OXFORD

Briefings in Functional Genomics, 2026, **25**, elaf025
https://doi.org/10.1093/bfgp/elaf025
Review Paper

Bioinformatics insights into plant genomic imprinting: approaches, challenges, and future perspectives

Xiaotong Jing [1,2], Xi Su [3], Quan Zou [1,2], Mengting Niu [1,4,*]

[1]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, No. 2006 Xiyuan Avenue, West Hi-Tech Zone, Chengdu 611731, China
[2]Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, No. 1 Chengdian Road, Kecheng District, Quzhou 324003, China
[3]The Affiliated Foshan Women and Children Hospital, Guangdong Medical University, No. 11 Renminxi Road, Chancheng District, Foshan 528000, China
[4]School of Electronic and Communication Engineering, Shenzhen Polytechnic University, No. 7098 Liuxian Avenue, Nanshan District, Shenzhen 518055, China

*Corresponding author. Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China. E-mail: niumt@uestc.edu.cn

Abstract

Genomic imprinting is an epigenetic occurrence that results in the expression of alleles specific to the parent of origin, plays pivotal roles in plant development, stress adaptation, and agronomic trait regulation. While imprinting has been intensively investigated in model plants (e.g. *Arabidopsis*, maize, and rice), its dynamic regulatory mechanisms and evolutionary implications remain enigmatic. Recent advances in bioinformatics—including single-cell omics, machine learning, and deep learning—have revolutionized the identification, functional annotation, and network modeling of imprinted genes. This review not only provides a detailed summary of the identification, functions and regulatory mechanisms of plant imprinted genes, but also systematically summarizes methodologies for studying plant genomic imprinting, highlights challenges in multi-omics data integration, and envisions artificial intelligence–driven strategies for epigenetic breeding.

Keywords: genomic imprinting; bioinformatics; epigenetics; multi-omics integration; crop epigenomics

Introduction

Genomic imprinting serves as a cornerstone of epigenetic regulation, where alleles are expressed in a distinct manner depending on their parental origin. This particular phenomenon is not common in eukaryotes and is solely detected in mammals, flowering plants, and certain insects [1]. In mammals, the imprinting phenomenon happens in embryos, placentas, and adult tissues [2], while in flowering plants, imprinting predominantly takes place in the endosperm, and a small quantity of imprinted genes have also been detected in embryos as well [3]. Studies have shown that after double fertilization, the interaction between the maternal and paternal genomes within the endosperm regulates seed development and is marked by genomic imprinting [4]. In plants, imprinted genes are predominantly implicated in seed development (e.g. endosperm cellularization) [5], environmental adaptation [6], and offering untapped potential for crop improvement [7].

Imprinted genes were first identified in maize through phenotypic recognition in plants. Kermicle [8] discovered that when R^rR^{8t} (female parent) was crossed with r^gr^g (male parent), the F1 generation seeds showed an almost equal number of solid-colored and spotted seeds; however, in the reciprocal cross ($r^gr^g \times R^rR^{8t}$), half of the F1 seeds had spots and the other half had weak spots. This phenotype of *R* was not caused by a dosage effect but was determined by its own genetic pattern. Besides phenotypic recognition, traditional experimental methods such as GeneCalling, Allelic Message Display, and low-throughput analysis of single-nucleotide polymorphism (SNP) allelic expression levels were also employed to identify imprinted genes [9]. Given the specificity of genomic imprinting occurring in the endosperm, only a small number of imprinted genes were identified for a long period.

In recent years, along with the progress of high-throughput sequencing technology, numerous imprinted genes have been

Xiaotong Jing is a postdoctoral researcher at the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China. Her research is centered on the epigenetic regulation of imprinted genes.
Xi Su is a chief physician at the Affiliated Foshan Women and Children Hospital, Guangdong Medical University. His primary research focuses on the maternal-fetal medicine.
Quan Zou is a professor at the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China. His research interests encompass machine learning and multiple sequence alignment.
Mengting Niu is an associate professor at the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China. She works on the deep learning and bioinformatics.

Received: November 5, 2025. **Revised:** December 9, 2025. **Accepted:** December 11, 2025

discovered in a relatively short time based on RNA sequencing (RNA-seq) analysis in plants [10]. Moreover, experimental verification using whole-genome bisulfite sequencing (WGBS) and chromatin immunoprecipitation sequencing (ChIP-seq) has gradually clarified the regulatory mechanism of imprinted genes. The formation of imprinting results from the disparities in epigenetic marks on alleles originating from parental gametes. These research results provide a research basis for imprinting omics in the study of grain development in food and oil crops, and they offer a technical foundation for overcoming interspecific hybridization and ploidy hybridization obstacles in horticultural crops. This review focuses on cutting-edge computational tools, databases, and predictive models that address these challenges, positioning bioinformatics as a catalyst for imprinting research.

Computational strategies for identifying imprinted genes

Allele-specific expression analysis

RNA-seq remains the gold standard for imprinting detection. In the realm of plants, the assessment of imprinting on a genome-wide scale typically involves using RNA-seq of tissues obtained from reciprocal crosses between two species or inbred lines. The selected species or lines should have the ability to interbreed, yet be distinct enough such that SNPs or other polymorphisms occur frequently. SNPs specific to a particular strain are utilized to determine the parental origin of reads, and these allele-specific reads are then counted for each gene (Fig. 1).

Based on these strain-specific SNPs and RNA-seq, hundreds of thousands of imprinted genes have been detected in diverse monocotyledonous and dicotyledonous plants, such as gramineaceous crops [9, 11–14], cruciferous crops [15–17], oil crops [18, 19], horticultural plants [5, 20–22], and other crops such as flax [23] (Fig. 2). In addition, some tools, like QuASAR [24], AlleleSeq [25], MBASED [26], and Ornaments [27], leverage parental SNPs to quantify allele-specific expression (ASE) ratios in hybrid offspring. In addition to SNP-based allele resolution, some statistical modeling approaches—such as Bayesian mapping methods [28, 29]—account for technical biases and low-coverage data. However, these tools and models are mostly used in the analysis of imprinting in animals. Although SNPs have been widely applied to identify ASE in research areas such as gene regulation and imprinted genes, they also have several limitations and drawbacks. For instance, if the density of SNPs in the genome is low or unevenly distributed, it may be difficult to accurately identify allele-specific expression. Also, RNA-seq data may contain technical noise (e.g. sequencing errors, amplification bias) and biological noise (e.g. transcript degradation), which can lead to false-positive or false-negative ASE signals.

Epigenomic feature integration

The mechanism of imprinted genes in plants bears resemblance to that in animals, primarily achieved through parental epigenetic asymmetries to ensure parent-specific expression. Among them, the asymmetries in DNA methylation between the parental genomes in the offspring are the main factor regulating the imprinting state of numerous genes [1, 30–32]. A simple mechanism of imprinting has been demonstrated in many plants, where maternal DNA demethylation occurs in the progeny endosperm while paternal DNA methylation is maintained [3, 11, 33–35]. These imprinted genes are significantly correlated with differentially methylated regions (DMRs). Notably, these DMRs are reliant on the parent of origin, and thus referred to as parent-of-origin-dependent DMRs (pDMRs). For example, the maternally expressed gene (MEG), *FLOWERING WAGENINGEN*, is imprinted by the activated maternal gametophyte-specific gene in the endosperm, a process that depends on a DNA glycosylase, *DEMETER* (*DME*) [36]. Likewise, DMRs have been identified in the promoter regions of the MEG, *FERTILIZATION-INDEPENDENT SEED2* (*FIS2*) [37], as well as in the 3′-downstream region of the paternally expressed gene (PEG), *PHERES1* (*PHE1*) [38].

Therefore, a genome-wide map of allele-specific DNA methylation with high resolution will be of great significance for achieving a deeper comprehension of the regulation of genetic imprinting. WGBS can be used not only to analyze methylation patterns but also to identify new imprinted genes. A variety of analysis tools have also been developed for WGBS, such as BSMAP [39], Bismark [40], BS-Seeker2 [41], and BiSulfite Bolt [42], *etc.* Among them, BSMAP demonstrated the fastest running speed and highest alignment quality in plant DNA methylation studies [43]. In the analysis of plant imprinted genes, combining parental SNP information with DMRs not only enables the identification of imprinted genes but also allows for a preliminary determination of their imprinting mechanisms [44].

