Paternally expressed imprinted genes under positive Darwinian selection in *Arabidopsis thaliana*

Reetu Tuteja[1†], Peter C. McKeown[1,*], Pat Ryan[1], Claire C. Morgan[2=], Mark T.A. Donoghue[1,+], Tim Downing[2], Mary J. O'Connell[3,4*], Charles Spillane[1,*]

[1]Genetics & Biotechnology Lab, Plant & AgriBiosciences Research Centre (PABC), School of Natural Sciences, Ryan Institute, National University of Ireland Galway, University Road, Galway H91 REW4, Ireland.

[2]School of Biotechnology, Faculty of Biological Sciences, Dublin City University, Ireland.

[3]Computational and Molecular Evolutionary Biology Research Group, School of Biology, Faculty of Biological Sciences, The University of Leeds, LS2 9JT, UK.

[4]Computational and Molecular Evolutionary Biology Group, School of Life Sciences, University of Nottingham, NG7 2RD, UK

[†]current address: Center for Genomics and Systems Biology, New York University, New York, NY, USA.

[=]current address: Division of Diabetes, Endocrinology and Metabolism, Imperial College London, UK.

[+]current address: Memorial Sloan Kettering Cancer Center, New York, NY, USA.

*to whom correspondence should be addressed:

*Professor Charles Spillane, Genetics & Biotechnology Lab, Plant & AgriBiosciences Research Centre (PABC), Ryan Institute, National University of Ireland Galway, University Road, Galway H91 REW4, Ireland.
*Dr. Peter C. McKeown, Genetics & Biotechnology Lab, Plant & AgriBiosciences Research Centre (PABC), Ryan Institute, National University of Ireland Galway, University Road, Galway H91 REW4, Ireland.
*Dr. Mary J. O'Connell, Computational and Molecular Evolutionary Biology Group, School of Life Sciences, University of Nottingham, NG7 2RD, UK.

charles.spillane@nuigalway.ie
peter.mckeown@nuigalway.ie
mary.o'connell@nottingham.ac.uk

Keywords: Genomic imprinting; genomic conflict; positive Darwinian selection; endosperm; plant evolution

Abstract

Genomic imprinting is an epigenetic phenomenon where autosomal genes display uniparental expression depending on whether they are maternally or paternally inherited. Genomic imprinting can arise from parental conflicts over resource allocation to the offspring, which could drive imprinted loci to evolve by positive selection. We investigate whether positive selection is associated with genomic imprinting in the inbreeding species *Arabidopsis thaliana*. Our analysis of 140 genes regulated by genomic imprinting in the *A. thaliana* seed endosperm demonstrates they are evolving more rapidly than expected. To investigate whether positive selection drives this evolutionary acceleration, we identified orthologs of each imprinted gene across 34 plant species and elucidated their evolutionary trajectories. Increased positive selection was sought by comparing its incidence among imprinted genes with non-imprinted controls. Strikingly, we find a statistically significant enrichment of imprinted paternally expressed genes (iPEGs) evolving under positive selection, 50.6% of the total, but no such enrichment for positive selection among imprinted maternally expressed genes (iMEGs). This suggests that maternally- and paternally-expressed imprinted genes are subject to different selective pressures. Almost all positively selected amino acids were fixed across 80 sequenced *A. thaliana* accessions, suggestive of selective sweeps in the *A. thaliana* lineage. The imprinted genes under positive selection are involved in processes important for seed development including auxin biosynthesis and epigenetic regulation. Our findings support a genomic imprinting model for plants where positive selection can affect paternally-expressed genes due to continued conflict with maternal sporophyte tissues, even when parental conflict is reduced in predominantly inbreeding species.

Introduction

Rapid evolution under Positive Selection (PS) is a feature of many reproductive proteins in both plants and animals, occurring either as a result of adaptive radiation or of sexual conflict within and between genomes (Clark, et al. 2006). For example, tests of selective pressure have shown that genes expressed in the highly reduced male gametophyte of flowering plants (the pollen grain) display elevated PS (Arunkumar, et al. 2013; Gossmann, et al. 2014). These increased levels of PS are observed in genes expressed in the pollen tube but not the sperm cell, and are interpreted to be a consequence of conflict driven by competition between pollen grains for access to ovules (Bernasconi, et al. 2004). Conflict is also expected to occur at loci regulated by genomic imprinting, in which genes are monoallelically expressed under epigenetic regulation in a parent-of-origin specific manner, in violation of the Mendelian rules of genetic inheritance (Haig 1997; Wilkins 2011). Indeed, genomic imprinting is widely considered to have evolved due to conflict between parentally-derived genomes over resource allocation to developing offspring which lead to genes evolving different optimal expression levels depending upon whether they are maternally- or paternally-derived (Willson and Burley 1983; Wilkins and Haig 2003b; Haig 2004). Imprinting has been reported from both mammals and flowering plants, in which it principally occurs in the endosperm (Gehring and Satyaki 2017), the second product of double fertilization which provides maternally-derived resources to the developing embryo in the seed (Walbot and Evans 2003). Imprinting leads to the occurrence of imprinted maternally expressed genes (iMEGS) and imprinted paternally expressed genes (iPEGS) (Haig and Westoby 1991; Garnier, et al. 2008; Köhler, et al. 2012). Kin conflict between iPEGs and iMEGs in plants is expected to arise from differences in the optimal level of offspring resource allocation, and resulting offspring size, between the maternal and paternal genomes as selection on the maternal genome favours equal provision to all offspring (and iMEGs near-equal provision; see (Trivers 1974)) while the paternal genome promotes growth of its own offspring alone (Haig 2000; Costa, et al. 2012; Haig 2013; Willi 2013).

Such conflict can have different consequences at the molecular level, including conflict relating to expression level and rapid evolution of nucleotide sequence (or epigenetic

signatures) associated with gene expression (Haig, et al. 2014). At the level of the coding sequence, one prediction is that conflict can lead to positive selection on pairs of reciprocally imprinted genes expressed from the maternally and paternally inherited genomes, each having antagonistic effects on offspring growth (Wilkins and Haig 2001; Mills and Moore 2004). We illustrate this occurring inside the endosperm of the seed (yellow) in fig. 1A, within which iMEGs and iPEGs mutually interact. Some support for this particular form of parental conflict has been found in mammals, for example at the *Igf-2* and *callipyge* loci (Georges, et al. 2003; Reik, et al. 2003; Crespi and Semeniuk 2004). Signatures of positive selection have also been detected at the imprinted *MEDEA* locus in the flowering plant *Arabidopsis lyrata* (Spillane, et al. 2007; Miyake, et al. 2009) which may support the hypothesis that imprinting can cause positive selection on coding sequences of the loci concerned. On the other hand, conflict can have other molecular effects, including selection for stable equilibria of iMEG and iPEG expression levels (Haig 2014), and co-evolutionary scenarios between iMEGs and cytoplasmic factors (Wolf and Hager 2006), as shown in fig. 1B. It has also been suggested that conflict could occur between iPEGs and the tissues of the maternal sporophyte (Willi 2013): the genes of the seed coat (SC) are also maternally-derived and could therefore act in a manner antagonistic to iPEGs – this scenario of 'indirect conflict' between the genes of the maternal seed coat (which we denote scMEGs) and iPEGs in the endosperm is shown in fig. 1C. It has been alternatively suggested that imprinting in plants could be related to the biology of gene expression in triploid endosperm, for example as a dosage control mechanism, although a recent study of gene expression in triploid embryos did not support this (Fort, et al. 2017).

Genomic imprinting also occurs in the model plant, *Arabidopsis thaliana* (L.) Heynh, which is the sister species to *A. lyrata,* at *MEDEA* and several hundred other loci (Gehring, et al. 2011; Hsieh, et al. 2011; McKeown, et al. 2011; Wolff, et al. 2011). Furthermore, a subset of imprinted genes which are expressed early in *A. thaliana* seed development (four days after pollination) display accelerated evolutionary rates compared to non-imprinted genes (Wolff, et al. 2011) as measured by D_N/D_S. The rate of nonsynonymous mutations per nonsynonymous site (D_N) 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_N/D_S (also denoted ω), is therefore approximate to the selective pressure on the protein product of a gene. A value of $\omega>1$ signifies positive selection (PS) at a site, $\omega\approx1$ implies neutral evolution, while $\omega<1$ indicates purifying selection. It should be noted that positive selection typically only acts at a subset of amino acid sites while other sites are typically still under purifying selection, so ω is still generally <1 at the level of the whole gene even when PS has occurred. Hence, comparisons between sets of candidate genes and relevant control sets are needed to identify elevated levels of ω. Enrichment for sites with $\omega>1$ in the dataset of Wolff et al. (Wolff, et al. 2011) when compared with controls in this way was therefore interpreted as a possible signature for conflict-driven selection within plant imprinted genes.

Evidence of elevated rates of adaptive substitution have also been reported for imprinted genes of the outcrossing Brassicaceae species, *Capsella rubella* (Hatorangan, et al. 2016). This suggests that increased PS could be a general phenomenon for imprinted genes, supporting models of the parental conflict theory in which conflict leads to rapid evolution of coding sequences. However, it is important to note that elevated D_N/D_S values can be caused by other factors such as variable effective population size, N_e (Kryazhimskiy and Plotkin 2008; Jensen and Bachtrog 2011) and selection on silent sites (Chamary, et al. 2006). It is also unclear whether potential PS in *A. thaliana* or *C. rubella* is acting equally on iMEGs or iPEGs as would be consistent with models of parental conflict involving direct interactions between the proteins which they encode (fig. 1A): iMEGs and iPEGs both showed higher D_N/D_S in the study of (Wolff, et al. 2011), although in *C. rubella* increased accumulation of nearly neutral non-synonymous variants was restricted to iPEGs (Hatorangan, et al. 2016). Nor has it been shown whether past positive selection has led to fixation within current plant populations, as would be expected if the selection acting on amino acids is functionally significant for protein function.

To determine whether genomic imprinting in the seed endosperm is associated with positive selection in plant genomes, we analyzed the selective pressures acting on a comprehensive group of all confirmed imprinted genes of *A. thaliana* (Gehring, et al. 2011; Hsieh, et al. 2011; McKeown, et al. 2011; Wolff, et al. 2011). Specifically, we

addressed the following questions: 1) What selective pressures are imprinted genes evolving under in *A. thaliana*? 2) If imprinted genes are evolving under positive selection, does this lead to overall positive selection in iMEGs and/or iPEGs being elevated compared to similar sets of biallelically expressed genes? And 3) Is there evidence for fixation of positively selected sites in imprinted genes across sequenced *A. thaliana* accessions? Our findings in relation to these questions extend our understanding of the evolutionary drivers of genomic imprinting and the consequences of parental conflict during reproduction.

