a process of balancing selection. However, consistent with the study on *C. rubella* (Hatorangan et al., 2016*)*, we did not find a homogenous signal of selection on all PEG/MEG based on population genomics. In detail, we did not find a specific signal of selection on MEG_e and on PEG_o, whereas the PEG_e and MEG_o presented an increase of the π_{NS}/π_S and a decrease of the FI, translating a signal of recent relaxation of selection. However, we need to be aware that MEG_e exhibited a decrease of gene size and GC content, and a decrease in GC content is expected to promote the decrease in the number of synonymous mutations (Eyre-Walker & Hurst, 2001). If the effect of gene size in selection is poorly documented, an analysis of the human genome also observed fewer synonymous mutations in small genes (Gordo & Charlesworth, B., 2000). Consistent with these assumptions, we found a decrease in the π_S of MEG_e compared to non-imprinted genes. This decrease could have overestimated the π_{NS}/π_S, and finally underestimated the fixation index.

The PEG_e present also singular decrease of Tajima's D. This observation translated opposite senses of selection than the π_{NS}/π_S because it is associated with directional selection. In *Mycobacterium tuberculosi*s, a subset of resistance genes also presented an increase of π_{NS}/π_S but a decrease of Tajima's D (Hough and al., 2013). The authors explained this profile by a relaxation of these genes after a long period of directional selection. This hypothesis of relaxation of selection is interesting in our context because these populations of *A. lyrata* colonised recently North America from Europe (Ross-Ibarra et al., 2008; Otto et al., 2015; Willi et al., 2018). With colonisation, the change in environmental pressures (biotic and abiotic) could relax selection on genes highly selected in ancestral populations.

We found also that some imprinted genes exhibited signal of directional selection in allogamous populations. These genes could be potentially under high parental conflict and deserve more study.

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

The autogamy promotes the relaxation of the selection on imprinted genes.

Parental conflict is expected only in case of multiple mating (Haig, 2000). Thus, we supposed relaxation of selection on imprinted genes in the autogamous population. Consistent with this assumption, signatures of positive selection have been detected at the imprinted *MEDEA* locus in the allogamous plant *Arabidopsis lyrata* (Spillane et al., 2007) but not in the autogamous sister species *A. thaliana* (Miyake et al., 2009). However, positive selection on PEGs was reported on the autogamous species *A. thaliana* and *C. rubella* (Hatorangan et al., 2016; Tuteja et al., 2019). This observation could be explained if, in autogamous populations, the polymorphism on imprinted genes will be like non-imprinted endosperm genes, because the absence of parental conflict should promote directional selection for all genes implicated in resource allocation. These opposite observations raise

the question of the general effect of the mating system, and more specifically of the selfing rate, on the selection of imprinted genes in autogamous species or populations.

In the autogamous populations, as expected, we observed an equivalent Tajima's D, a decrease of the π_{NS}/π_S and an increase of the FI in endosperm genes compared to control gene, suggesting a general directional positive selection on these genes. According to our expectations, we detect singular and homogenous variations of polymorphism in imprinted genes upregulated in endosperm only in autogamous populations. Exactly, we detected an increase of the Tajima's D on MEG_e compared to endosperm genes, whereas the Tajima's D becomes equivalent to endosperm genes for the PEG_e. All these observations suggest a relaxation of selection related to the mating system on imprinted genes.

An important argument to counteract our conclusion could be that the colonisation event of our populations is so recent (Ross-Ibarra et al., 2008; Foxe et al., 2010; Hough et al., 2013; Willi et al., 2018) that variation in selection process is not yet detectable. In addition, the transition to selfing is recent and partial (significant rates of outcrossing) in these populations, potentially limiting the impact of relaxed parental conflict on imprinted genes sequence evolution. However, in non-imprinted endosperm genes, also expected under parental conflict, we observed significant differences between the two mating types, suggesting that the effect of the recent shift to selfing is sufficient to be detected.

Some imprinted genes present a particular signal of coevolution.

Another expected consequence of parental conflict is the coevolution of the reciprocally imprinted genes having antagonistic effects on offspring growth by direct or indirect interactions (Wilkins & Haig, 2001; Mills, & Moore, 2004). To detect coevolution between imprinted and non-imprinted genes, population genomic and phylogenetic methods were used (Dutheil & Galtier, 2007). We detected direct coevolution signals for twenty-seven pairs of PEG/MEG, and we were able to find coevolving groups of sites for eighteen of them. We also found indirect coevolution signals between pairs of PEG/MEG. Together, these imprinted genes are involved in biological functions related to the transport of macromolecules and the endomembrane system. Further research may explore the functional significance of this coevolution signal.