In addition to the simplest imprinting model mentioned above, imprintome studies indicate that the imprinting of PEGs depends on parental DNA methylation asymmetries and histone marks, such as H3K9me2 and H3K27me3 (associated with transcriptional repression) and H3K4me3 and H3K36me3 (associated with transcriptional activation) [1, 45]. Among these histone modifications, H3K27me3 is considered a secondary imprint controlling parental gene expression asymmetries, with DNA methylation being the primary one [1]. In the endosperm of maize, PEGs are characterized by the presence of H3K27me3, H3K4me3, and H3K36me3 modifications altogether [45]. In addition, H3K4me3 and H3K36me3 are also significantly associated with MEGs in maize endosperm [45]. These histone modification markers are mainly identified *via* ChIP-seq, and SNPs are subsequently used to distinguish maternal/paternal signals of ChIP-seq. Currently, a variety of regulatory models for genome-imprinted genes have been reported, which range from simple to complex [1]. Therefore, researchers often combine histone marks with DMRs to identify correlations.

Machine learning and deep learning in imprinting

In the past several years, machine learning and deep learning have been utilized to multi-omics (including transcriptomics, metabolomics, proteomics, and epigenomics) analysis to effectively accelerate the process of plant breeding. For instance, dimensionality reduction algorithms (machine learning) are an essential step prior to the application of genome-wide association studies (GWASs) and genomic selection (GS) in plant germplasm research [46–48]. Linear Bayesian models (deep learning) have been used for complex trait genomic prediction in allopolyploid strawberry and autopolyploid blueberry [49]. DeepGS, a convolutional neural network-based approach, has demonstrated high predictive accuracy in inferring phenotypes from genotypes [50]. Moreover, machine learning and deep learning have shown superior performance in bioinformatics–epigenomics problems [51–54]. These problems mainly focus on chromatin accessibility [55, 56], DNA methylation predictions [57–62], biosequence function prediction [59, 63–65], and histone modification [66, 67]. To date, no studies have used machine or deep learning to predict imprinted genes in plants, but the

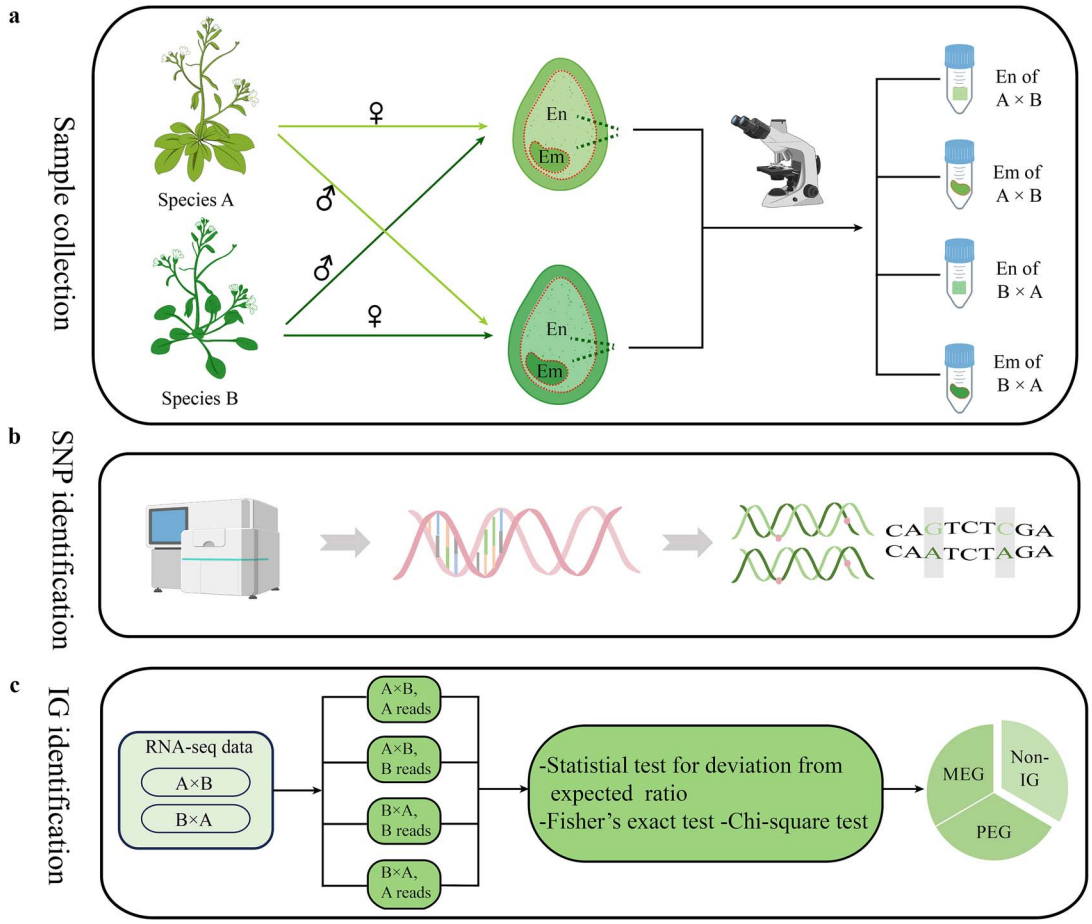

Figure 1. The pipeline for identifying imprinted genes based on RNA-seq and strain-specific SNPs in plants. (a). Two strains, A and B are carried out for reciprocal crosses. The developing seeds were harvested and the embryo and endosperm tissues (and sometimes the seed coat as well to reduce false positives) were dissected under a stereomicroscope. (b). SNPs between A and B were acquired by mapping the high-throughput sequencing data (either the genome resequencing of A and B or RNA-seq of reciprocal crosses between A and B) to the reference genome. (c). Utilizing RNA-seq and SNPs, the reads of A and B for each gene in both A × B and B × A were tallied. These counts were subsequently used to analyze the expression ratios of maternal and paternal alleles. Ultimately, a gene is classified as imprinted based on the parental allele expression deviation ratio in conjunction with statistical tests and predefined criteria. En: endosperm; Em: embryo; IG: imprinted gene; MEG: maternally expressed gene; PEG: paternally expressed gene.

solution of these epigenetic problems provides an indirect method for applying these approaches to the prediction of imprinted genes.

Mechanisms underlying imprinting genes in plants

Based on the research of imprinting mechanisms in plants, Batista and Kohler [1] proposed five regulatory models of imprinted genes. Among them, the imprinting regulatory mechanism of MEGs is divided into two types: DNA glycosylase DME-dependent and DME-independent (Fig. 3a, b). In the DME-dependent type, DNA methylation marked in both the maternal allele and paternal allele, therefore which been silenced in the central cells and sperm cells respectively. However, in the endosperm, for maternal genes to be expressed, maternal DNA methylation was removed by DME, while paternal DNA methylation has to be maintained by METHYLTRANSFERASE 1 [36]. In addition, research shows that H3K36me3 and H3K4me3 promoted the expression of maternal genes in the maize [45]. In the DME-independent model of MEG, maternal-specific expression is attained by silencing the paternal allele. This process might potentially be mediated by the RNA-directed DNA methylation activity in pollen. The establishment of imprinting in PEGs involves distinct mechanisms, which are mainly regulated by the parental DNA methylation asymmetry and H3K27me3. Here, we outline two classical regulatory models of PEGs. In the context of DME involvement, maternal-specific demethylation facilitated by DME enables the deposition of H3K27me3 in these alleles. As a result, their transcriptional activity is inactivated within the endosperm. In contrast, the existence of DNA methylation in paternal alleles hinders the deposition of H3K27me3 by FIS–PRC2, thereby permitting the transcription of this allele [68, 69]. Eventually, the gene show imprinting with paternal allele expressed in the endosperm. Moreover, H3K4me3 and H3K36me3 deposit in both the maternal allele and paternal allele in the somatic tissues of maize [45]. In central cells, demethylation promotes the association of the PRC2. This might attract H3K4me3 and H3K36me3 demethylases to remove H3K4me3 and H3K36me3 (Fig. 3c). In another PEG model, maternal allele and paternal allele were marked by the silencing factor, H3K27me3. However, the reduced activity of the PRC2 in sperm cells results in the removal of H3K27me3, thereby triggering the transcriptional activation of paternal alleles. And H3K4me3 and H3K36me3 modifications tend to be deposited on paternal chromosomes that lack H3K27me3 modifications in maize [45] (Fig. 3d).