Results

Imprinted *Arabidopsis thaliana* genes are rapidly evolving

Genomic imprinting has been predicted to evolve due to parental conflicts over provision of maternal resources to offspring, which has been hypothesised to lead to positive selection at loci involved in this conflict. The model eudicot *Arabidopsis thaliana* has been reported to display genomic imprinting on at least 436 genes in its seed endosperm (Gehring et al., 2011; Hsieh et al., 2011; McKeown et al., 2011; Wolff et al., 2011), with growing consensus over a core set which appear to be stably imprinted in many accessions (Gehring and Satyaki 2017; Schon and Nodine 2017; Wyder, et al. 2017). The identification of genes subject to monoallelic expression in the seed endosperm can be confounded by parent-of-origin specific expression patterns that can also arise during early seed development from gametophytic deposition of mRNA in the fertilised egg cell (zygote) or fertilised central cell (endosperm), or from maternal-expression from genes expressed in the sporophytic seed coat, which may be present as contaminants during RNA-seq analyses. To determine the selective pressures acting on imprinted genes, while avoiding these confounding scenarios, we focused our analyses on those genes with strong evidence for uniparental expression in seeds due to imprinting. We classified these as genes identified from RNA-seq-based studies (Gehring, et al. 2011; Hsieh, et al. 2011; Wolff, et al. 2011) which are expressed from the paternal genome (iPEGs), and which therefore cannot be due to contamination from maternal tissues; and those iMEGs for which experimental validation of monoallelic expression and/or epigenetic regulation in the endosperm has been performed *in planta* (Vielle-Calzada, et al. 1999; Kinoshita, et al. 2004; Köhler, et al. 2005; Tiwari, et al. 2008; Gehring, et al. 2009; Hsieh, et al. 2011; McKeown, et al. 2011; Shirzadi, et al. 2011; Wolff, et al. 2011). This produced a set of 140 high-confidence imprinted genes (supplementary table S1A,1B) of which 63 were iPEGs and 77 were iMEGs. By comparing the *A. thaliana* and *A. lyrata* orthologs, we determined that both iPEGs and iMEGs within the 140 imprinted genes had mean values of ω significantly higher than that of the background representing all other remaining *A. thaliana* genes (table 1; U-test: iPEGs p=9.9e-07, iMEGs p=1.9e-06). This provides large-scale empirical evidence that rapid evolution previously observed in

imprinted genes detected in seed offspring at 4 days after pollination from one set of reciprocal crosses (Wolff, et al. 2011) applies more generally to the imprinted genes of *A. thaliana*.

Imprinted genes are evolving under positive selection in *A. thaliana*

PS can be detected at the population genomic level by assessing allele frequency and coalescence time as variation subject to PS is expected to go to fixation (Nielsen 2005; Sabeti, et al. 2006). Genes can display elevated ω for a range of reasons other than PS, however, such as reduced functional constraint or pseudogenization. To test whether the increase in ω observed across the imprinted iMEGs and iPEGs was due to positive selection, we analyzed the evolutionary rates of iMEGs and iPEGs in the context of clusters of orthologous genes from across the plant kingdom. This analysis was conducted using an in-house plant database containing ortholog clusters from 34 sequenced plant species, either Embryophyte or Chlorophyte (supplementary fig. S1). To further ensure the robustness of our analysis, we only considered clusters for which orthologous genes could be identified from at least six species, in addition to *A. thaliana* (see Methods), following recommended best practice for PAML analyses derived from simulation studies (Anisimova, et al. 2001). Applying this filter, suitable clusters for PAML (codeML) analyses were obtained for 64 of the 140 imprinted genes (30 iMEGs and 34 iPEGs; fig. 2; supplementary table S1B). Sequence alignment quality is also critical for correct sequence analysis (Markova-Raina and Petrov 2011) so all alignments were also assessed using the norMD score as a proxy for alignment quality (Thompson, et al. 2001) – see Methods for details. Two genes (iPEG AT4G11400, iMEG AT5G53870) that had poor sequence alignment quality (norMD score <0.6) were excluded from further analyses.

Applying standard codeML models to the remaining 62 imprinted genes, we identified 30 that are evolving under PS (table 2; fig. 3A; supplementary table S1). For 6 of the 30 positively selected imprinted genes, the PS was specific to the *A. thaliana* lineage (i.e. lineage-specific PS; supplementary table S1A), while for 16 imprinted genes positive selection was detected at individual codons in cross-lineage comparisons (i.e. site-specific PS, supplementary table S1A). Eight imprinted genes displayed both lineage-

specific and site-specific PS (fig. 3A). To ensure that these results have not been biased by any of the assumptions inherent in PAML, we also performed a HyPhy analysis (Pond and Muse 2005) on these 62 genes, using a combination of FEL (Fixed Effects Likelihood), SLAC (Single-Nucleotide Ancestor Counting), and MEME (Mixed Effects Model of Evolution) packages, as described in the Methods. From these analyses, we determined that PS is also predicted to be occurring on all 30 genes identified by PAML (supplementary table S2). HyPhy and codeml-based models such as PAML differ fundamentally in how they estimate site-specific rates: PAML models use random effects likelihood while HyPhy models use fixed-effects likelihood, hence the congruence between the results of the two approaches provides strong confirmation of the robustness of the PS signature at the 30 imprinted loci.

Recently, a methodology has been published for directly estimating possible confounding of imprinting gene analysis by contamination with maternal tissues (Schon and Nodine 2017) . Two of the datasets, of Gehring et al. (2011) and Hsieh et al. (2011), were analysed by Schon and Nodine who suggested that 20 iMEGs from these studies used in our analysis should be considered 'low-confidence' (although variation in gene expression patterns under different growth conditions could itself confound these conclusions). The RNA-seq dataset of Wolff et al. (2011) was not analysed by the Schon and Nodine (2017), so we performed the tissue-enrichment test of Schon and Nodine on the datasets used by Wolff et al. (2011) to determine expression pattern (Belmonte et al., 2012). We conclude that these datasets do not suffer from significant levels of cross-tissue contamination (supplementary fig. S2): only the suspensor showing any potential contamination from non-suspensor specific transcripts while none of the endosperm datasets used to identify imprinted genes showed any enrichment for other tissues, including the maternal seed coat. We conclude that the remaining 57/77 iMEGs used in our PAML and HyPhy analyses are 'high-confidence' imprinted genes, while a further 20 may be due to the presence of maternally-derived transcripts (supplementary table S3). These include four genes which are under positive selection according to both codeML and HyPhy, 10 others which showed no evidence for PS and 6 which were not tested due to lack of sufficient orthology clusters. We conclude that positive selection acts upon 19 iPEGs and 11

iMEGs, and that all of the iPEGs and at least 7 of the iMEGs are high-confidence imprinted genes. Taken together, these results indicate that positive selection acts on protein-coding genes regulated by genomic imprinting in the seeds of *A. thaliana*.

Imprinted genes are preferentially affected by positive selection

The large number of imprinted genes subject to positive selection suggested that genes epigenetically regulated by genomic imprinting could be under stronger positive selection than biallelically-expressed genes. To test this hypothesis, we compared the extent of positive selection in imprinted genes to that observed in randomly sampled gene sets from across the whole genome (supplementary table S4A). Genomic imprinting in plants mainly occurs in the seed endosperm, which can be subject to different selective pressures related to its triploid genome dosage independent of imprinting (Baroux, et al. 2002). Hence, we also conducted analysis of positive selection for random samples of known endosperm-specific *A. thaliana* genes (Belmonte, et al. 2013) (supplementary table S4B). For iPEGs, the odds ratio score for lineage-specific positive selection indicated 3.3- and 2.6- fold enrichment in positive selection in imprinted genes compared to whole-genome and endosperm controls, respectively. These ratios equate to a significant enrichment of lineage-specific positive selection in iPEGs when compared with either the genome-wide or endosperm-specific controls (Fisher's test, p=0.014 and p=0.041 respectively; Fig. 3B). Strikingly, no enrichment was found for iMEGs in either lineage-specific (p=0.531 vs. genome-wide controls, p=0.688 vs. endosperm genes) or site-specific selective pressure variation (p=0.542 vs. genome-wide controls, p=0.764 vs. endosperm genes) (Fig. 3C), whether lower-confidence iMEGs were included or not. To determine if the bias in enrichment of position selection in iPEGs as compared to iMEGs is due to statistical threshold effect we identified an additional set of imprinted genes where the significance level following LRT fell just below the cut off p-value of 0.05 (but above 0.10): out of the set of six imprinted genes identified with this relaxed criteria, only one imprinted gene is annotated as an iMEG, while the other five were iPEGs, therefore we can discount any potential bias of this results due to thresholding. We further tested the strength of the difference between the selective pressures acting on iMEGs and iPEGs by performing a chi-squared test directly on the ω-values as

extracted from the branch site models (using likelihood ratio tests values from (Morgan, et al. 2010)). We conclude that iPEGs, but not iMEGs, are subject to higher levels of positive selective pressure, revealing a difference in the evolutionary trajectory of imprinted genes depending on the parental genome from which they are expressed.

Most imprinted genes exhibit fixation of positively selected sites

If the sites determined to be under positive selection in the *A. thaliana* lineage improved plant fitness, then we could expect that these substitutions would be fixed or exist at high frequency within *A. thaliana* populations due to full or partial selective sweeps (Patwa and Wahl 2008). Hence, we tested the percentage conservation of *A. thaliana*-specific amino acid sites under either lineage-specific PS or site-specific PS (supplementary tables S5, S6). For almost all imprinted genes subject to lineage-specific PS, the associated sites showed 100% conservation across the 80 *A. thaliana* accessions for which full sequence data was available (posterior probability >0.95) (supplementary table S7) (Cao, et al. 2011), with no difference observed between iMEGs and iPEGs. Only two imprinted genes (AT1G48910 and AT1G55050) displayed nonsynonymous mutations at the otherwise conserved positively selected position. AT1G48910 encodes YUCCA 10, which is a flavin monooxygenase involved in auxin biosynthesis predicted to have roles in morphogenetic development of pollen grains, while AT1G55050 is a widely-conserved gene of unknown function. If variation at the amino acids subject to positive selection confers phenotypic effects, this requires distinct *A. thaliana* populations with known population histories to test for differing intra-specific selection signatures driven by local environments (Huber, et al. 2014). We consider that positive selective pressures at imprinted loci in the *A. thaliana* lineage has been sufficiently strong, (i.e. with a selective advantage for these alternative amino acids), to cause the fixation of these amino acid variants.

Positive selection on the imprinted *NRPD1a* gene involved in sRNA regulation

We noted that the imprinted genes subject to lineage-specific positive selection included *NRPD1a*, which encodes a component of the RNA Pol IV complex response for transcribing small RNA and, subsequently, transcriptional balance between

maternally and paternally inherited genomes in endosperm (supplementary table S8) (Kanno, et al. 2005; Eamens, et al. 2008; Erdmann, et al. 2017). It has previously been reported that nucleotide substitution rate of the Pol IV polymerase subunit encoded by *NRPD1a* is 20 times higher than that observed in the equivalent subunit of Pol II (Luo and Hall 2007), supporting a scenario whereby the *NRPD1a* gene is under positive selection and suggesting a possible functional relationship between sRNA processing and (imprinted) genes under positive selection. We assessed if positive selection at *NRPD1a* might be due to selection occurring more generally on sRNA-processing genes, perhaps because of their roles in controlling the balance of maternal and paternal gene expression, and not due to the imprinting status of this gene specifically. However, when we analysed the selective pressures acting on 23 non-imprinted genes encoding components of the sRNA processing pathway, none displayed any signature of positive selection (supplementary table S8). We consider that the positive selection acting on *NRPD1a* is associated with its status as an imprinted gene involved in small RNA production and, likely, with subsequent control of gene expression in the endosperm.

iMEGs and iPEGs have similar evolutionary ages

One potential confounding factor in our analysis would be if iMEGs and iPEGs had different evolutionary ages. To address this possibility, we determined the evolutionary ages of the 140 imprinted genes using a phylostratigraphy approach (Domazet-Loso, et al. 2007) (fig. 4). Nine Age Classes (AC) were defined for available plant genome sequences (http://www.phytozome.net/) where AC 0 includes the youngest genes (i.e. those which have evolved since the divergence of *A. thaliana*) and AC 9 the oldest, or most conserved. We then assigned imprinted genes to different age classes using an e-value cutoff of $<10^{-3}$ (supplementary table S9). Notably, no significant difference was observed between the age distributions of iMEGs and iPEGs (Fisher's exact test, p=0.7), suggesting that differences in age are unlikely to explain the differing levels of PS observed in these categories.