Acknowledgements

This work has been supported by the Czech Science Foundation (GACR 22-20240S granted to CLP). Computational resources were supplied by the project 'e-Infrastruktura CZ' (e-

INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Competing interests

The authors declare no competing interests.

Author contributions

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

Data availability

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

References

1) Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. https://doiOrg/10.1016/S0022-2836(05)80360-2

2) Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25: 25–29. https://doiOrg/10.1038/75556

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

4) Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doiOrg/10.1093/bioinformatics/btr330

5) DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498. https://doiOrg/10.1038/ng.806

6) Dutheil J, Galtier N. 2007. Coevolution within and between Genes. Genome Dyn 3:1-12. https://doiOrg/10.1159/000107599

7) Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797. https://doiOrg/10.1093/nar/gkh340

8) Eyre-Walker A, Hurst LD. 2001. The evolution of isochores. Nat Rev Genet 2: 549–555. https://doiOrg/10.1038/35080577

9) Foxe JP, Stift M, Tedder A, Haudry A, Wright SI, Mable BK. 2010. Reconstructing Origins of Loss of Self-Incompatibility and Selfing in North American Arabidopsis Lyrata: A Population Genetic Context. Evolution 64: 3495–3510. https://doiOrg/10.1111/j.1558-5646.2010.01094X

10) Gehring M, Missirian V, Henikoff S. 2011. Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. PLOS ONE 6: e23687. https://doiOrg/10.1371/journalPone.0023687

11) Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40: D1178-1186. https://doiOrg/10.1093/nar/gkr944

12) Gordo I, Charlesworth B. 2000. On the speed of Muller's ratchet. Genetics 156: 2137–2140. https://doiOrg/10.1093/genetics/156.4.2137

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

14) Guindon S, Dufayard JF, 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://doiOrg/10.1093/sysbio/syq010

15) 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://doiOrg/10.1098/rspb.1997.0230

16) Haig D. 2000. The Kinship Theory of Genomic Imprinting. Annual Review of Ecology and Systematics 31: 9–32. https://doiOrg/10.1146/annurevEcolsys.31.1.9

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

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

19) Hatorangan MR, Laenen B, Steige KA, Slotte T, Köhler C. 2016. Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. The Plant Cell 28: 1815–1827. https://doiOrg/10.1105/tpc.16.00304

20) Hough J, Immler S, Barrett SCH, Otto SP. 2013. Evolutionarily stable sex ratios and mutation load. Evolution 67: 1915–1925. https://doiOrg/10.1111/evo.12066

21) Hsieh TF, Shin J, Uzawa R, Silva P, Cohen,S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D et al. 2011. Regulation of imprinted gene expression in Arabidopsis endosperm. Proceedings of the National Academy of Sciences 108: 1755–1762. https://doiOrg/10.1073/pnas.1019273108

22) Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H et al. The Arabidopsis lyrata genome sequence and the

basis of rapid genome size change. Nat Genet 43: 476–481. https://doiOrg/10.1038/ng.807

23) İltaş Ö, Čertner M, Lafon-Placette C. 2024. Natural variation in sexual traits and gene expression between selfing and outcrossing Arabidopsis lyrata suggests sexual selection at work. Plant and Cell Physiology, in press.

24) Klosinska M, Picard CL, Gehring M. 2016. Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. Nature Plants 2: 1–8. https://doiOrg/10.1038/nplants.2016.145

25) Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359. https://doiOrg/10.1038/nmeth.1923

26) 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://doiOrg/10.1093/bioinformatics/btp352

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

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

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

30) Otto SP, Scott MF, Immler S. 2015. Evolution of haploid selection in predominantly diploid organisms. Proceedings of the National Academy of Sciences 112. 15952–15957. https://doiOrg/10.1073/pnas.1512004112

31) Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. 2014. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. eLife 3: e03198. https://doiOrg/10.7554/eLife.03198

32) Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doiOrg/10.1093/bioinformatics/btq033

33) Raissig MT, Baroux C, Grossniklaus U. 2011. Regulation and Flexibility of Genomic Imprinting during Seed Development. The Plant Cell 23: 16–26. https://doiOrg/10.1105/tpc.110.081018

34) Romiguier J, Roux C. 2017. Analytical Biases Associated with GC-Content in Molecular Evolution. Frontiers in Genetics 8.