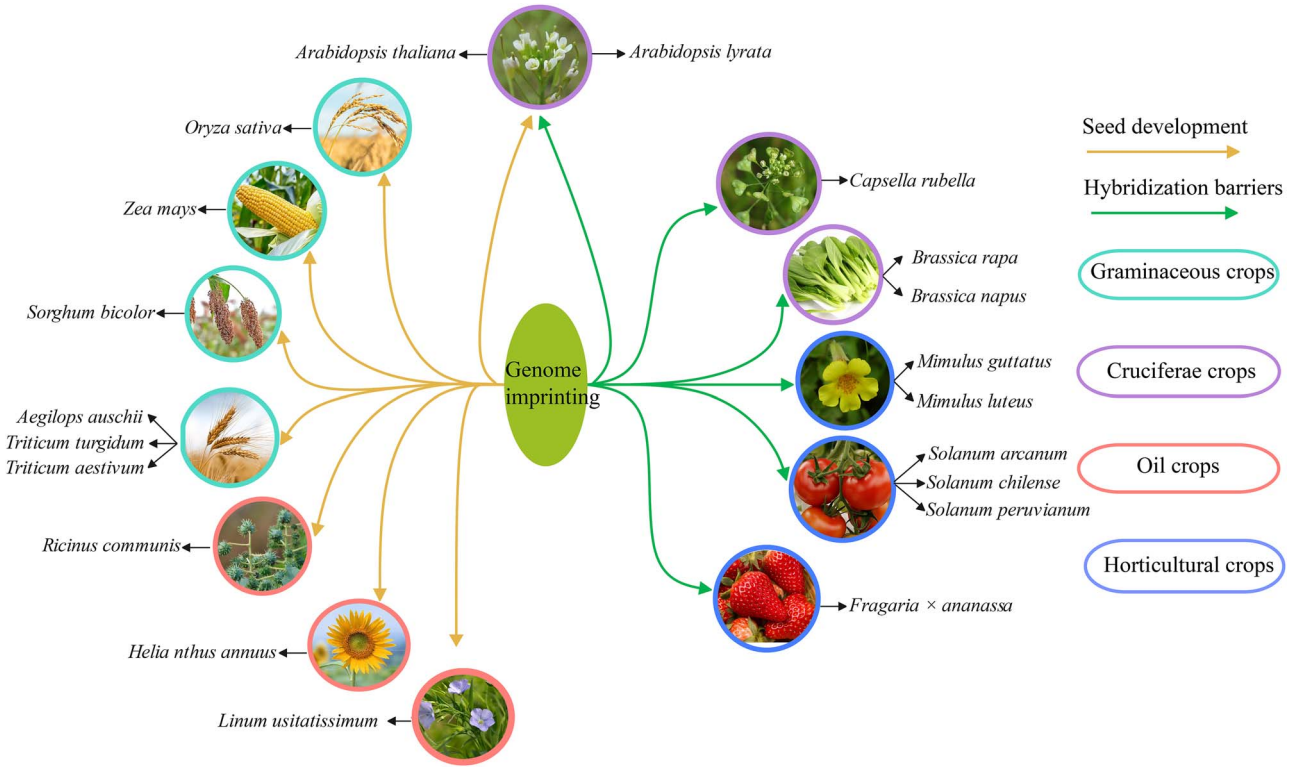

Figure 2. The imprinted genes of species were identified based on RNA-seq.

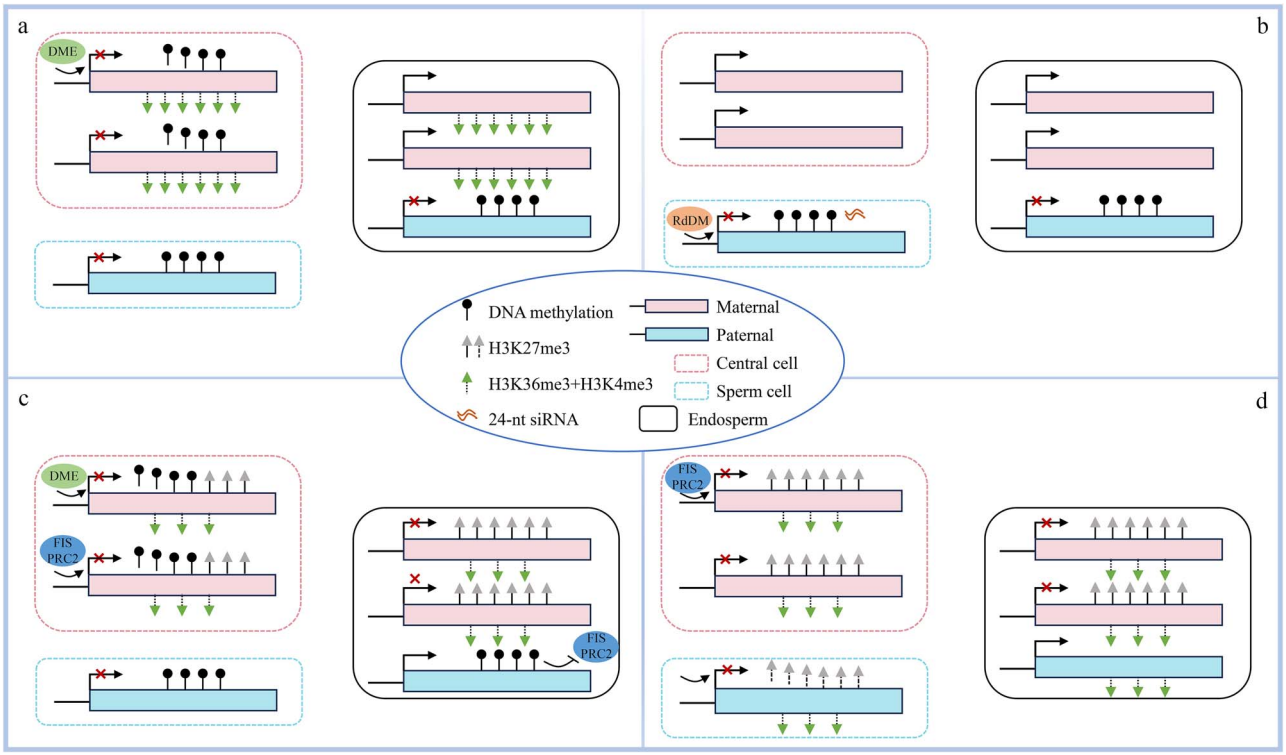

Figure 3. Models of epigenetic regulation in MEGs (a, b) and PEGs (c, d) in plants.

In addition to H3K27me3, methylation at other sites of the H3 histone also regulates the occurrence of imprinting. For example, MEGs are strongly associated with CG_pDMRs, and the target sites of mH3K4me3 and mH3K36me3 in maize endosperm [45] (Fig. 3a). It is possible that the maternal allele of MEGs undergoes DNA demethylation and acquires activating histone modifications, such as H3K4me3 and H3K36me3, thus being transcribed, while the paternal allele does not have these target sites and remains silent. Furthermore, sRNAs are of great significance in the epigenetic regulation of imprinting. Erdmann *et al.* [70] found

that disruption of the sRNA pathway mediated by *RNA polymerase IV* (*PolIV*) led to a global bias towards the maternal genome in the endosperm transcriptome. Mutations in the paternal *PolIV* caused an excessive paternal genome dosage within the endosperm, ultimately leading to seed abortion. This reveals the role of the *PolIV*-mediated sRNA pathway in regulating the global transcriptional balance between the maternal and paternal genomes in the endosperm. In rice, it was found that the 24-nt sRNAs of imprinted expression were transcribed from different parental alleles of imprinted genes, and they regulated the expression of MEGs by controlling the differential methylation of parental alleles [71] (Fig. 3b). Similarly, in *Arabidopsis*, it was found that paternal epigenetic-related siRNAs established 'endosperm balance' by enhancing sensitivity to increased paternal genome dosage, thereby avoiding hybridization barriers and maintaining genome stability and seed viability in the endosperm after fertilization [72].

Moreover, recent studies have demonstrated distinct epigenetic status in the rice endosperm between persistent and stage-specific imprinted genes. Specifically, the paternal alleles of persistent MEGs exhibit higher CG and CHG methylation levels compared to those of stage-specific MEGs. Furthermore, persistent PEGs show greater H3K27me3 accumulation than stage-specific PEGs [12]. These findings contribute to a more comprehensive understanding of the regulatory mechanisms underlying genomic imprinting. However, deciphering imprinting mechanisms faces unique hurdles, such as tissue specificity—imprinting often occurs transiently in specialized tissues (e.g. endosperm), necessitating high-resolution spatiotemporal omics data; species divergence—imprinting patterns vary significantly between monocots and dicots, complicating cross-species generalization; and multi-layered regulation—imprinted expression is orchestrated by DNA methylation, histone modifications, and non-coding RNAs, requiring integrative bioinformatics frameworks.