Interestingly, 11 of the imprinted *A. thaliana* genes have been shown to have homologs regulated by imprinting in the sister species, *A. lyrata* (table S1A), according

to the analysis of (Klosinska, et al. 2016). These include three iMEGs and eight iPEGs, including three iPEGs which we find to be under PS; these three all belonged to the most conserved age classes (8 or 9; table S9) so may be good candidates for highly conserved imprinting. In contrast, a total of seven imprinted genes did not show any sequence similarity outside Brassicaceae (fig. 4), *i.e.* they were Brassicaceae-specific orphans according to our previous definition (Donoghue, et al. 2011). Of these Brassicaceae-specific imprinted orphan genes, one (AT4G31060) was found in *A. thaliana* only and so represents the most recently-arisen imprinted gene known for this species. The fact that some imprinted genes date from the evolution of the angiosperms may indicate roles for these genes in the accompanying double fertilization event by which the endosperm evolved (Gehring, et al. 2011), although this remains to be tested.

We found that the imprinted gene set as a whole showed enrichment for participation in the *At*-α whole genome duplication (WGD; 52 imprinted genes, Fisher's test, p=0.02), whereas only 21 genes were found to have participated in either the *At*-β or *At*-γ WGD events (Fisher's test, p=0.14) (fig. 4). The *At*-α WGD pre-dated the diversification of core Brassicaceae from *Aethionema* (Franzke, et al. 2011), while *At*-β and *At*-γ are older WGD events predating the emergence of Brassicaceae within the Eurosids (Bowers et al., 2003). These findings are in agreement with the models of Qiu et al., who suggested that many imprinted genes are descended from loci formed by WGD during the evolution of Brassicales (Qiu, et al. 2014). However, there was again no difference in this distribution between iMEGs and iPEGs across different WGD events. In summary, we found no evidence for differing evolutionary histories or recent iPEG diversification that could confound our molecular evolutionary comparison between iPEGs and iMEGs.

Most imprinted genes are functionally constrained

Even if imprinted genes have been subject to positive selection in their evolutionary histories, it is possible that their recent evolution has been more constrained, for example by purifying selection. To estimate the relative roles of ancestral PS (*i.e.* predating the most recent common ancestor of *A. thaliana* and *A. lyrata)* PS and

recent selective constraint, we performed McDonald-Kreitman tests (McDonald and Kreitman 1991) on our entire set of 140 imprinted orthologs from *A. lyrata* and *A. thaliana* (this included the imprinted genes for which orthologs were identified in fewer than six other plant species, and which we had not been able to analysis by PAML or HyPhy). Unambiguous *A. lyrata* orthologs were detected for 110 out of the 140 total imprinted *A. thaliana* genes (56 iPEGs and 54 iMEGs) on the basis of BLASTP alignments (supplementary table S10A). This approach assumed that the number of substitutions fixed between *A. thaliana* and *A. lyrata* was driven by ancestral positive selection and neutral substitution at nonsynonymous sites (D_N), and by neutral processes only at synonymous ones (D_S). As a result, a large D_N/D_S ratio may indicate PS. We compared these D_N and D_S counts to the numbers of nonsynonymous (P_N) and synonymous (P_S) polymorphisms within the population of 80 genome-sequenced *A. thaliana* accessions to determine the fixation index (*FI*) such that $FI=(D_N/D_S)/(P_N/P_S)$. Both P_N and P_S reflect a combination of neutral and deleterious alleles and thus represent an expected value for a neutral D_N/D_S if no ancestral PS has occurred. If $FI>1$, then ancestral adaptation through beneficial non-synonymous changes in the most recent common ancestor of *A. thaliana* and *A. lyrata* can be concluded to have occurred; alternately, if $FI<1$, then it implies that purifying selection on the ancestral lineage was the predominant selective force. For the 110 imprinted genes, we found that D_N/D_S (1.139) approximated P_N/P_S (1.196) with FI=0.952 (table 3) and conclude that there is no evidence of relaxed selective constraints. (We note that neither D_N/D_S and P_N/P_S ratios of these imprinted gene sets were biased by outliers (Daub, et al. 2014)). To further examine the recent selective pressures acting on *A. thaliana* imprinted genes, we also performed Direction of Selection (DoS) analysis which can produce more accurate estimates of selection, especially for highly conserved genes. In agreement with the results of the McDonald-Kreitman test, DoS analysis did not indicate any evidence of relaxed selective constraints (supplementary table S10B) according to the Tarone and Greenland Neutrality Index (NI_{TG}=1.237; table 3). Here, NI >1 indicates that negative selection is preventing fixation of harmful mutations.

We also compared these values to those of the *A. thaliana* genome as a whole and found no evidence for imprinted genes differing from the genome-wide pattern (fig.

5). This suggests that the imprinted genes have been subject to similar selective processes as other genes since the divergence of *thaliana-lyrata* (supplementary fig. S3): the same relative proportions showed patterns of PS ($D_N/D_S >> P_N/P_S$), ancestral purifying selection (low D_N/D_S), neutrality ($D_N/D_S \sim P_N/P_S$), or potential pseudogenization evidenced by relaxed selective constraint (high P_N/P_S and high D_N/D_S) (Yang, et al. 2011; Wang, et al. 2012). In contrast to the PAML and HyPhy analysis of selection from before the *thaliana-lyrata* divergence, no difference was apparent between iMEGs and iPEGS (supplementary fig. S3). Both McDonald-Kreitman and DoS analysis identified signatures of purifying selection on the same group of 13 genes (12% of the total, supplementary table S10A and B) while six putative pseudogenes were discovered (5% of the total, supplementary table S11): as expected, none of these showed any evidence of PS. As imprinted pseudogenes could potentially bias the overall analysis, their effect was assessed by comparing the baseline *FI* (0.952) to the expected fixation index (*eFI,* 1.205) determined from the expected contingency table values of D_N, D_S, P_N, P_S for each of the 110 imprinted genes (Axelsson and Ellegren 2009). This higher *eFI* suggested population-level mutations were negatively correlated with purifying selection, presumably due to deleterious alleles segregating within the 80 accessions and supporting previous reports of high P_N values in *A. thaliana* (Huber et al. 2014). This is also important as relaxed selective constraints (evident from a high level of within-*A. thaliana* nonsynonymous changes) would have confounded our interspecies tests for positive selection, and because previous work has shown that the average effect of nonsynonymous changes in *A. thaliana* is slightly deleterious (Bustamante, et al. 2002).

Comparison of the results of PAML and HyPhy analysis, McDonald-Kreitman tests and DoS demonstrates that the imprinted genes subject to positive selection in interspecies analysis using at least six genomes do not show any strong evidence of positive selection since the divergence of *A. thaliana* and *A. lyrata*. We conclude that genes with different evolutionary trajectories are regulated by genomic imprinting in *A. thaliana*, including some subject to pseudogenization while non-pseudogenized genes show signatures of ancestral PS with stronger signatures of PS predating the *thaliana-lyrata* split. Estimating the timing of these events with greater accuracy, and

449 determining their effects in extant populations, will provide a basis for future
450 determination of the selective pressures involved in the evolution of imprinted genes
451 in plants.

452

453 ## Discussion

454 Evolutionary trajectories of genes in mammals and angiosperms can be influenced by
455 their association with tissues involved in maternal provisioning, creating the possibility
456 for conflict over resource allocation and positive selection (PS) on the loci involved,
457 among other molecular signatures (fig. 1). In this study we have concentrated on the
458 molecular signatures of conflict acting on coding sequences of imprinted genes in
459 which alleles are expressed at different levels depending on whether they are
460 maternally- or paternally-derived (denoted iMEGs and iPEGs respectively; (Köhler, et
461 al. 2012)). The phenotypes associated with certain imprinted genes under PS in
462 animals (*Igfr*) and plants (*AlMEDEA*) supports the possibility of conflict-driven PS
463 (Spillane, et al. 2007; Miyake, et al. 2009; Wawrzik, et al. 2010; McCole, et al. 2011).
464 However, we have previously demonstrated that there is no strict concordance
465 between evidence of positive selection and imprinting status in mammals (O'Connell,
466 et al. 2010), and how conflict affects imprinted plant genes in general remains
467 unknown.

468 In this study, we have performed a comprehensive ortholog-based analysis of
469 selective pressures on genes subject to genomic imprinting in the seed endosperm of
470 *A. thaliana* and have demonstrated signatures of elevated PS (tables 1 and 2; figs. 2
471 and 3; fig. S1). To ensure these conclusions are robust, we have considered and
472 accounted for the effects of possible endosperm-specific effects and of differences in
473 gene age (fig. 4) and have accounted for potential confounding by genes expressed
474 uniparentally from maternal tissues (fig. S2). As approaches for inferring selection
475 pressures may be limited by their own inherent assumptions, we took a multiple-
476 methodology approach. For example, PAML makes the assumption that selective
477 pressures do not change on the branches where it is inferred, while HyPhy allows
478 branch-specific selection to change across all branches. We used two methodologies
479 for our ortholog-based analyses (PAML and HyPhy) and for our analysis of extant *A.*

480 *thaliana* populations (McDonald-Kreitman and Direction of Selection tests). In fact, the
481 30 imprinted genes founds to be under PS by PAML analysis were confirmed in every
482 case confirmed as such by at least two HyPhy methods (tables S1, S2), while similar
483 conclusions were derived from both McDonald-Kreitman and DoS approaches (table
484 3). We also note that it is not currently feasible to assess such changes at gene
485 regulatory sequences across lineages, so our estimates for selection levels across loci,
486 based as they are on coding-sequences alone, may in fact be underestimates.

487 It should be noted that some assumptions still remain within our analyses. For
488 example, all Dn/Ds based methods for estimating selective pressure variation from
489 sequence data assume that Ds is a proxy for neutral evolution, *i.e.* silent sites are not
490 under selective pressure, even though we know for example that exon splice sites can
491 be subject to selection to function the spliceosomal machinery (albeit mostly in intron-
492 rich genomes (Warnecke, et al. 2008)). To control for this, we made use of non-
493 imprinted controls, both from genome-wide data and from genes specifically
494 expressed in the endosperm in which genomic imprinting occurs in flowering plants
495 (supplementary table S4). The robustness of the results from these analyses is
496 furthermore supported by the robustness of the phylogeny used, which is
497 uncontroversial (fig. 4; https://phytozome.jgi.doe.gov/pz/portal.html), and on the
498 number of species used in each alignment, which was set at a minimum of six,
499 following experimentally-determined best practice (Anisimova et al., (2001)).

500 Combining together these analyses, and their comparison with relevant controls, we
501 conclude that accelerated evolution and preferential tendency to PS are general
502 features of imprinted genes in *A. thaliana*.