35) Ross-Ibarra J, Wright S.I, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008. Patterns of Polymorphism and Demographic History in Natural Populations of Arabidopsis lyrata. PLOS ONE 3: e2411. https://doiOrg/10.1371/journalPone.0002411

36) Scott RJ, Spielman M, Bailey J, Dickinson HG. 1998. Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125: 3329–3341. https://doiOrg/10.1242/dev.125.17.3329

37) Spencer HG, Clark, AG. 2014. Non-conflict theories for the evolution of genomic imprinting. Heredity 113. 112–118. https://doiOrg/10.1038/hdy.2013.129

38) Spillane C, Schmid KJ, Laoueillé-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C, Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at the imprinted MEDEA locus in plants. Nature 448: 349–352. https://doiOrg/10.1038/nature05984

39) Tuteja R, McKeown PC, Ryan P, Morgan CC, Donoghue MTA, Downing T, O'Connell MJ, Spillane C. 2019. Paternally Expressed Imprinted Genes under Positive Darwinian Selection in Arabidopsis thaliana. Molecular Biology and Evolution 36: 1239–1253. https://doiOrg/10.1093/molbev/msz063

40) Vishnoi A, Kryazhimskiy S, Bazykin GA, Hannenhalli S, Plotkin JB. 2010. Young proteins experience more variable selection pressures than old proteins. Genome Res 20: 1574–1581. https://doiOrg/10.1101/gr.109595.110

41) 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://doiOrg/10.1086/670196

42) 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://doiOrg/10.1093/molbev/msy003

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

44) Wilkins JF. 2011. Genomic imprinting and conflict-induced decanalization. Evolution 65: 537–553. https://doiOrg/10.1111/j.1558-5646.2010.01147X

45) Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, 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://doiOrg/10.1371/journalPgen.1004622

46) Wolf Y.I, Novichkov PS, Karev GP, Koonin EV, Lipman DJ. 2009. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. Proceedings of the National Academy of Sciences 106: 7273–7280. https://doiOrg/10.1073/pnas.0901808106

47) Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, 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://doiOrg/10.1371/journalPgen.1002126

Main Figures legends

Figure 1: Phylogenetic characteristic of the imprinted genes compared to endosperm genes across the Brassicaceae. *A: boxplot of the variations of the number of species with homologous genes by group of genes. B: phylogenetic tree obtained by HKY85 model*[41] *based on the 1055 endosperm genes found in all species after 1000 bootstrap. The percentage of trees in which the associated haplotypes clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. C: boxplots of the variations of the median length branch by gene based on all the species. D: boxplot of the variations of the D_{NS}/D_S by gene. The light grey lines represent the threshold value for neutrality. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. Only the significant p-values with endosperm genes were specified.*

Figure 2: Variation of the polymorphism in the allogamous and autogamous populations in the imprinted genes compared to endosperm and control genes. *A: variation of the median Tajima's D by gene. B: variation of the median π_{NS}/π_S by gene. The grey bars represent the median value obtained for each group of genes. The intervalles represent the 90% of the distribution for each group of genes.The light grey lines represent the threshold value for neutrality. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. Only the significant p-values with endosperm genes were specified.*

Figure 3: Variation of the divergence and of the fixation index in the imprinted genes compared to endosperm and control genes. *Variation of the median FI by gene based on the allogamous populations. The grey bars represent the median value obtained for each group of genes. The intervals represent 90% of the distribution for each group of genes. The light grey lines represent the threshold value for neutrality. ***: p-values U-test < 0.001. **: p-values U-test < 0.01. *: p-values U-test < 0.05. Only the significant p-values with endosperm genes were specified.*

Figure 4: Networks of the imprinted and non-imprinted genes based on A) population polymorphism, B) length of branches and C) phylogenetic topologies. *A) Network based on the Pearson's correlations of π found on genes in the populations. B) Networks based on the Pearson's correlations of the length of branches for each gene estimated with the global topology. C) Networks based on the similarities of topologies on pairs of genes higher than 99% of the similarities found on the pair of non-imprinted endosperm genes and not related to the global topology. Only the networks implicating more than one PEG/MEG are represented. The grey lines represent the relations found. Blue circles written in blue:*