Challenges and bioinformatics solutions

Accurate identification of imprinted genes

Imprinted genes are often expressed only in specific tissues or developmental stages. The endosperm serves as the main tissue in which gene imprinting takes place, followed by the embryo. During the collection of endosperm and embryo samples, the seed coat surrounding the seed tissue may serve as a source of contamination, potentially leading to false-positive MEG identifications, as the seed coat is a maternal tissue [73]. Consequently, minimizing contamination from maternal tissues has become a critical challenge in such studies. Furthermore, species with complex genetic architectures—such as octoploid cultivated strawberries—or those with low levels of homozygosity pose significant analytical challenges that can compromise the accuracy of results. Moreover, the quality of reference genome assembly plays a decisive role in RNA-seq data analysis and SNP identification. Through high-throughput RNA-seq technology, genome-wide studies in the hybrid endosperms of *Arabidopsis*, rice, and maize have detected several hundred imprinted genes. However, the real number of imprinted genes is likely to have been significantly underestimated, as most of these experiments only covered one developmental stage [9, 16, 74–77]. For instance, B73 and Mo17 are classical strains for identifying the imprintome in maize. During the development of maize endosperm, imprinted genes show dynamic expression. The majority of PEGs were observed at 7 days after pollination (DAP) in the B73 and Mo17 reciprocal crosses. Conversely, ~80% of the MEGs were specific to 10 DAP. This was chiefly because their expression levels increased significantly compared to other stages [78]. Moreover, in the studies of *Arabidopsis*, rice, and maize, only a small fraction of the imprinted genes was commonly recognized. This is probably due to diverse ecotypes, different developmental stages, and the bioinformatic criteria for defining imprinted genes [9, 16, 74, 75]. In addition, hybrid experiments are required to isolate parental allele-specific expression; however, hybridization is challenging in some plants. Studies have shown that imprinting loss is directly related to intraspecific, interspecific, and ploidy hybridization failure [4, 15, 79, 80]. These hybrid seed failures are considered postzygotic barriers. However, imprinting is difficult to study in prezygotic post-pollination barriers of plants. Cross incompatibility and unilateral cross incompatibility have been documented in numerous plants, such as tomato [81], *Cucumis* [82], and maize [83]. Pollen is rejected during the pollen–pistil interaction in these crosses. Furthermore, allelic expression differences may arise from genetic variation or stochastic expression fluctuations, necessitating differentiation from true imprinting effects.

Complexity of epigenetic mechanisms

Imprinting may involve DNA methylation, histone modifications, non-coding RNAs, and other mechanisms. A large number of candidate imprinted genes were identified by RNA-seq analysis based on parental expression differences, but it was impossible to distinguish which were imprinted genes and which were maternal effect genes. Therefore, integrating data from different omics levels (such as genome, transcriptome, epigenome, *etc.*) for analysis aims to fully understand the interrelationships between different molecular levels in organisms and their contributions to genomic imprinting phenomena. However, due to the large differences in the generation, processing, and analysis of different omics data, how to efficiently integrate these multi-source data has become a challenge in bioinformatics. For instance, there are significant differences in the measurement scale, data type, and quality control of various omics data, and how to combine these different sources of data is still a technical problem. Moreover, in multi-omics data, missing values and noise are unavoidable. An important issue is how to handle missing data and eliminate noise to guarantee the quality of integrated data. In addition, although some tools and algorithms can be used for multi-omics data integration, there are still some limitations such as high computational demands, low operating efficiency, and poor interpretability. At present common methods of multi-omics data integration mainly include data standardization, principal component analysis, joint analysis, and network analysis. In recent years, machine learning has been widely used to integrate and analyze multi-omics data. For example, machine learning can automatically learn underlying patterns and patterns from complex multi-omics data to help discover new biomarkers, predict diseases, and more [84, 85].

Functional validation and phenotypic linking

Imprinted genes possess a diverse array of functions and have been intensively studied in mammals. These genes frequently appear in clusters on autosomes and sex chromosomes, and they regulate the growth and development of animals, biological evolution, genetic diseases, and the occurrence of tumors by regulating the growth and development of embryos and fetuses after birth. For example, *XIST* is a PEG located on the sex chromosome X. If this gene is abnormally expressed on the maternal X chromosome, the maternal X chromosome will be inactivated, resulting in abnormal early embryonic development [86].

At present, only a few functions of imprinted genes have been clarified in plants. Some of these are closely related to seed development, and their mutation can lead to embryo abortion or delayed development, seed reduction, and endosperm defects, such as *MEDEA* (*MEA*) [87], *MULTICOPY SUPRESSOR OF IRA1* (*MSI1*) [88], *FIS2* [37], *MATERNALLY EXPRESSED PAB C-TERMINAL* (*MPC*) [89], *FH5* [90], *LORELEI* (*LRE*) [91], and *NUWA* [92] in *Arabidopsis*; *FIE1* in rice [93], *maternally expressed gene 1* (*meg1*) in maize [94], *etc.* (Table 1). Among these imprinted genes, *MEA*, *FIS2*, and *FIE* also act as important PcG chromosomal remodeling factors by mediating H3K27me3 to bind the maternal allelic hypomethylation region of *PHE1*, which is a PEG in *Arabidopsis*. The nearby histones are methylated to inhibit the expression of maternal alleles, resulting in the paternal imprinting state of *PHE1* [38]. This pattern of methylation is similar to the paternal imprinting mechanism of *Upward Curly Leaf 1* (*UCL1*), in which the maternal allele of *UCL1* is silenced by the FIS2-PRC2 complex in the endosperm before and after fertilization [95]. It was found that overexpression of *UCL1* not only made the leaf curl upward and flower early but also reduced the level of Curly Leaf (CLF) protein and changed the H3K27me3 chromatin marker of the *CLF* target gene [96]. In addition, the loss of imprinting is directly related to the failure of intraspecific, interspecific, and ploidy hybridization [4, 15, 79, 80]. In the hybridization of *Arabidopsis thaliana* Col-0 and C24, the loss of eight PEG imprinting states, such as *PHE1* and *HDG3*, played a fatal role in hybrid incompatibility [79]; in the reciprocal crosses of wild tomatoes *Solanum peruvianum* and *S. chilense*, when *S. peruvianum* was the maternal parent, PEGs were almost completely eliminated in the endosperm genome, and the change in the paternal expression ratio caused endosperm development failure, leading to hybrid seed abortion [80]. In the genus *Cardamine*, the number and effective ploidy of PEGs decrease with the selfing history of the species, and these PEGs involved in hybridization barriers are related to the silencing marks of transposable elements and CHH methylation [15, 97]. *Arabidopsis*, rice, and maize are the model plants for studying plant gene function, and these species all have mature transgenic technology and the clustered regularly interspaced short palindromic repeats (CRISPR) technology to verify gene function. However, transgenic and CRISPR efficiency varies across plant species, and epigenetic modifications may lack heritable stability. In addition, imprinted genes often function within regulatory networks, leading to subtle phenotypes in single-gene knockouts.

Future directions: from discovery to design

Diverse methodologies for identifying imprinted genes

In the past five years, high-throughput sequencing technologies at single-cell resolution have been rapidly applied to study plant tissues. Among them, single-cell RNA sequencing (scRNA-seq) has been most widely used, followed by single-nucleus RNA sequencing (snRNA-seq), single-cell/nucleus assay for transposase accessible chromatin sequencing (scATAC-seq/snATAC-seq), and spatial transcriptome technologies [109]. Based on scRNA-seq, atlases of diverse tissues in *Arabidopsis* were constructed, including the root [110], vegetative shoot apex [111], leaf vasculature [112], and sperm cells in pollen [113], *etc.* In addition, a comprehensive map of transcriptional complexity in the endosperm of developing *Arabidopsis* seeds was built by snRNA-seq [114]. It showed that imprinting exhibits spatial and temporal heterogeneity. Specifically, the extent of parental bias in MEGs varies with the cell-cycle phase, and imprinting of PEGs is strongest in the chalazal endosperm [114, 115]. In addition, imprinted genes were also identified through the integration of methods of isolating nuclei tagged in specific cell types (INTACT) and high-throughput chromatin conformation capture (Hi-C) [116]. The Plant Cell Atlas was launched and implemented in 2021; it is devoted to addressing unresolved questions in plant science, such as plant development, response to the environment, and cellular evolution [117]. Based on single-cell omics platforms, a 4D developing representation that shows the growth of a plant from root to fruit will be created in the future [117, 118]. Therefore, single-cell multi-omics profiling, coupling epigenomic, transcriptomic, and proteomic data at cellular resolution, offers a new approach to dissect the imprinting establishment [119, 120].

The surge in omics data poses a challenge to the long-standing methodologies for data analysis. Statistics and artificial intelligence (AI) are both used for prediction and inference. In general, traditional statistical methods are based on certain assumptions regarding the data-generating systems. For instance, the endosperm exhibits a maternal/paternal (m:p) genome dosage ratio of 2 m:1p in most angiosperms; therefore, existing methods for evaluating imprinting take into account this maternal and paternal dosage in the endosperm [114]. AI focuses on the development and utilization of algorithms that enhance performance as they gain experience. A variety of machine learning methods are capable of generating models for pattern recognition, classification, and prediction from available data, without depending on strict assumptions regarding data-generating systems [121]. For example, based on machine learning algorithms and attribute weighting methods, researchers found that several factors were the most crucial attributes for differentiating paternal and maternal imprinted genes in species such as cattle, humans, and sheep. These factors included GC content within 100 kb upstream of genes, the existence of LINE in that particular region, and the frequency of amino acids such as Gly, Gln, and Met [122]. As the research on plant genomic imprinting lags behind that on animals, there are currently few studies using AI to predict plant imprinted genes.