503

504 **Fixation of selected sites and significance of mating system**

505 Extant plant lineages have undergone multiple transitions between self-fertilising and
506 out-crossing reproduction. It is expected that parental conflict will be minimized by
507 increased levels of self-fertilization, which reduces or eliminates the genetic
508 divergence between maternally- and paternally-derived genomes (Haig 1997, 2013;
509 Gehring and Satyaki 2017), as well as slightly reducing the efficacy of purifying
510 selection across the genome (Payne and Alvarez-Ponce 2018). Consistent with this,

previous investigations of the imprinted maternally expressed gene (iMEG) *MEDEA* found that *MEDEA* was under positive selection in the outcrossing Brassicaceae species, *Arabidopsis lyrata,* while its non-imprinted paralog *SWINGER* was not; but that neither gene was under positive selection in the largely inbreeding congener, *A. thaliana* (Spillane, et al. 2007; Miyake, et al. 2009). This was interpreted as a consequence of reduced genomic conflict due to inbreeding (Garnier, et al. 2008; McKeown, et al. 2013). The findings of our present study indicate that almost all of the positively selected sites are now fixed across populations in extant *A. thaliana* which may indicate that conflict has been reduced in this largely self-pollinated species: while the levels of outcrossing in *A. thaliana* can reach 18% in natural populations in exceptional cases, it is generally much lower (Bomblies, et al. 2010).

The fixation of sites under positive selection in imprinted genes of *A. thaliana* is consistent with hypotheses that imprinting may in some cases be a relic of its outbreeding past (Brandvain and Haig 2005), perhaps because loss of imprinting to protect against deleterious recessive mutations only occurs very slowly (Wilkins and Haig 2003b). In other words, the signatures of selection detected by non-synonymous changes to coding sequences retain evidence of past conflict even after any such equilibrium has been reached: our PAML analysis is in fact identifying sites which have changed under positive selection but are now at a stable equilibrium, and which no longer show signatures of such pressures in current populations (whether measured by McDonald-Kreitman tests or by Direction of Selection tests; tables S10 and S11). Whether amino acid changes at these sites have also become fixed across other plant lineages with different levels of inbreeding would be an interesting test of this hypothesis, and will be possible to test empirically when once genomic data from multiple accessions of sufficient numbers of outcrossing and inbreeding plant species becomes available. It should also be noted that clonal interference arising from inbreeding is expected to marginally reduce the efficiency of selection across the genome (Neher, et al. 2013) and potentially mask signatures of positive selection, although rates of neutral evolution at silent sites should not be affected (Good, et al. 2014), provided that the beneficial alleles co-occur in the same period of selection.

Therefore, clonal interference would mean tests for positive selection would be more prone to false negatives rather than false positives.

In addition, we have compared our rates of positive selection in imprinted loci to the genome-wide pattern for *A. thaliana*, which also adjusts for any potential confounding effects of inbreeding. Whether fixed or not, imprinted genes which have been under PS are likely to have been important for plant fitness and represent strong candidates for future functional investigations.

Imbalance between selective pressures acting on iMEGs and iPEGs

Imprinted genes in mammals can undergo different evolutionary trajectories (O'Connell, et al. 2010; McCole, et al. 2011). Our results from this study in plants demonstrate that differential selective pressures act on imprinted genes that are expressed from either the maternal or the paternal genomes. Specifically, iPEGs display higher D_N/D_S values, and are significantly more likely to be subject to PS. This finding of asymmetric selection pressures on iPEGs vs iMEGs does not fit neatly with expectations of kin conflict which predict that any PS driven by intra-genomic conflict should likely act on both genomes due to the mutual antagonism between the parents over resource allocation to the offspring, possibly on pairs of reciprocally imprinted genes encoding physically interacting offspring growth regulators (Moore and Haig 1991; Mills and Moore 2004).

Our identification of PS in iPEGs also lacks concordance with theories that propose that imprinting results from maternal-offspring co-adaptation or cytonuclear co-evolution as illustrated in fig. 1B (Wolf and Hager 2006), in line with the lack of experimental support for this model (Haig 2013; Haig 2014). Although co-evolutionary scenarios can lead to rapid evolution of genes (Wolf and Brandvain 2014), both of these scenarios would be expected to preferentially affect iMEGs (assuming maternal cytonuclear inheritance). Nor is PS in iPEGs due to genome dosage effects in the endosperm, as the levels of positive selection for iPEGs are significantly higher than biallelically-expressed endosperm genes (fig. 3). We can also rule out the possibility that PS in iPEGs could be an artifact of these genes being younger than iMEGs, because (1) there is no significant age difference between iPEGs and iMEGs, and (2) PS

572 does not affect the more recently evolved iPEGs (figs. 2 and 5). We do note that levels
573 of PS in the endosperm-expressed control set are slightly greater than the background
574 control set (fig. 3B), which could indicate the existence of unreported iPEGs within this
575 dataset, or other causes related to the role of the endosperm in seeds. Finally, our
576 results do not support an evolutionary scenario where imprinted genes arise as a
577 result of pseudogenization following gene duplication (Wolff et al., 2011), as we could
578 only identify six possible examples of this (fig. 2).
579 The finding that *A. thaliana* iPEGs are preferentially affected by PS compared with
580 iMEGs provides an interesting parallel with the evolutionary flexibility of iPEGs
581 observed in comparisons to *A. thaliana's* sister species, *Arabidopsis lyrata*. Analysis of
582 *A. lyrata* endosperm found that iPEGs were more highly expressed in *A. lyrata* than *A.*
583 *thaliana,* while expression levels of iMEGs were more highly conserved (Klosinska, et
584 al. 2016). These changes were also associated with greater variation in CHG
585 methylation and histone modification marks between at least some conserved iPEGs
586 in the two species (Klosinska, et al. 2016). Furthermore, a study in *Capsella rubella*
587 showed that iPEGs display higher levels of non-synonymous substitution, a possible
588 indicator of PS (Hatorangan, et al. 2016), suggesting that this pattern may not be
589 restricted to the *Arabidopsis* genus either but may be a common feature of imprinting
590 in, at least, the Brassicaceae. One possible explanation for the differences between
591 selective pressures acting on iMEGs and iPEGs is that kin conflict more commonly
592 involves interactions between iPEGs and genes expressed in maternal tissues such as
593 the sporophytic seed coat (which are also involved in maternal provisioning (Orozco-
594 Arroyo, et al. 2015)), rather than with iMEGs in the endosperm. This would lead to
595 conflict that was indirect in nature, rather than involving physical interactions
596 between antagonistic pairs of iMEGs and iPEGs (McVean and Hurst 1997). Intriguingly,
597 an analysis of parental conflict in *A. lyrata* populations with different levels of
598 outbreeding suggested that conflict involving indirect interactions between paternal
599 factors and the female sporophyte ('the kinship model') was favoured in more self-
600 fertile populations, while direct interactions between proteins encoded by imprinted
601 genes in the endosperm tended to be lost as outcrossing reduced (Willi 2013). This
602 would also fit with the discovery that genes which are strongly expressed in the seed
603 coat of *A. thaliana* can also evolve under positive selection (Schon and Nodine 2017).

We also note that antagonism between the developing endosperm and another maternal tissue, the nucellus, has been proposed as a key characteristic of seed development in *A. thaliana* (Xu, et al. 2016). Analysis of the genetic interactions between maternal seed coat or nucellus with iPEGs which regulate seed size (such as *ADMENTOS;* (Kradolfer, et al. 2013)) will therefore be required to clarify whether parental conflict occurs in *A. thaliana* and related species, and if so by what mechanism.

Further possible explanations for the differences in selective pressures acting on iMEGs and iPEGs could include differential breadth of expression patterns (including in somatic tissues) or wider interaction networks which could theoretically place iMEGs under greater constraints due to risk of pleiotropic interactions. Alternatively PS could also be due to so-called 'arms races' between siblings that do not share the same paternal parent (Sadras and Denison 2009), which is more likely among paternally-derived 'patrigenes' than maternally-derived 'matrigenes' (Haig 2013). It has been shown that PS in flowering plants can be driven by pre-fertilization sexual conflict between male genomes during pollen tube competition (Gossmann, et al. 2014), in a manner analogous to competition between animal sperm (Torgerson, et al. 2002), such that positive selection at iPEGs could be triggered by conflict between the paternal genomes of endosperm tissues within seeds developing on the same plant (or in the same fruit). Paternal genetic variation is known to influence resource allocation in embryos by up to 10% in *A. thaliana* (House, et al. 2010), which could be sufficient to drive conflict between paternal alleles. Finally, if this pattern was also conserved in monocots, it could explain reports that paternally-derived expression-QTLs (eQTLs) have major roles in determining transcription levels in hybridized maize seed (Swanson-Wagner, et al. 2009). Finally, the most active evolutionary signatures acting at iPEGs in different species of Brassicaceae (this study; (Hatorangan, et al. 2016; Klosinska, et al. 2016)), in which multiple shifts of mating system have occurred, could suggest that shifting patterns of paternal relatedness, and hence, patrigenic phenotypic optima for seed size, could lead to continual evolutionary pressure manifested in different ways, such as changes to transcription level, epigenetic marks, and changes to the nucleotide and amino sequence. More generally, models of imprinting and conflict suggest that matrigenes typically favour phenotypes

intermediate to those favoured by patrigenes and maternal alleles (Burt and Trivers 1998; Wilkins and Haig 2002, 2003a; Haig 2013), in which case, positive selection for conflict with maternal tissues would be stronger on paternally expressed imprinted genes than on maternally expressed ones. If so, the same trend might be expected to be common across seed plants: analysis of selective pressures acting on imprinted genes in a more distantly related group such as the cereals could be instructive in testing this hypothesis.

Given these different, and non-mutually exclusive possibilities, careful analysis of the functions of the genes and codons subject to PS will be needed to clarify the underlying impacts of the patterns we observe on the biology of the plant. Although experimental characterization for many genes has yet to be fully performed, we note that one of the iPEGs we have identified to be under PS is *NRPD1a*, which encodes a subunit of RNA Pol IV, while other sRNA genes are not subject to PS (table S8). RNA Pol IV is involved in control of transposable elements via RNA directed DNA methylation (RdDM) and has recently also been identified as a regulator of allelic dosage in the endosperm (Erdmann, et al. 2017). Interestingly, the largest subunits of PolV (NRPE1), which is also implicated in the activity of 24-nt sRNAs in RNA-directed DNA methylation (RdDM), has also been reported to evolve rapidly through restructuring of intrinsically disordered repeats within its Argonaute-binding platform (Trujillo, et al. 2016). In the case of NRPD1a, this subunit is involved in physically binding transposable elements including those expressed in maternal tissues in seeds (Mosher, et al. 2009). Hence, it is possible that PS could be driven by conflict between paternally-expressed proteins and maternally-controlled transposable elements, or to interactions with the maternally-derived genomes of the endosperm in the case of dosage control (Erdmann, et al. 2017). Interestingly, *NRPD1a* does not appear to be an iPEG in *A. lyrata,* although two other genes encoding subunits of complexes involved in the RdDM pathway are (Klosinska, et al. 2016). Further functional characterization of the positively selected subunits will be needed to distinguish these possibilities.

We note that positive selection has been reported from the iMEG *MEDEA* in the predominantly outcrossing *A. lyrata*, but that this selective pressure has been lost in the inbreeding *A. thaliana* lineage (Spillane, et al. 2007). This lends further support to the hypothesis that positive selection persists between iPEGs and the maternal

668 sporophyte but not between iPEGs and iMEGs during the transition to self-fertilization
669 (Willi 2013). Analysis of signatures of selective pressure on the components of the FIS
670 complex across multiple plant species will be essential for clarifying the effects of
671 parental conflict in imprinting, endosperm development and speciation.