*PEG_e. Blue circles written in black: PEG_o. Red circles written in red: MEG_e. Red circles written in black: MEG_o. Purple circles written in black: non-imprinted endosperm genes. *: Non-imprinted endosperm genes with other relations were found with the PEG/MEG genes in the networks but not symbolised to simplify.*

Supporting Information

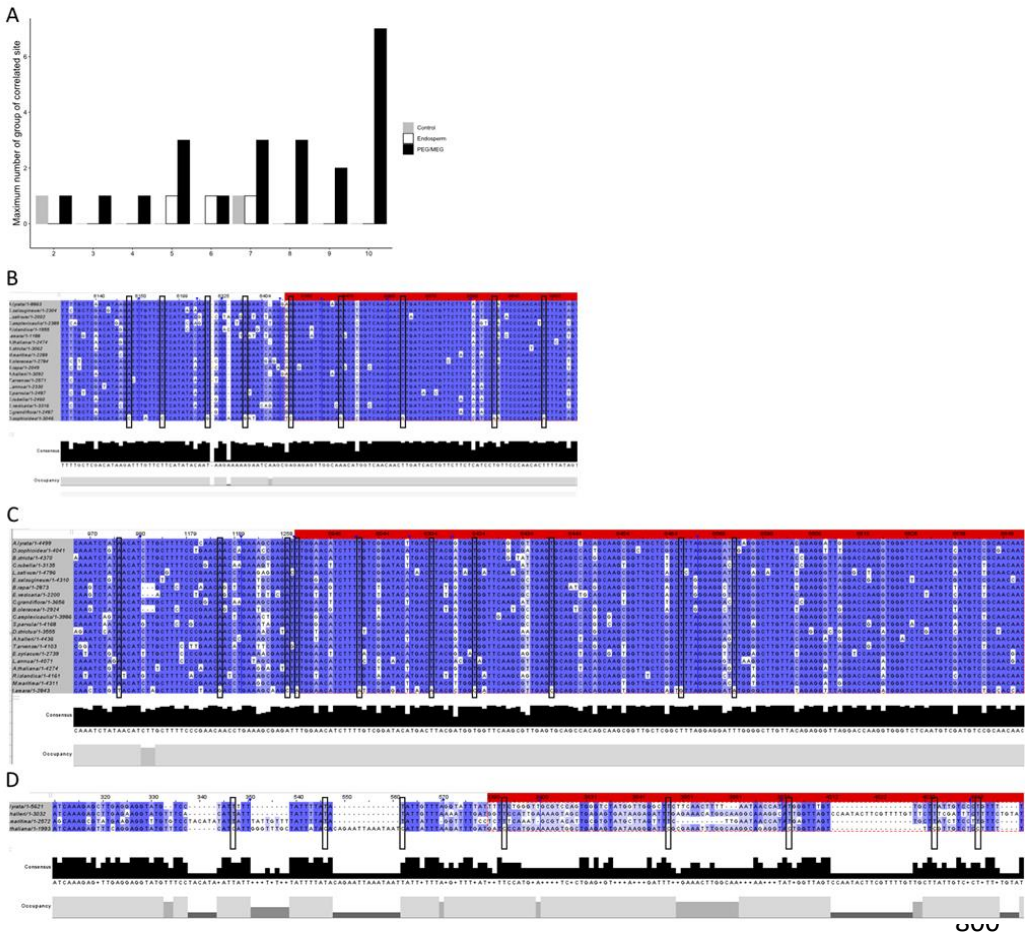

Figure S1: Correlated sites in imprinted genes directly associated in the networks in 22 Brassicaceaes. *A) Maximum number of groups with at least 2 correlated sites in each pair of genes. We distinguished the groups of correlated sites by the number of sites found in each group (between two and ten) and by type of pair of genes. We compared the real pair of imprinted genes directly associated in the networks (black bars) of the pair of imprinted genes never associated (control for false association; grey bars) and of the pair of non-imprinted endosperm genes directly associated in the networks (white bars). B) Example of alignment for one group of correlated sites for a pair of imprinted genes associated by polymorphism in populations. C) Example of alignment for one group of correlated sites for a pair of imprinted genes associated by length of branches. D) Example of alignment for one group of correlated sites for a pair of imprinted genes associated by topologies. The identity on each position is represented by a blue gradient. The arrows represent the hidden part of the alignment. The region on alignment for first and the second genes is delimited by the red scare (top). The correlated positions are represented by black scares. The consensus sequences and the occupancy are represented at the bottom.*