The biological significance of genomic imprinting

Genomic imprinting is a genetic phenomenon that does not conform to the classical Mendelian theory. Based on the occurrence of this special phenomenon, researchers have proposed various hypotheses from different perspectives. Among them, the most mainstream one is the parental conflict theory [123]. In angiosperms, seeds require nutrients from the mother from fertilization to maturity. This theory holds that paternal gene expression aims to maximize the transmission of maternal nutrients to the offspring, while maternal gene expression aims to increase the possibility of reproductive success by limiting the maximization of maternal nutrient flow to ensure a reasonable distribution of nutrients to more offspring, thereby increasing seed set. Therefore, the parental conflict theory predicts that under adverse conditions, MEGs play a role in reducing seed size and seed number, while PEGs have the opposite effect. This prediction has been verified in both plants and mammals. In plants, maternal imprinted genes *MEA* and *FIS2* have defects in excessive reproduction during seed development [37, 87]. The effects of interspecific hybridization between plants with different ploidies on endosperm development and seed morphology are also consistent with this theory. Ploidy hybridization found that an excess of maternal genome results in slow endosperm development and smaller seeds, while an excess of paternal genome leads to an extended endosperm proliferation period and larger seeds

Table 1. The information of known imprinted genes in the model plants.

Gene	Full Name	Epigenetic Mark	Annotation	Mutant Phenotype	Reference
Arabidopsis					
ADM[P, a]	*ADMETOS*	Not known	J-domain family	Seed abortion	[98]
FH5[M, a]	*FORMIN HOMOLOGUE 5*	H3K27me3	Actin nucleator	Endosperm posterior pole structures are defective	[90]
FIS2[M, a]	*FERTILIZATION INDEPENDENT SEED 2*	DNA-me	Subunit of the FIS–PRC2 complex	Seed abortion	[37]
FWA[M, a]	*FLOWERING WAGENINGEN*	DNA-me	HD-ZIP transcription factor	Late flowering	[36]
HDG3[P, a]	*HOMEODOMAIN GLABROUS 3*	DNA-me	HD-ZIP transcription factor	Seed weight and area are slightly reduced	[44]
HDG8[M, a]	*HOMEODOMAIN GLABROUS 8*	DNA-me	HD-ZIP transcription factor	No visible phenotype	[44]
HDG9[M, a]	*HOMEODOMAIN GLABROUS 9*	DNA-me	HD-ZIP transcription factor	No visible phenotype	[44]
LRE[M, a]	*LORELEI*	Not known	Glycosylphosphatidylinositol—123 anchored membrane protein	Delayed seed development	[91]
MEA[M, a]	*MEDEA*	H3K27me3, DNA-me	Subunit of the FIS–PRC2 complex	Seed abortion	[87]
MPC[M, a]	*MATERNALLY EXPRESSED PAB C-TERMINAL*	DNA-me	Poly(A) binding C-terminal domain	Small seeds, and abnormal embryo	[89]
MYB3R2[M, a]	*MYB3R2*	DNA-me	MYB transcription factor	Not known	[44]
NRPD1A[P, a]	*Nuclear RNA polymerase D1A*	Not known	Subunit of RNA POLYMERASE IV	Seed abortion	[70, 72]
NUWA[M, b]	*NUWA*	Not known	Mitochondrial-localized protein	Seed abortion	[92]
PEG2[P, a]	*PEG2*	Not known	Unknown protein	Seed abortion	[98, 99]
PHE1[P, a]	*PHERES1*	H3K27me3, DNA-me	MADS-box transcription factor	No visible phenotype	[38]
PKR2[P, a]	*PICKLE RELATED 2*	H3K27me3	CHD3 chromatin remodeler	Seed size is slightly reduced	[100]
SUVH7[P, a]	*SU(VAR)3–9 HOMOLOG 7*	Not known	Putative histone-lysine N-methyltransferase	Seed abortion	[98]
UCL1[P, a]	*UPWARD CURLY LEAF 1*	H3K27me3, DNA-me	E3 ligase	Not known	[95]
Maize					
dzr1[M, a]	*dzr1*	Not known	Regulate the accumulation of 10-kDa zein	Not known	[101]
fie1[M, a]	*FERTILIZATION INDEPENDENT ENDOSPERM 1*	H3K27me3, DNA-me, H3/H4-Ac	Homolog of *AtFIE*	Not known	[102]
fie2[M, a]	*FERTILIZATION INDEPENDENT ENDOSPERM 2*	DNA-me	Homolog of *AtFIE*	Not known	[102]
Mee1[M, c]	*maternally expressed in embryo 1*	DNA-me	Conserved protein	Not known	[103]
meg1[M, a]	*maternally expressed gene 1*	DNA-me	Cysteine-rich polypeptide	Small seeds	[94]
Mez1[M, a]	*maize Enhancer of zeste1*	DNA-me	Homolog of *AtMEA*	Not known	[104]
nrp1[M, a]	*no-apical-meristem related protein1*	H3K27me3, DNA-me, H3/H4-Ac	Transcription factor	Not known	[105]
peg1[P, a]	*paternally expressed gene 1*	Not known	Not known	Not known	[101, 106]
R[M, a]	*R*	Not known	Transcription factor	Not known	[8]
tubα3[M, a]	*α-tubulin 3*	DNA-me	Tubulin	Not known	[107]
tubα4[M, a]	*α-tubulin 4*	DNA-me	Tubulin	Not known	[107]
Zein[M, a]	*zein*	DNA-me	Zein polypeptide	Not known	[108]
Rice					
FBDUF48[P, a]	*FBDUF48*	DNA-me	DUF295-domain protein	Small seeds	[11]
FBX365[P, a]	*FBX365*	DNA-me	F-box domain protein	Seeds showed chalkiness	[11]
FIE1[M, a]	*FERTILIZATION INDEPENDENT ENDOSPERM 1*	DNA-me	Homolog of *AtFIE*	No visible phenotype	[93]
MEG2[M, a]	*MEG2*	DNA-me	Unknown protein	Width and length of seeds are reduced	[11]
MEG3[M, a]	*MEG3*	DNA-me	Unknown protein	Weight and thickness of seeds are reduced	[11]
PEG1[P, a]	*PEG1*	DNA-me	Oxygenase	Small seeds	[11]

Note: [M]MEG. [P]PEG. [a]the genes showed imprinting in the endosperm. [b]the genes were demonstrated to be imprinted within the embryo. [c]the genes were imprinting in both the embryo and endosperm.

[80]. Whole-genome imprinting analysis of different species has revealed an imbalance in the number of MEGs and PEGs, with a significant difference in quantity. Based on this phenomenon, researchers have proposed the maternal-offspring coadaptation theory, suggesting that the imprinted expression of maternal alleles is a result of natural selection that benefits the health of the offspring, providing the greatest fitness for the offspring and the mother in their mutual adaptation [124]. The research by Raissig *et al.* [125] provides support for this hypothesis. The study found that MEGs in *A. thaliana* embryos function in maternal tissues (seed coat) and at the junction of the embryo and maternal tissues (basal endosperm and hypocotyl), possibly by mediating metabolic connections between the seed coat and the embryo, thereby keeping the embryo's genes under maternal control. Furthermore, some studies suggest that genomic imprinting results from the genome's defensive reaction against the invasion of exogenous genetic fragments, such as transposons, thus proposing the host-defense hypothesis [126]. Research has found that the establishment of initial imprinting is closely related to transposons, and many imprinted genes in plants are located in transposon and repetitive sequence regions [3, 127]. Among them, transposon movement can generate new imprinted genes [75]; transposon deletion can lead to the loss of imprinting [127]; transposon insertion can also disrupt the imprinting status, causing imprinted genes to show biallelic expression [128]; however, the generation of imprinting due to transposon insertion has not been fully confirmed yet. Currently, it has only been found in maize that the insertion of *Mutator* transposons imprints two potential MEGs [129]. Based on the above, plant genomic imprinting can be studied from an evolutionary systems biology perspective using phylogenomic frameworks to trace its origin and understand its role in plant adaptations.

Key Points

- Plant genomic imprinting research is transitioning from mechanistic exploration to bioinformatics-driven innovation.
- By harnessing multi-omics integration, machine learning, and collaborative data sharing, the field is poised to unlock novel breeding strategies and address global food security challenges.
- Future efforts must prioritize scalable algorithms, species-tailored pipelines, and ethical artificial intelligence applications to fully realize the promise of epigenetic agriculture.