672

673 ## Conclusions

674 The study of imprinted genes in both plants and mammals has identified examples of
675 positive Darwinian selection (Spillane, et al. 2007; O'Connell, et al. 2010; Wawrzik, et
676 al. 2010). Our study demonstrates that while imprinted genes expressed in the
677 endosperm of *Arabidopsis thaliana* are rapidly evolving due to positive selection, such
678 positive selection is preferentially associated with imprinted paternally expressed
679 genes (iPEGs). This raises the possibility that ongoing intra-genomic conflicts between
680 paternally-expressed imprinted genes (iPEGs), or between iPEGs and genes
681 functioning in the maternal sporophyte, could be evolutionary drivers and maintainers
682 of imprinting in plants. The iPEG and iMEG genes we have identified under positive
683 selection are involved in processes such as auxin biosynthesis (e.g. *YUCCA10, TAR1*)
684 and epigenetic regulation involving small RNAs and chromatin remodelling (*NRPD1a*).
685 Overall, our results identify the subset of imprinted genes, both iPEGs and iMEGs,
686 which are strong candidates for having functional effects that are antagonistic with
687 other molecular factors, in a manner that results in their evolution under positive
688 selection.

689

691 # Methods

692 ## Identification of imprinted genes and orthologs

693 An *A. thaliana* imprinted gene set was compiled from a number of high-throughput
694 expression screens (Gehring, et al. 2011; Hsieh, et al. 2011; McKeown, et al. 2011;
695 Wolff, et al. 2011), supplemented by other studies (Vielle-Calzada, et al. 1999;
696 Kinoshita, et al. 2004; Köhler, et al. 2005; Jullien, et al. 2006; Tiwari, et al. 2008;
697 Gehring, et al. 2009; Gerald, et al. 2009) to yield 140 high-confidence imprinted genes
698 (supplementary table S1). Orthologs were identified across 34 plant species for which
699 assembled whole genome sequences were publically available (fig. 4). Peptide and
700 CDS sequences for 32 species were downloaded from Phytozome v8.0 (Goodstein, et
701 al. 2012); *Cajanus cajan* sequences were accessed from (Varshney, et al. 2012) and
702 *Lotus japonicus* from the PlantGDB database (Dong et al., 2004). In all cases, the
703 longest transcript was used as the representative transcript for each gene. To
704 minimize the number of false positives and ensure tight clustering of genes families,
705 we detected orthologous relationships between sequences using OrthoMCL (Li, et al.
706 2003; Chen, et al. 2007). We also chose to use maximum likelihood methods based on
707 codon models of sequence evolution as these are considered to be more robust than
708 alternative methods such as sliding window approaches (Schmid and Yang 2008). As
709 the power of maximum likelihood methods increases with greater taxonomic
710 representation and breadth (Anisimova, et al. 2001), we considered only the 62
711 imprinted genes for which orthologous genes could be identified from at least six
712 other species (in addition to *A. thaliana* itself). As controls, random sets of 100 genes
713 were generated representing the entire *A. thaliana* genome, and a subset of
714 endosperm-specific genes derived from (Belmonte, et al. 2013) (Supplementary table
715 S4). To ensure a valid comparison with the imprinted dataset, only genes belonging to
716 orthology clusters present in at least six other species (Anisimova, et al. 2001) were
717 included in these control sets.
718

719 ## Multiple sequence alignments

720 Multiple sequence alignments for each gene family were constructed using MUSCLE
721 (Edgar 2004) and MAFFT (Katoh and Toh 2008) and were compared in AQUA (Muller,

et al. 2010). RASCAL (Thompson, et al. 2003) was used to refine the alignments and norMD (Thompson, et al. 2001) was used to assess their quality. Alignments with a norMD score <0.6 were considered as low quality. Poorly aligned sequences were removed from alignments with norMD <0.6 and norMD was recalculated: if the norMD score subsequently increased to >0.6, the alignment was retained for further analysis. Nucleotide sequence alignments were generated for each family using the amino acid alignment and original nucleotide sequence files, using in-house software. Recombinant sequences were also removed identified using RDP3 (Martin, et al. 2010) with two substitution-based methods – GENECONV (Sawyer 1989) and MaxChi (Smith 1992) – and two phylogenetic-based methods – BOOTSCAN (Martin, et al. 2005) and SiScan (Gibbs, et al. 2000). Sequences were considered as recombinant if a recombination event was significantly predicted by at least one substitution-based method *and* at least one phylogenetic-based method. The percentage of gaps in the alignments were calculated using TrimAL (Capella-Gutierrez, et al. 2009) (-sgc option) and predicted sites of positive selection which overlapped with regions of poor alignment (gaps > 40%) were discarded.

Tree building

Models for protein sequence evolution were generated using modelgenerator (Keane, et al. 2006). Phylogenetic trees were inferred using RAxML (Randomized Axelerated Maximum Likelihood) version 7.2.6 (Stamatakis 2006) with 1000 bootstrap replicates and the rapid bootstrapping algorithm. The codeML analysis was run on all clades of interest for genes with >80 sequences in their orthology clusters (supplementary table S12A) and on control genes from genome-wide and endosperm-expressed datasets (supplementary table S12B).

Selective pressure analysis

Selective pressure analysis was conducted using PAML version 4.4e (Yang 2007). Both lineage-specific models (Yang 1998; Yang and Nielsen 2002) and site-specific models (Yang and Swanson, 2002) were evaluated using likelihood ratio test (LRT). Sequences were considered to exhibit lineage-specific selective pressure if the likelihood ratio test for ModelA was significant in comparison to both ModelA null and M1Neutral,

where M1Neutral is a neutral model that allows two site classes: ω_0=0 and ω_1=1. Model A assumes the two site classes are the same in both foreground and background lineages (ω_0=0 and ω_1=1) and ω_1 was calculated from the data. Model A null is the null hypothesis for this model and allows sites to be evolving under either purifying selection, or to be neutrally evolving in the background lineages. For site-specific analyses, LRTs were conducted to compare models M7 and M8a with model M8. The test compared the neutral model M7, which assumes a β distribution for ω over sites and the alternative model M8 (β and ω), which adds an extra site class of positive selection. M8a is the null hypothesis of M8 where the additional category is neutral, i.e. ω=1. An automated CodeML wrapper (VESPA, (Webb, et al. 2017)) was used to prepare all the codeML files, to parse the PAML output and perform the likelihood ratio test. After ML estimates of model parameters were obtained, we used two bayesian approaches to infer the posterior probability of the positively selected sites: Bayes Empirical Bayes (BEB) and Naïve Empirical Bayes (NEB). BEB reduces the rate of false positives when analyzing small datasets and retains the power of NEB when analyzing large datasets (Yang and Nielsen 2002). Therefore if NEB and BEB were both predicted the results from BEB were preferred.

Use of HyPhy to estimate rates of Darwinian selection

A second positive selection pressure analysis of genes which were predicted to be under positive selective pressure by PAML was conducted using HyPhy version 2.2.4 (Pond and Muse 2005). We employed the following three approaches from the HyPhy package: FEL (Fixed effects Likelihood), SLAC (Single-Nucleotide Ancestor Counting), and MEME (Mixed Effects Model of Evolution). FEL tests for both positive and negative selection per individual site, and can identify individual sites that have undergone pervasive diversifying selection while SLAC is an approximate method similar to FEL (Kosakovsky Pond and Frost 2005). We also applied the MEME model from the HyPhy package which tests for episodic selection at individual sites and on specific branches: MEME does not assume that the strength and direction of selection is constant across all lineages (Murrell, et al. 2012). Only sites resolved as being under PS by at least two methods were considered confirmed by HyPhy.

Tests including population-level variation

A. lyrata orthologs of 140 imprinted *A. thaliana* genes were identified using reciprocal best hits (RBH) of which 110 were also derived as the best hits of the *A. thaliana* genes in reciprocal BLAST. *A. thaliana* and *A. lyrata* CDS were aligned as described above. 80 accession SNP data for *A. thaliana* was downloaded from the 1001 genome project (http://1001genomes.org/data/MPI/MPICao2010/releases/current/genome_matrix) and SNPs mapped to the reference genome using a custom-made python script. McDonald-Kreitman tests were performed on each imprinted gene using a python script that uses egglib library to calculate D_N, D_S, P_N and P_S values and calculated the ratio using Fisher's exact test. Fixation indices (*FI*) were determined as *FI*= (D_N/D_S)/(P_N/P_S) with expected fixation index (*eFI*) calculated as reported previously (Axelsson and Ellegren 2009). Genes with zero D_N/D_S and P_N/P_S were not considered for *FI* calculations. Direction of selection (*DoS*) ((Stoletzki and Eyre-Walker 2011) was calculated using $D_N/(D_N+D_S)-P_N/(P_N+P_S)$; the Tarone and Greenland Neutrality Index (NI_{TG}) was calculated using the Distribution of Fitness Effect (DoFE) package.

Acknowledgements

This work was supported by grant funding from Science Foundation Ireland (SFI) to CS (Principal Investigator Grants 08/IN.1/B193 and 13/IA/1820) and MJO'C (Research Frontiers Programme Grant EOB2673). We would like to thank the DJEI/DES/SFI/HEA funded Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. The support of the NUI Galway Thomas Crawford Hayes Trust is also acknowledged.

Tables

Table 1

Gene Class	Mean ω (D_N/D_S)	Median ω (D_N/D_S)
iPEGs	0.4265±0.053	0.3339
iMEGs	0.5045±0.061	0.3314
whole genome	0.2436±0.002	0.1814

Table 1. D_N/D_S ratios (ω) of iPEGs and iMEGs compared to whole genome.

818 **Table 2**

819

	iMEGs	iPEGs	Total
Total number of genes tested	**30**	**32**	**62**
Genes subject to lineage-specific selection only	2 (6.7%)	4 (12.5%)	**6 (0.9%)**
Genes subject to site-specific selection only	7 (23.3%)	9 (28.1%)	**16 (25.8%)**
Genes subject to both lineage- and site-specific selection	2 (6.7%)	6 (18.8%)	**8 (12.9%)**
TOTAL	**11** (36.7%)	**19** (59.4%)	**30 (48.4%)**

820

821 **Table 2:** Numbers of iMEGs and iPEGs determined to be under positive selection.

Table 3.

Parameter	Polymorphism	Divergence
Non-synonymous substitutions *(D_N)*	1988	4740
Synonymous substitutions *(D_s)*	1662	4161
Ratio of Non-synonymous /Synonymous *(D_N/D_S)* substitutions	1.196	1.139
Fixation Index (*FI*)[1]	0.952	
Expected Fixation Index (*eFI*)[2]	1.205	
Neutrality Index (NI_{TG})[3]	1.237	
α[4]	-0.210	

Table 3. Calculations derived from McDonald-Kreitman analyses of genes regulated by genomic imprinting in the *A. thaliana* endosperm; values were derived from comparisons between 80 sequenced *A. thaliana* accessions, using *A. lyrata* as outgroup. Full gene-by-gene results from which these figures were derived are presented in Suppl. table S7. [1]Observed fixation index, calculated according to $FI=(D_N/D_S)/(P_N/P_S)$. [2]Expected fixation index (*eFI*). [3]The Tarone and Greenland Neutrality Index (NI_{TG}). [4]Proportion of fixed nonsynonymous mutations driven by fixed positive selection fixed in *A. thaliana,* α= (*FI-eFI*)/*eFI*.

Figure legends

Fig. 1. Summary of scenarios for selection on imprinted plant genes. Schematic of *A. thaliana* seed summarising the impacts of genomic imprinting on genetic selection as predicted by major hypotheses for genomic imprinting. In each case, the diploid F1 embryo is shown in dark green, surrounded by the triploid F1 endosperm, shown in yellow) in which imprinting occurs, and the diploid seed coat (SC) which is part of the maternal sporophyte, shown in light green. (A) Intra-Genomic conflict in which antagonism between matrigenes and patrigenes over resource allocation results in physical interactions between iMEGs and iPEGs (Spillane, et al. 2007). (B) Co-adaptation models predict that any selective pressure should be concentrated on iMEGs which are co-inherited with cytoplasmic genomes in *A. thaliana* (Wolf and Brandvain 2014). (C) Indirect conflict or "Kinship Model" predicts that conflict between iPEGs and genes expressed in maternal tissues (e.g. seed coat, scMEG, or other sporophyte tissues) leads to positive selection on iPEGs (Willi 2013).

Fig. 2. Size of orthology clusters to which imprinted *A. thaliana* genes belong. Orphans are defined according to (Donoghue, et al. 2011); genes present in orthology clusters >6 were considered for further selective pressure variation analysis.

Fig. 3. Summary of the number of genes under positive selection in the dataset. (A) Numbers of imprinted *A. thaliana* genes under site and/or lineage specific PS; (B, C) the percentages of *A. thaliana* iMEGs and iPEGs subject to lineage-specific (B) or site-specific (C) PS compared to the percentages in control sets of endosperm-expressed ('Endosperm') or genome-wide ('Genome') biallelic genes; control gene-sets are listed in supplementary table S4.

Fig. 4: Phylogeny of the 34 species included in our analyses and the age distribution of iMEGs and iPEGs. (A) shows the frequency of age class (AC) for the iMEGs and iPEGs tested. AC0, *A. thaliana* specific; AC1, *A. lyrata*; AC2, Brassicaceae; AC3, Brassicales-Malvales; AC4, Rosid; AC5, Eudicot; AC6, Angiosperm; AC7, Tracheophyte;

AC8, Embryophyte; AC9, Viridiplantae. (B) Consensus phylogenetic relationships of all 34 species; the phylogenetic position of the age classes and the known whole genome duplication events for the species included in the study are also highlighted (Vanneste, et al. 2014).

Fig. 5. Distribution of D_N/D_S and P_N/P_S ratios for imprinted genes compared to all protein-coding genes in *A. thaliana*. X-axis depicts Pn/Ps ratios, Y-Axis represents D_N/D_S ratios. Green dots = genes under purifying selection; red dots = genes under positive selection; yellow dots = genes under neutral evolution; black triangles = *A. thaliana* imprinted genes; blue triangles = pseudogenes with high D_N/D_S and high P_N/P_S. No clustering was observed.

Supplementary material legends

Suppl. fig. S1. Schematic of workflow for identifying positive Darwinian selection.

Suppl. fig. S2. Tissue specific transcript enrichment of previously published datasets. The test for tissue specific transcript contamination (Schon and Nodine 2017) was applied to the datasets generated by (Wolff, et al. 2011), indicating that only the suspensor datasets exhibited an enrichment for non-tissue specific transcripts. Endosperm datasets exhibited no significant enrichments for other tissues.

Suppl. fig. S3. Distribution of D_N/D_S and P_N/P_S ratios for imprinted genes compared to all protein-coding genes in *A. thaliana* disaggregated between iPEGs and iMEGs.

Suppl. table S1. 140 experimentally-validated genes subject to genomic imprinting in the *A. thaliana* endosperm and original references.

Suppl. table S2. Confirmation of positive selection by HyPhy.

Suppl. table S3. Comparison of imprinted genes with methodology of Schon and Nodine (2017) for identifying potential maternal contaminants.

Suppl. table S4. Results of selection analysis for control datasets.

Suppl. table S5. Details of all sites subject to positive selection in imprinted genes of *A. thaliana*. Genes whose imprinted status is predicted to be 'low-confidence' according to the methodology of Schon and Nodine (2017) are indicated.

Suppl. table S6. Amino acids encoded by sites subject positive selection in imprinted genes of *A. thaliana*.

Suppl. tables S7. Percentage of fixation of lineage-specific positively selected sites in Ath 80 population data identified by codeml. Data for all sites with posterior probability >0.5 is displayed.

Suppl. table S8. Results of tests for positive selection on genes of the sRNA processing pathway.

Suppl. table S9. Age classes of iMEGs and iPEGs (as shown in fig. 3).

Suppl. table S10. Results of all McDonald-Kreitman (10A) and Direction of Selection (DoS, 10B) tests.

Suppl. table S11. High Dn/Ds and High Pn/Ps of *A. thaliana* imprinted genes identified as pseudogenes.

Suppl. table S12. Gene trees for (A) imprinted genes and (B) non-imprinted control genes tested for positive selection.

References

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. Molecular biology and evolution 18:1585-1592.

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. Molecular biology and evolution 30:2475-2486.

Axelsson E, Ellegren H. 2009. Quantification of adaptive evolution of genes expressed in avian brain and the population size effect on the efficacy of selection. Molecular biology and evolution 26:1073-1079.

Baroux C, Spillane C, Grossniklaus U. 2002. Evolutionary origins of the endosperm in flowering plants. Genome Biology 3:1026.1021-1026.1025.

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD. 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proceedings of the National Academy of Sciences USA 110:E435-E444.

Bernasconi G, Ashman T-L, Birkhead T, Bishop J, Grossniklaus U, Kubli E, Marshall D, Schmid B, Skogsmyr I, Snook R. 2004. Evolutionary ecology of the prezygotic stage. Science 303:971-975.

Bomblies K, Yant L, Laitinen RA, Kim ST, Hollister JD, Warthmann N, Fitz J, Weigel D. 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. PLOS Genetics 6:e1000890.

Brandvain Y, Haig D. 2005. Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. The American Naturalist 166:330-338.

Burt A, Trivers R. 1998. Genetic conflicts in genomic imprinting. Proceedings of the Royal Society of London B: Biological Sciences 265:2393-2397.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. 2002. The cost of inbreeding in Arabidopsis. Nature 416:531-534.

954 Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle
955 O, Lippert C. 2011. Whole-genome sequencing of multiple *Arabidopsis thaliana*
956 populations. Nature genetics 43:956-963.
957 Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated
958 alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.
959 Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at
960 synonymous sites in mammals. Nat Rev Genet 7:98-108.
961 Chen F, Mackey AJ, Vermunt JK, Roos DS. 2007. Assessing performance of orthology
962 detection strategies applied to eukaryotic genomes. PLoS ONE 2:e383.
963 Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from
964 animals and plants. Reproduction 131:11-22.
965 Costa Liliana M, Yuan J, Rouster J, Paul W, Dickinson H, Gutierrez-Marcos Jose F. 2012.
966 Maternal control of nutrient allocation in plant seeds by genomic imprinting. Current
967 Biology 22:160-165.
968 Crespi B, Semeniuk C. 2004. Parent-offspring conflict in the evolution of vertebrate
969 reproductive mode. The American Naturalist 163:635-653.
970 Daub JT, Dupanloup I, Robinson-Rechavi M, Excoffier L. 2014. Inference of
971 evolutionary forces acting on human biological pathways.
972 Domazet-Loso T, Brajkovic J, Tautz D. 2007. A phylostratigraphy approach to uncover
973 the genomic history of major adaptations in metazoan lineages. Trends Genet 23:533-
974 539.
975 Donoghue MT, Keshavaiah C, Swamidatta SH, Spillane C. 2011. Evolutionary origins of
976 Brassicaceae specific genes in *Arabidopsis thaliana*. BMC evolutionary biology 11:47.
977 Eamens A, Vaistij FE, Jones L. 2008. NRPD1a and NRPD1b are required to maintain
978 post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis.
979 The Plant Journal 55:596-606.
980 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
981 throughput. Nucleic Acids Research 32:1792-1797.
982 Erdmann RM, Satyaki PR, Klosinska M, Gehring M. 2017. A small RNA pathway
983 mediates allelic dosage in endosperm. Cell reports 21:3364-3372.

Fort A, Tuteja R, Braud M, McKeown PC, Spillane C. 2017. Parental-genome dosage effects on the transcriptome of F1 hybrid triploid embryos of Arabidopsis thaliana. The Plant Journal 92:1044-1058.

Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K. 2011. Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in Plant Science 16:108-116.

Garnier O, Laoueille-Duprat S, Spillane C. 2008. Genomic imprinting in plants. Epigenetics 3:14-20.

Gehring M, Bubb KL, Henikoff S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324:1447-1451.

Gehring M, Missirian V, Henikoff S. 2011. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PLoS One 6:e23687.

Gehring M, Satyaki P. 2017. Endosperm and imprinting, inextricably linked. Plant Physiology 173:143-154.

Georges M, Charlier C, Cockett N. 2003. The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. Trends in Genetics 19:248-252.

Gerald JNF, Hui PS, Berger F. 2009. Polycomb group-dependent imprinting of the actin regulator *AtFH5* regulates morphogenesis in *Arabidopsis thaliana*. Development 136:3399-3404.

Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573-582.

Good BH, Walczak AM, Neher RA, Desai MM. 2014. Genetic diversity in the interference selection limit. PLOS Genetics 10:e1004222.

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 Research 40:D1178-D1186.

Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ. 2014. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. Molecular biology and evolution 31:574-583.

Haig D. 2014. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. Heredity 113:96.

Haig D. 2004. Genomic imprinting and kinship: how good is the evidence? Annu. Rev. Genet. 38:553-585.

Haig D. 2013. Kin conflict in seed development: an interdependent but fractious collective. Annual review of cell and developmental biology 29:189-211.

Haig D. 2000. The kinship theory of genomic imprinting. Annual review of ecology and systematics 31:9-32.

Haig D. 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. Proceedings of the Royal Society of London B: Biological Sciences 264:1657-1662.

Haig D, Úbeda F, Patten MM. 2014. Specialists and generalists: the sexual ecology of the genome. Cold Spring Harbor perspectives in biology 6:a017525.

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 of the Royal Society of London Series B-Biological Sciences 333:1-13.

Hatorangan MR, Laenen B, Steige K, Slotte T, Köhler C. 2016. Rapid evolution of genomic imprinting in two species of the Brassicaceae. The Plant Cell:tpc. 00304.02016.

House C, Roth C, Hunt J, Kover PX. 2010. Paternal effects in Arabidopsis indicate that offspring can influence their own size. Proceedings of the Royal Society B: Biological Sciences 277:2885-2893.

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 USA 108:1755-1762.

Huber CD, Nordborg M, Hermisson J, Hellmann I. 2014. Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana*. Mol Biol Evol 31:3026-3039.

Jensen JD, Bachtrog D. 2011. Characterizing the Influence of Effective Population Size on the Rate of Adaptation: Gillespie's Darwin Domain. Genome Biology and Evolution 3:687-701.

Jullien PE, Kinoshita T, Ohad N, Berger F. 2006. Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell 18:1360-1372.