Acknowledgements

The authors thank Pinyu Zhu for providing some suggestions on the manuscript.

Author contributions

Xiaotong Jing (Conceptualization, Writing—original draft), Xi Su (Conceptualization, Writing—review & editing), and Quan Zou (Conceptualization, Writing—review & editing), Mengting Niu (Writing—review & editing). All authors have read and agreed to the published version of the manuscript.

Conflict of interest

Authors declare no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China (No. 62425107, 62402088, 62473268, 32270786), the Key Science and Technology Research Project of Foshan (No. 2120001008276), the Foshan Birth Defects Prevention and Control Engineering Technology Research Center (No. 2420001000293), the Municipal Government of Quzhou (No. 2024D042), and Shenzhen Science and Technology Program (No. RCBS20231211090800004).

Data availability

Not applicable.

References

1. Batista RA, Kohler C. Genomic imprinting in plants-revisiting existing models. *Genes Dev* 2020;**34**:24–36. https://doi.org/10.1101/gad.332924.119.
2. Pires ND, Grossniklaus U. Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000prime Rep* 2014;**6**:63. https://doi.org/10.12703/P6-63.
3. Pignatta D, Erdmann RM, Scheer E. *et al.* Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife* 2014;**3**:e03198. https://doi.org/10.7554/eLife.03198.
4. Kinser TJ, Smith RD, Lawrence AH. *et al.* Endosperm-based incompatibilities in hybrid monkeyflowers. *Plant Cell* 2021;**33**:2235–57. https://doi.org/10.1093/plcell/koab117.
5. Joldersma D, Guo L, Alger EI. *et al.* Identification and analysis of imprinted genes in wild strawberry uncover a regulatory pathway in endosperm development. *Plant Physiol* 2024;**196**:2599–613. https://doi.org/10.1093/plphys/kiae496.
6. Torres JR, Botto JF, Sanchez DH. Canonical transcriptional gene silencing may contribute to long-term heat response and recovery through MOM1. *Plant Cell Environ* 2024;**47**:372–82. https://doi.org/10.1111/pce.14722.
7. Zhang H, Zhu JK. Epigenetic gene regulation in plants and its potential applications in crop improvement. *Nat Rev Mol Cell Biol* 2025;**26**:51–67. https://doi.org/10.1038/s41580-024-00769-1.
8. Kermicle JL. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* 1970;**66**:69–85. https://doi.org/10.1093/genetics/66.1.69.
9. Zhang M, Zhao H, Xie S. *et al.* Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A* 2011;**108**:20042–7. https://doi.org/10.1073/pnas.1112186108.
10. Sudarsanan SKV, Sreekumar N. Bioinformatic resources for plant genomic research. *Curr Bioinf* 2024;**19**:513–29. https://doi.org/10.2174/1574893618666230725123211.
11. Yuan J, Chen S, Jiao W. *et al.* Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol* 2017;**216**:373–87. https://doi.org/10.1111/nph.14510.
12. Tonosaki K, Susaki D, Morinaka H. *et al.* Multilayered epigenetic control of persistent and stage-specific imprinted genes in rice endosperm. *Nat Plants* 2024;**10**:1231–45. https://doi.org/10.1038/s41477-024-01754-4.
13. Yang G, Liu Z, Gao L. *et al.* Genomic imprinting was evolutionarily conserved during wheat polyploidization. *Plant Cell* 2018;**30**:37–47. https://doi.org/10.1105/tpc.17.00837.

14. Zhang M, Li N, He W. *et al.* Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J* 2016;**85**:424–36. https://doi.org/10.1111/tpj.13116.
15. Lafon-Placette C, Hatorangan MR, Steige KA. *et al.* Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. *Nat Plants* 2018;**4**:352–7. https://doi.org/10.1038/s41477-018-0161-6.
16. Hsieh TF, Shin J, Uzawa R. *et al.* Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc Natl Acad Sci U S A* 2011;**108**:1755–62. https://doi.org/10.1073/pnas.1019273108.
17. Yoshida T, Kawanabe T, Bo Y. *et al.* Genome-wide analysis of parent-of-origin allelic expression in endosperms of Brassicaceae species. *Brassica rapa Plant Cell Physiol* 2018;**59**:2590–601. https://doi.org/10.1093/pcp/pcy178.
18. Han B, Li Y, Wu D. *et al.* Dynamics of imprinted genes and their epigenetic mechanisms in castor bean seed with persistent endosperm. *New Phytol* 2023;**240**:1868–82. https://doi.org/10.1111/nph.19265.
19. Zhang Z, Yu S, Li J. *et al.* Epigenetic modifications potentially controlling the allelic expression of imprinted genes in sunflower endosperm. *BMC Plant Biol* 2021;**21**:1–11. https://doi.org/10.1186/s12870-021-03344-4.
20. Roth M, Florez-Rueda AM, Paris M. *et al.* Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. *Plant J* 2018;**95**:1084–101. https://doi.org/10.1111/tpj.14012.
21. Flores-Vergara MA, Oneal E, Costa M. *et al.* Developmental analysis of *mimulus* seed transcriptomes reveals functional gene expression clusters and four imprinted, endosperm-expressed genes. *Front. Plant Sci* 2020;**11**:132. https://doi.org/10.3389/fpls.2020.00132.
22. Povilus RA, Martin CA, Bechen LL. *et al.* Imprinting and DNA methylation in water lily endosperm: implications for seed evolution. *Mol Biol Evol* 2025;**42**:271.
23. Jiang H, Guo D, Ye J. *et al.* Genome-wide analysis of genomic imprinting in the endosperm and allelic variation in flax. *Plant J* 2021;**107**:1697–710. https://doi.org/10.1111/tpj.15411.
24. Harvey CT, Moyerbrailean GA, Davis GO. *et al.* QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics* 2014;**31**:1235–42.
25. Rozowsky J, Abyzov A, Wang J. *et al.* AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol* 2011;**7**:522. https://doi.org/10.1038/msb.2011.54.
26. Mayba O, Gilbert HN, Liu J. *et al.* MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome Biol* 2014;**15**:1–21. https://doi.org/10.1186/s13059-014-0405-3.
27. Adduri A, Kim S. Ornaments for efficient allele-specific expression estimation with bias correction. *Am J Hum Genet* 2024;**111**:1770–81. https://doi.org/10.1016/j.ajhg.2024.06.014.
28. Runqing Y, Xin W, Yuehua C. Bayesian inference for genomic imprinting underlying developmental characteristics. *Brief Bioinform* 2012;**13**:555–68. https://doi.org/10.1093/bib/bbr079.
29. Zhang Q, Liu G. SNP-based computational analysis reveals recombination-associated genome evolution in humans. *Curr Bioinf* 2023;**18**:192–204. https://doi.org/10.2174/1574893618666221226142329.
30. Li Y, Zou Q, Dai Q. *et al.* Identifying the DNA methylation preference of transcription factors using ProtBERT and SVM. *PLoS Comput Biol* 2025;**21**:1012513. https://doi.org/10.1371/journal.pcbi.1012513.
31. Wei R, Zhang L, Zheng H. *et al.* A systematic review of the application of machine learning in CpG Island (CGI) detection and methylation prediction. *Curr Bioinf* 2024;**19**:235–49. https://doi.org/10.2174/1574893618666230508104341.
32. Luo X, Zou Q. Identifying the "stripe" transcription factors and cooperative binding related to DNA methylation. *Commun Biol* 2024;**7**:1265. https://doi.org/10.1038/s42003-024-06992-y.
33. Peng X, Cui W, Zhang W. *et al.* A metric to characterize differentially methylated region sets detected from methylation array data. *Curr Bioinf* 2024;**19**:571–83. https://doi.org/10.2174/1574893618666230816141723.
34. Dao F, Xie X, Zhang H. *et al.* PlantEMS: a comprehensive database of epigenetic modification sites across multiple plant species. *Plant Commun* 2025;**6**:101228. https://doi.org/10.1016/j.xplc.2024.101228.
35. Jin J, Yu Y, Wang R. *et al.* iDNA-ABF: multi-scale deep biological language learning model for the interpretable prediction of DNA methylations. *Genome Biol* 2022;**23**:1–23. https://doi.org/10.1186/s13059-022-02780-1.
36. Kinoshita T, Miura A, Choi Y. *et al.* One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* 2004;**303**:521–3. https://doi.org/10.1126/science.1089835.
37. Jullien PE, Kinoshita T, Ohad N. *et al.* Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *Plant Cell* 2006;**18**:1360–72. https://doi.org/10.1105/tpc.106.041178.
38. Batista RA, Moreno-Romero J, Qiu Y. *et al.* The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *Elife* 2019;**8**:e50541. https://doi.org/10.7554/eLife.50541.
39. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinf* 2009;**10**:232. https://doi.org/10.1186/1471-2105-10-232.
40. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 2011;**27**:1571–2. https://doi.org/10.1093/bioinformatics/btr167.
41. Guo W, Fiziev P, Yan W. *et al.* BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* 2013;**14**:774.
42. Farrell C, Thompson M, Tosevska A. *et al.* BiSulfite bolt: a bisulfite sequencing analysis platform. *GigaScience* 2021;**10**:33. https://doi.org/10.1093/gigascience/giab033.
43. Wu Q, Yang M, Yang Y. *et al.* Assessment of bisulfite sequencing alignment tools for whole genome analysis in plants. *Int J Biol Macromol* 2025;**305**:140940. https://doi.org/10.1016/j.ijbiomac.2025.140940.
44. Gehring M, Bubb KL, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 2009;**324**:1447–51. https://doi.org/10.1126/science.1171609.
45. Dong X, Zhang M, Chen J. *et al.* Dynamic and antagonistic allele-specific epigenetic modifications controlling the expression of imprinted genes in maize endosperm. *Mol Plant* 2017;**10**:442–55. https://doi.org/10.1016/j.molp.2016.10.007.
46. Wang T, Yan Z, Zhang Y. *et al.* postGWAS: a web server for deciphering the causality post the genome-wide association studies. *Comput Biol Med* 2024;**171**:108108. https://doi.org/10.1016/j.compbiomed.2024.108108.
47. Alemu A, Åstrand J, Montesinos-López OA. *et al.* Genomic selection in plant breeding: key factors shaping two decades of progress. *Mol Plant* 2024;**17**:552–78. https://doi.org/10.1016/j.molp.2024.03.007.
48. Hong Z, Zeng X, Wei L. *et al.* Identifying enhancer-promoter interactions with neural network based on pre-trained DNA

vectors and attention mechanism. *Bioinformatics* 2020;**36**:1037–43. https://doi.org/10.1093/bioinformatics/btz694.
49. Zingaretti LM, Gezan SA, Ferrão LFV. *et al.* Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. *Front. Plant Sci* 2020;**11**:25. https://doi.org/10.3389/fpls.2020.00025.
50. Ma W, Qiu Z, Song J. *et al.* A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta* 2018;**248**:1307–18. https://doi.org/10.1007/s00425-018-2976-9.
51. Lv H, Zhang Y, Wang JS. *et al.* iRice-MS: an integrated XGBoost model for detecting multitype post-translational modification sites in rice. *Brief Bioinf* 2022;**23**:486.
52. Xu J, Lu C, Jin S. *et al.* Deep learning-based cell-specific gene regulatory networks inferred from single-cell multiome data. *Nucleic Acids Res* 2025;**53**:gkaf138. https://doi.org/10.1093/nar/gkaf138.
53. Liu B, Gao X, Zhang H. BioSeq-Analysis2.0: an updated platform for analyzing DNA, RNA and protein sequences at sequence level and residue level based on machine learning approaches. *Nucleic Acids Res* 2019;**47**:e127. https://doi.org/10.1093/nar/gkz740.
54. Li H, Pang Y, Liu B. BioSeq-BLM: a platform for analyzing DNA, RNA, and protein sequences based on biological language models. *Nucleic Acids Res* 2021;**49**:e129. https://doi.org/10.1093/nar/gkab829.
55. Shen Y, Chen LL, Gao J. CharPlant: a de novo open chromatin region prediction tool for plant genomes. *Genomics, Proteomics Bioinf* 2021;**19**:860–71. https://doi.org/10.1016/j.gpb.2020.06.021.
56. Yao Z, Zhang W, Song P. *et al.* DeepFormer: a hybrid network based on convolutional neural network and flow-attention mechanism for identifying the function of DNA sequences. *Brief Bioinf* 2023;**24**:bbad095.
57. Tang W, Wan S, Yang Z. *et al.* Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* 2018;**34**:398–406. https://doi.org/10.1093/bioinformatics/btx622.
58. Ni P, Huang N, Nie F. *et al.* Genome-wide detection of cytosine methylations in plant from nanopore data using deep learning. *Nat Commun* 2021;**12**:5976.
59. Wang Y, Zhai Y, Ding Y. *et al.* SBSM-pro: support bio-sequence machine for proteins. *Sci China-Inf Sci* 2024;**67**:212106. https://doi.org/10.1007/s11432-024-4171-9.
60. Xie H, Wang L, Qian Y. *et al.* Methyl-GP: accurate generic DNA methylation prediction based on a language model and representation learning. *Nucleic Acids Res* 2025;**53**:gkaf223. https://doi.org/10.1093/nar/gkaf223.
61. Wang L, Qian Y, Xie H. *et al.* Structured sparse regularization-based deep fuzzy networks for RNA N6-Methyladenosine sites prediction. *IEEE Trans Fuzzy Syst* 2025;**33**:131–44. https://doi.org/10.1109/TFUZZ.2024.3428402.
62. Wei L, He W, Malik A. *et al.* Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. *Brief Bioinf* 2021;**22**:bbaa275. https://doi.org/10.1093/bib/bbaa275.
63. Luo Y, Shi L, Li Y. *et al.* From intention to implementation: automating biomedical research via LLMs. *Sci China-Inf Sci* 2025;**68**:170105. https://doi.org/10.1007/s11432-024-4485-0.
64. Qiao J, Jin J, Yu H. *et al.* Towards retraining-free RNA modification prediction with incremental learning. *Inform Sci* 2024;**660**:120105. https://doi.org/10.1016/j.ins.2024.120105.
65. Li H, Liu B. BioSeq-diabolo: biological sequence similarity analysis using diabolo. *PLoS Comput Biol* 2023;**19**:e1011214. https://doi.org/10.1371/journal.pcbi.1011214.
66. Hsieh CH, Chang YS, Yen MR. *et al.* Predicting protein synergistic effect in Arabidopsis using epigenome profiling. *Nat Commun* 2024;**15**:9160. https://doi.org/10.1038/s41467-024-53565-y.
67. Hamdy R, Omar Y, Maghraby F. DeepEpi: deep learning model for predicting gene expression regulation based on epigenetic histone modifications. *Curr Bioinf* 2024;**19**:624–40. https://doi.org/10.2174/1574893618666230818121046.
68. Köhler C, Wolff P, Spillane C. Epigenetic mechanisms underlying genomic imprinting in plants. *Annu Rev Plant Biol* 2012;**63**:331–52. https://doi.org/10.1146/annurev-arplant-042811-105514.
69. Satyaki PR, Gehring M. DNA methylation and imprinting in plants: machinery and mechanisms. *Crit Rev Biochem Mol Biol* 2017;**52**:163–75. https://doi.org/10.1080/10409238.2017.1279119.
70. Erdmann RM, Satyaki PRV, Klosinska M. *et al.* A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep* 2017;**21**:3364–72. https://doi.org/10.1016/j.celrep.2017.11.078.
71. Rodrigues JA, Ruan R, Nishimura T. *et al.* Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci U S A* 2013;**110**:7934–9. https://doi.org/10.1073/pnas.1306164110.
72. Martinez G, Wolff P, Wang Z. *et al.* Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat Genet* 2018;**50**:193–8. https://doi.org/10.1038/s41588-017-0033-4.
73. Schon MA, Nodine MD. Widespread contamination of Arabidopsis embryo and endosperm transcriptome data sets. *Plant Cell* 2017;**29**:608–17. https://doi.org/10.1105/tpc.16.00845.
74. Waters AJ, Makarevitch I, Eichten SR. *et al.* Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* 2011;**23**:4221–33. https://doi.org/10.1105/tpc.111.092668.
75. Wolff P, Weinhofer I, Seguin J. *et al.* High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. *PLoS Genet* 2011;**7**:e1002126. https://doi.org/10.1371/journal.pgen.1002126.
76. Wei L, Liao M, Gao Y. *et al.* Improved and promising identification of human microRNAs by incorporating a high-quality negative set. *IEEE/ACM Trans Comput Biol Bioinform* 2014;**11**:192–201. https://doi.org/10.1109/TCBB.2013.146.
77. Wei L, Xing P, Zeng J. *et al.* Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier. *Artif Intell Med* 2017;**83**:67–74. https://doi.org/10.1016/j.artmed.2017.03.001.
78. Xin M, Yang R, Li G. *et al.* Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *Plant Cell* 2013;**25**:3212–27. https://doi.org/10.1105/tpc.113.115592.
79. Burkart-Waco D, Ngo K, Lieberman M. *et al.* Perturbation of parentally biased gene expression during interspecific hybridization. *PloS One* 2015;**10**:e0117293. https://doi.org/10.1371/journal.pone.0117293.
80. Florez-Rueda AM, Paris M, Schmidt A. *et al.* Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol Biol Evol* 2016;**33**:2935–46. https://doi.org/10.1093/molbev/msw175.
81. Baek YS, Covey PA, Petersen JJ. *et al.* Testing the SI × SC rule: pollen-pistil interactions in interspecific crosses between members of the tomato clade (*solanum* section *Lycopersicon*, Solanaceae). *Am J Bot* 2015;**102**:302–11. https://doi.org/10.3732/ajb.1400484.
82. Ferriol M, Simó U, Mansanet CJ. *et al.* Pre- and post-zygotic barriers contribute to reproductive isolation and correlate with