1046 Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M,
1047 Matzke AJ. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA
1048 methylation. Nature genetics 37:761-765.
1049 Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence
1050 alignment program. Briefings in Bioinformatics 9:286-298.
1051 Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO. 2006. Assessment
1052 of methods for amino acid matrix selection and their use on empirical data shows that
1053 *ad hoc* assumptions for choice of matrix are not justified. BMC evolutionary biology
1054 6:29.
1055 Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T.
1056 2004. One-way control of *FWA* imprinting in Arabidopsis endosperm by DNA
1057 methylation. Science 303:521-523.
1058 Klosinska M, Picard CL, Gehring M. 2016. Conserved imprinting associated with unique
1059 epigenetic signatures in the Arabidopsis genus. Nature Plants 2:16145.
1060 Köhler C, Page DR, Gagliardini V, Grossniklaus U. 2005. The *Arabidopsis thaliana*
1061 MEDEA Polycomb group protein controls expression of *PHERES1* by parental
1062 imprinting. Nature genetics 37:28-30.
1063 Köhler C, Wolff P, Spillane C. 2012. Epigenetic mechanisms underlying genomic
1064 imprinting in plants. Annual Review of Plant Biology 63:331-352.
1065 Kosakovsky Pond SL, Frost SD. 2005. Not so different after all: a comparison of
1066 methods for detecting amino acid sites under selection. Molecular biology and
1067 evolution 22:1208-1222.
1068 Kradolfer D, Wolff P, Jiang H, Siretskiy A, Köhler C. 2013. An imprinted gene underlies
1069 postzygotic reproductive isolation in *Arabidopsis thaliana*. Developmental cell 26:525-
1070 535.
1071 Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLoS Genet
1072 4:e1000304.
1073 Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for
1074 eukaryotic genomes. Genome Research 13:2178-2189.
1075 Luo J, Hall BD. 2007. A multistep process gave rise to RNA Polymerase IV of land
1076 plants. Journal of Molecular Evolution 64:101-112.

Markova-Raina P, Petrov D. 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. Genome Research 21:863-874.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26:2462-2463.

Martin DP, Posada D, Crandall KA, Williamson C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21:98-102.

McCole RB, Loughran NB, Chahal M, Fernandes LP, Roberts RG, Fraternali F, O'Connell MJ, Oakey RJ. 2011. A case-by-case evolutionary analysis of four imprinted retrogenes. Evolution 65:1413-1427.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

McKeown PC, Fort A, Spillane C. 2013. Genomic imprinting: parental control of gene expression in higher plants. Polyploid and Hybrid Genomics:257.

McKeown PC, Laouielle-Duprat S, Prins P, Wolff P, Schmid MW, Donoghue MT, Fort A, Duszynska D, Comte A, Lao NT, et al. 2011. Identification of imprinted genes subject to parent-of-origin specific expression in *Arabidopsis thaliana* seeds. BMC Plant Biology 11:113.

McVean GT, Hurst LD. 1997. Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. Proceedings of the Royal Society of London B: Biological Sciences 264:739-746.

Mills W, Moore T. 2004. Polyandry, life-history trade-offs and the evolution of imprinting at Mendelian loci. Genetics 168:2317-2327.

Miyake T, Takebayashi N, Wolf DE. 2009. Possible diversifying selection in the imprinted gene, *MEDEA*, in Arabidopsis. Molecular biology and evolution 26:843-857.

Moore T, Haig D. 1991. Genomic imprinting in mammalian development: a parental tug-of-war. Trends in Genetics 7:45-49.

Morgan CC, Loughran NB, Walsh TA, Harrison AJ, O'Connell MJ. 2010. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. BMC evolutionary biology 10:39.

1109 Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC. 2009.
1110 Uniparental expression of PolIV-dependent siRNAs in developing endosperm of
1111 Arabidopsis. Nature 460:283-286.
1112 Muller J, Creevey CJ, Thompson JD, Arendt D, Bork P. 2010. AQUA: automated quality
1113 improvement for multiple sequence alignments. Bioinformatics 26:263-265.
1114 Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting
1115 individual sites subject to episodic diversifying selection. PLOS Genetics 8:e1002764.
1116 Neher RA, Kessinger TA, Shraiman BI. 2013. Coalescence and genetic diversity in
1117 sexual populations under selection. Proceedings of the National Academy of Sciences
1118 110:15836-15841.
1119 Nielsen R. 2005. Molecular signatures of natural selection. Annual Review of Genetics
1120 39:197-218.
1121 O'Connell MJ, Loughran NB, Walsh TA, Donoghue MT, Schmid KJ, Spillane C. 2010. A
1122 phylogenetic approach to test for evidence of parental conflict or gene duplications
1123 associated with protein-encoding imprinted orthologous genes in placental mammals.
1124 Mammalian Genome 21:486-498.
1125 Orozco-Arroyo G, Paolo D, Ezquer I, Colombo L. 2015. Networks controlling seed size
1126 in Arabidopsis. Plant reproduction 28:17-32.
1127 Patwa Z, Wahl LM. 2008. The fixation probability of beneficial mutations. Journal of
1128 The Royal Society Interface 5:1279-1289.
1129 Payne B, Alvarez-Ponce D. 2018. Higher rates of protein evolution in the self-fertilizing
1130 plant Arabidopsis thaliana than in the out-crossers Arabidopsis lyrata and Arabidopsis
1131 halleri. Genome Biology and Evolution:evy053-evy053.
1132 Pond SLK, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. In. Statistical
1133 methods in molecular evolution: Springer. p. 125-181.
1134 Qiu Y, Liu S-L, Adams KL. 2014. Frequent changes in expression profile and accelerated
1135 sequence evolution of duplicated imprinted genes in Arabidopsis. Genome Biology
1136 and Evolution 6:1830-1842.
1137 Reik W, Constância M, Fowden A, Anderson N, Dean W, Ferguson-Smith A, Tycko B,
1138 Sibley C. 2003. Regulation of supply and demand for maternal nutrients in mammals
1139 by imprinted genes. The Journal of physiology 547:35-44.

1140 Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A,
1141 Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human
1142 lineage. Science 312:1614-1620.
1143 Sadras VO, Denison RF. 2009. Do plant parts compete for resources? An evolutionary
1144 viewpoint. New Phytologist 183:565-574.
1145 Sawyer S. 1989. Statistical tests for detecting gene conversion. Molecular biology and
1146 evolution 6:526-538.
1147 Schmid K, Yang Z. 2008. The Trouble with Sliding Windows and the Selective Pressure
1148 in BRCA1. PLoS ONE 3:e3746.
1149 Schon MA, Nodine MD. 2017. Widespread Contamination of Arabidopsis Embryo and
1150 Endosperm Transcriptome Data Sets. The Plant Cell 29:608-617.
1151 Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M, Ungru A, Winge P, Koncz
1152 C, Aalen RB, Schnittger A. 2011. Genome-wide transcript profiling of endosperm
1153 without paternal contribution Identifies parent-of-origin–dependent regulation of
1154 AGAMOUS-LIKE36. PLOS Genetics 7:e1001303.
1155 Smith JM. 1992. Analyzing the mosaic structure of genes. Journal of Molecular
1156 Evolution 34:126-129.
1157 Spillane C, Schmid KJ, Laoueille-Duprat S, Pien S, Escobar-Restrepo JM, Baroux C,
1158 Gagliardini V, Page DR, Wolfe KH, Grossniklaus U. 2007. Positive darwinian selection at
1159 the imprinted *MEDEA* locus in plants. Nature 448:349-352.
1160 Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
1161 with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.
1162 Stoletzki N, Eyre-Walker A. 2011. Estimation of the Neutrality Index. Molecular biology
1163 and evolution 28:63-70.
1164 Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable
1165 PS. 2009. Paternal dominance of trans-eQTL influences gene expression patterns in
1166 maize hybrids. Science 326:1118-1120.
1167 Thompson JD, Plewniak F, Ripp R, Thierry JC, Poch O. 2001. Towards a reliable
1168 objective function for multiple sequence alignments. Journal of Molecular Biology
1169 314:937-951.
1170 Thompson JD, Thierry JC, Poch O. 2003. RASCAL: rapid scanning and correction of
1171 multiple sequence alignments. Bioinformatics 19:1155-1161.

Tiwari S, Schulz R, Ikeda Y, Dytham L, Bravo J, Mathers L, Spielman M, Guzman P, Oakey RJ, Kinoshita T, et al. 2008. *MATERNALLY EXPRESSED PAB C-TERMINAL*, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. Plant Cell 20:2387-2398.

Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Molecular biology and evolution 19:1973-1980.

Trivers RL. 1974. Parent-offspring conflict. Integrative and Comparative Biology 14:249-264.

Trujillo JT, Beilstein MA, Mosher RA. 2016. The Argonaute-binding platform of NRPE1 evolves through modulation of intrinsically disordered repeats. New Phytologist.

Vanneste K, Maere S, Van de Peer Y. 2014. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Phil. Trans. R. Soc. B 369:20130353.

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, et al. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology 30:83-89.

Vielle-Calzada JP, Thomas J, Spillane C, Coluccio A, Hoeppner MA, Grossniklaus U. 1999. Maintenance of genomic imprinting at the Arabidopsis *MEDEA* locus requires zygotic DDM1 activity. Genes & Development 13:2971-2982.

Walbot V, Evans MM. 2003. Unique features of the plant life cycle and their consequences. Nature Reviews Genetics 4:369-379.

Wang L, Si W, Yao Y, Tian D, Araki H, Yang S. 2012. Genome-wide survey of pseudogenes in 80 fully re-sequenced *Arabidopsis thaliana* accessions. PLoS One 7:e51769.

Wawrzik M, Unmehopa UA, Swaab DF, van de Nes J, Buiting K, Horsthemke B. 2010. The C15orf2 gene in the Prader–Willi syndrome region is subject to genomic imprinting and positive selection. Neurogenetics 11:153-161.

Webb AE, Walsh TA, O'Connell MJ. 2017. VESPA: Very large-scale evolutionary and selective pressure analyses. PeerJ Computer Science 3:e118.

Wilkins JF. 2011. Genomic imprinting and conflict-induced decanalization. Evolution 65:537-553.

Wilkins JF, Haig D. 2001. Genomic imprinting of two antagonistic loci. Proceedings of the Royal Society of London B: Biological Sciences 268:1861-1867.

Wilkins JF, Haig D. 2003a. Inbreeding, maternal care and genomic imprinting. Journal of Theoretical Biology 221:559-564.

Wilkins JF, Haig D. 2002. Parental modifiers, antisense transcripts and loss of imprinting. Proceedings of the Royal Society of London B: Biological Sciences 269:1841-1846.

Wilkins JF, Haig D. 2003b. What good is genomic imprinting: the function of parent-specific gene expression. Nature Reviews Genetics 4:359-368.

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.

Willson MF, Burley N. 1983. Mate Choice in Plants: Tactics, Mechanisms, and Consequences: Princeton University Press.

Wolf JB, Brandvain Y. 2014. Gene interactions in the evolution of genomic imprinting. Heredity (Edinb) 113:129-137.

Wolf JB, Hager R. 2006. A maternal–offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biology 4:e380.

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.

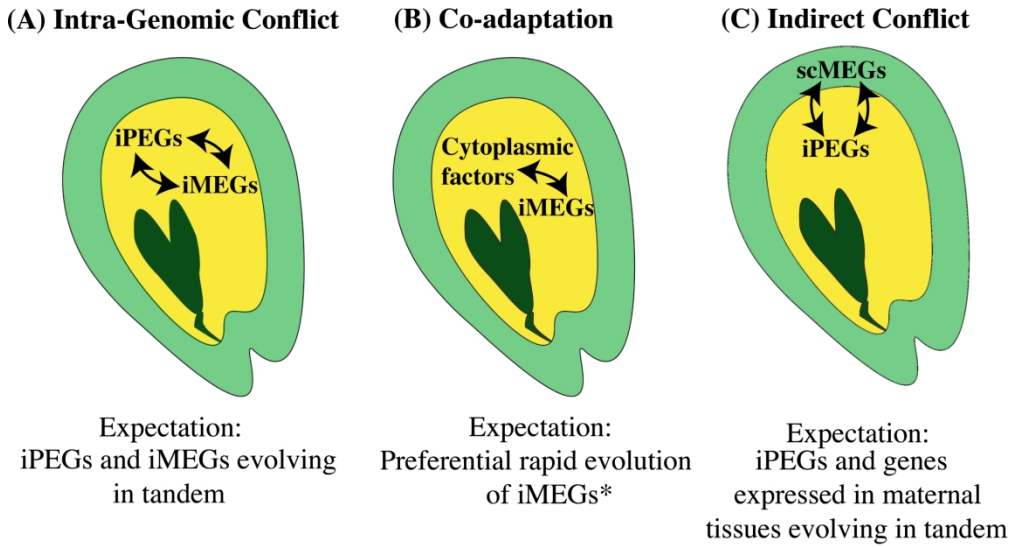
Wyder S, Raissig MT, Grossniklaus U. 2017. Consistent reanalysis of genome-wide imprinting studies in plants using generalized linear models increases concordance across datasets. bioRxiv:180745.

Xu W, Fiume E, Coen O, Pechoux C, Lepiniec L, Magnani E. 2016. Endosperm and nucellus develop antagonistically in Arabidopsis seeds. The Plant Cell:tpc. 00041.02016.

Yang L, Takuno S, Waters ER, Gaut BS. 2011. Lowly expressed genes in *Arabidopsis thaliana* bear the signature of possible pseudogenization by promoter degradation. Molecular biology and evolution 28:1193-1203.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular biology and evolution 15:568-573.

1235 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular
1236 biology and evolution 24:1586-1591.
1237 Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular
1238 adaptation at individual sites along specific lineages. Molecular biology and evolution
1239 19:908-917.
1240
1241
1242



*assumes maternal cytonuclear inheritance

Figure 1

198x117mm (300 x 300 DPI)

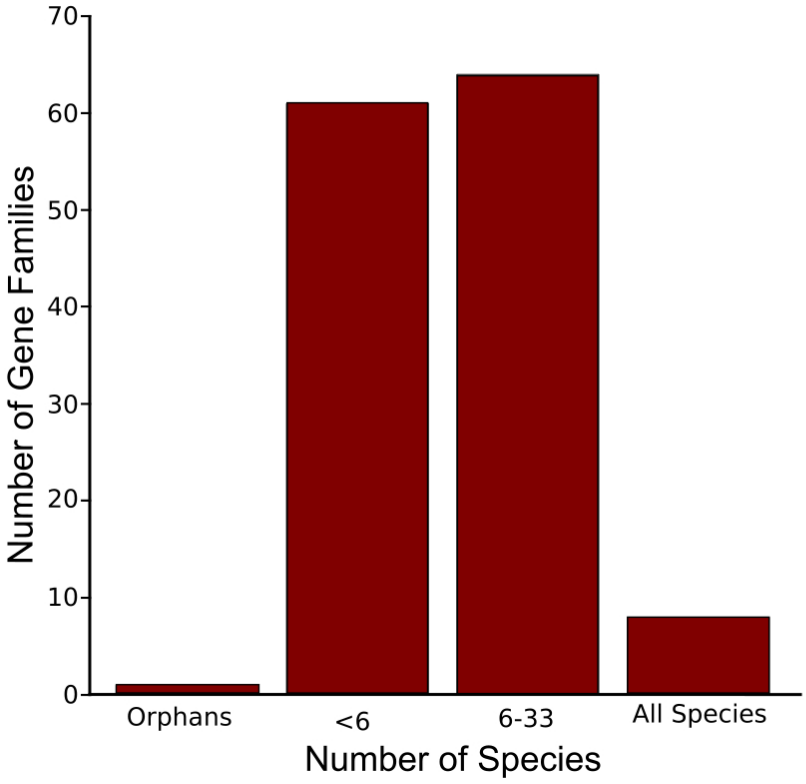

70
60
50
40
30
20
10
0
Number of Gene Families
Orphans
<6
6-33
All Species
Number of Species

Figure 2

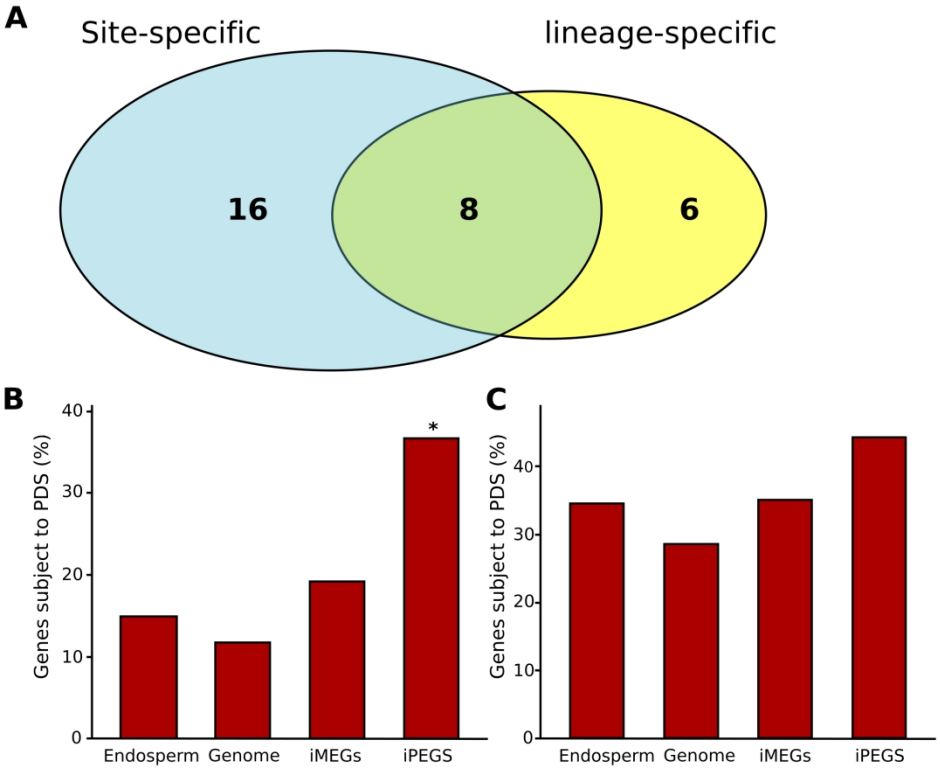

A
Site-specific
lineage-specific
16
8
6
B
Genes subject to PDS (%)
40
30
20
10
0
*
Endosperm
Genome
iMEGs
iPEGS
C
Genes subject to PDS (%)
40
30
20
10
0
Endosperm
Genome
iMEGs
iPEGS

Figure 3

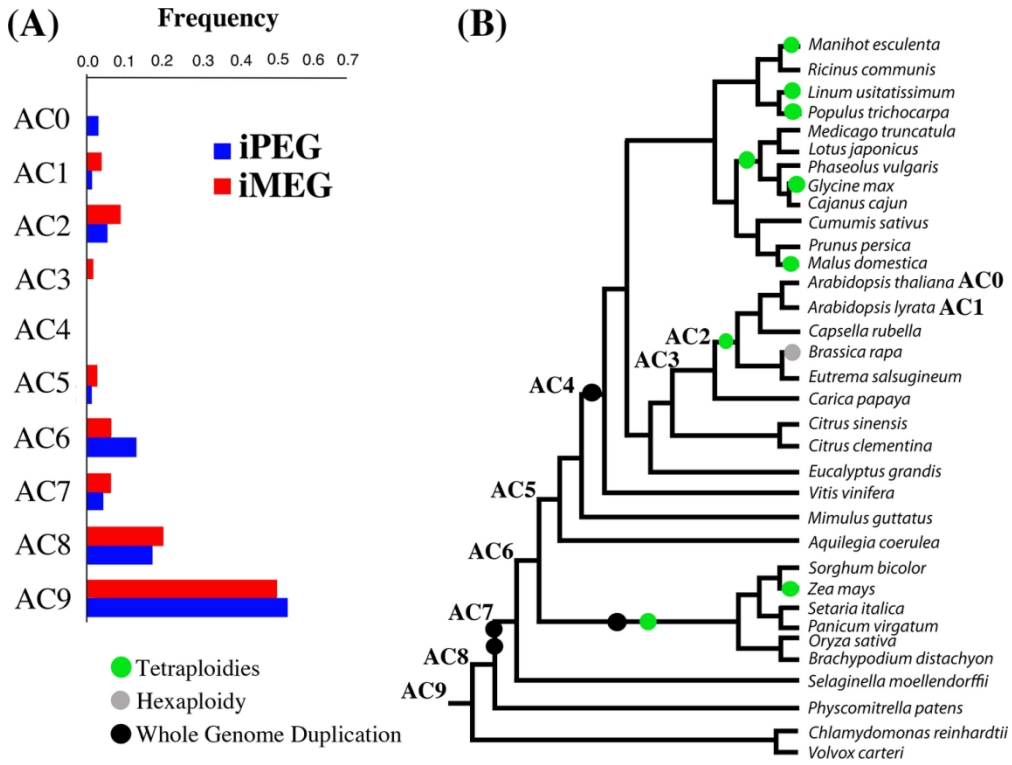


Figure 4

128x96mm (300 x 300 DPI)

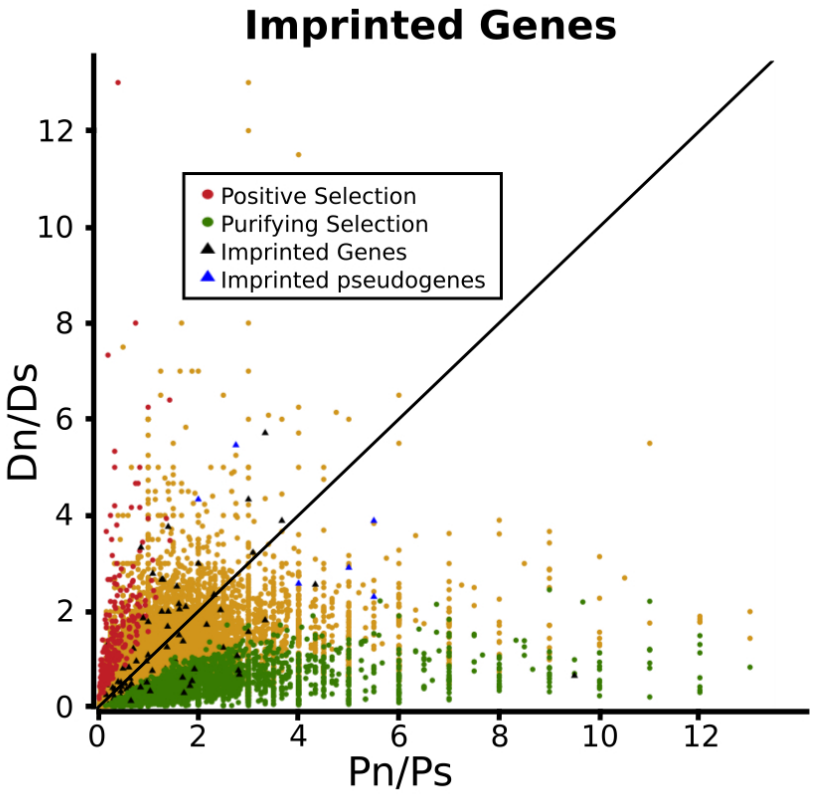

Imprinted Genes
Positive Selection
Purifying Selection
Imprinted Genes
Imprinted pseudogenes
Dn/Ds
Pn/Ps
0
2
4
6
8
10
12

Figure 5