genetic distance in *Cucumis*. *Plants* 2023;**12**:926. https://doi.org/10.3390/plants12040926.
83. Zhang Z, Li K, Zhang H. *et al.* A single silk- and multiple pollen-expressed *PMEs* at the *Ga1* locus modulate maize unilateral cross-incompatibility. *J Integr Plant Biol* 2023;**65**:1344–55. https://doi.org/10.1111/jipb.13445.
84. Sammut SJ, Crispin-Ortuzar M, Chin SF. *et al.* Multi-omic machine learning predictor of breast cancer therapy response. *Nature* 2022;**601**:623–9. https://doi.org/10.1038/s41586-021-04278-5.
85. Liu X, Shi J, Jiao Y. *et al.* Integrated multi-omics with machine learning to uncover the intricacies of kidney disease. *Brief Bioinf* 2024;**25**:bbae364. https://doi.org/10.1093/bib/bbae364.
86. Bao SQ, Miyoshi N, Okamoto I. *et al.* Initiation of epigenetic reprogramming of the X chromosome in somatic nuclei transplanted to a mouse oocyte. *EMBO Rep* 2005;**6**:748–54. https://doi.org/10.1038/sj.embor.7400461.
87. Kinoshita T, Yadegari R, Harada JJ. *et al.* Imprinting of the *MEDEA* polycomb gene in the Arabidopsis endosperm. *Plant Cell* 1999;**11**:1945–52. https://doi.org/10.1105/tpc.11.10.1945.
88. Köhler C, Hennig L, Bouveret R. *et al. Arabidopsis* MSI1 is a component of the MEA/FIE *Polycomb* group complex and required for seed development. *EMBO J* 2003;**22**:4804–14. https://doi.org/10.1093/emboj/cdg444.
89. Tiwari S, Schulz R, Ikeda Y. *et al.* MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in *Arabidopsis*, encodes the conserved C-terminal domain of polyadenylate binding proteins. *Plant Cell* 2008;**20**:2387–98. https://doi.org/10.1105/tpc.108.061929.
90. Fitz Gerald JN, Hui PS, Berger F. Polycomb group-dependent imprinting of the actin regulator *AtFH5* regulates morphogenesis in *Arabidopsis thaliana*. *Development* 2009;**136**:3399–404. https://doi.org/10.1242/dev.036921.
91. Wang Y, Tsukamoto T, Noble JA. *et al.* Arabidopsis *LORELEI*, a maternally expressed imprinted gene, promotes early seed development. *Plant Physiol* 2017;**175**:758–73. https://doi.org/10.1104/pp.17.00427.
92. He S, Sun Y, Yang Q. *et al.* A novel imprinted gene *NUWA* controls mitochondrial function in early seed development in *Arabidopsis*. *PLoS Genet* 2017;**13**:e1006553. https://doi.org/10.1371/journal.pgen.1006553.
93. Luo M, Platten D, Chaudhury A. *et al.* Expression, imprinting, and evolution of rice homologs of the polycomb group genes. *Mol Plant* 2009;**2**:711–23. https://doi.org/10.1093/mp/ssp036.
94. Gutiérrez-Marcos JF, Costa LM, Biderre-Petit C. *et al. Maternally expressed gene1* is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell* 2004;**16**:1288–301. https://doi.org/10.1105/tpc.019778.
95. Hong J, Lee J, Jeong CW. *et al.* Identification of elements responsible for maternally-silenced imprinted gene expression of *upward curly Leaf1*, an F-box protein gene that regulates curly leaf in *Arabidopsis*. *J Plant Biol* 2020;**63**:337–46. https://doi.org/10.1007/s12374-020-09256-4.
96. Jeong CW, Roh H, Dang TV. *et al.* An E3 ligase complex regulates SET-domain polycomb group protein activity in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 2011;**108**:8036–41. https://doi.org/10.1073/pnas.1104232108.
97. Del Toro-De LG, Köhler C. Rapid origin and turnover of genomic imprinting by transposable elements. *Curr Opin Plant Biol* 2025;**87**:102764. https://doi.org/10.1016/j.pbi.2025.102764.
98. Wolff P, Jiang H, Wang G. *et al.* Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *eLife* 2015;**4**:e10074. https://doi.org/10.7554/eLife.10074.
99. Wang G, Jiang H, Toro D. *et al.* Sequestration of a transposon-derived siRNA by a target mimic imprinted gene induces postzygotic reproductive isolation in Arabidopsis. *Dev Cell* 2018;**46**:696–705.e694. https://doi.org/10.1016/j.devcel.2018.07.014.
100. Huang F, Zhu QH, Zhu A. *et al.* Mutants in the imprinted *PICKLE RELATED 2* gene suppress seed abortion of *fertilization independent seed* class mutants and paternal excess interploidy crosses in Arabidopsis. *Plant J* 2017;**90**:383–95. https://doi.org/10.1111/tpj.13500.
101. Chaudhuri S, Messing J. Allele-specific parental imprinting of *dzr1*, a posttranscriptional regulator of zein accumulation. *Proc Natl Acad Sci U S A* 1994;**91**:4867–71. https://doi.org/10.1073/pnas.91.11.4867.
102. Danilevskaya ON, Hermon P, Hantke S. *et al.* Duplicated fie genes in maize: expression pattern and imprinting suggest distinct functions. *Plant Cell* 2003;**15**:425–38. https://doi.org/10.1105/tpc.006759.
103. Jahnke S, Scholten S. Epigenetic resetting of a gene imprinted in plant embryos. *Curr Biol* 2009;**19**:1677–81. https://doi.org/10.1016/j.cub.2009.08.053.
104. Haun WJ, Laoueillé-Duprat S, O'Connell MJ. *et al.* Genomic imprinting, methylation and molecular evolution of maize *enhancer of zeste* (*Mez*) homologs. *Plant J* 2007;**49**:325–37. https://doi.org/10.1111/j.1365-313X.2006.02965.x.
105. Guo M, Rupe MA, Danilevskaya ON. *et al.* Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. *Plant J* 2003;**36**:30–44. https://doi.org/10.1046/j.1365-313X.2003.01852.x.
106. Dickinson HG, Hiscock SJ, Crane PR. *et al.* Imprinting in the endosperm: a possible role in preventing wide hybridization. *Philos Trans R Soc B* 2003;**358**:1105–11. https://doi.org/10.1098/rstb.2003.1292.
107. Lund G, Messing J, Viotti A. Endosperm-specific demethylation and activation of specific alleles of α-tubulin genes of *Zea mays* L. *Mol Gen Genet* 1995;**246**:716–22. https://doi.org/10.1007/BF00290717.
108. Lund G, Ciceri P, Viotti A. Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of *Zea mays* L. *Plant J* 1995;**8**:571–81. https://doi.org/10.1046/j.1365-313X.1995.8040571.x.
109. Zheng D, Xu J, Lu Y. Recent progresses in plant single-cell transcriptomics. *Crop Design* 2023;**2**:100041. https://doi.org/10.1016/j.cropd.2023.100041.
110. Han E, Geng Z, Qin Y. *et al.* Single-cell network analysis reveals gene expression programs for *Arabidopsis* root development and metabolism. *Plant Commun* 2024;**5**:100978. https://doi.org/10.1016/j.xplc.2024.100978.
111. Zhang TQ, Chen Y, Wang JW. A single-cell analysis of the *Arabidopsis* vegetative shoot apex. *Dev Cell* 2021;**56**:1056–74.e1058. https://doi.org/10.1016/j.devcel.2021.02.021.
112. Kim JY, Symeonidi E, Pang TY. *et al.* Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* 2021;**33**:511–30. https://doi.org/10.1093/plcell/koaa060.
113. Misra CS, Santos MR, Rafael-Fernandes M. *et al.* Transcriptomics of Arabidopsis sperm cells at single-cell resolution. *Plant Reprod* 2019;**32**:29–38. https://doi.org/10.1007/s00497-018-00355-4.

114. Picard CL, Povilus RA, Williams BP. *et al.* Transcriptional and imprinting complexity in *Arabidopsis* seeds at single-nucleus resolution. *Nat Plants* 2021;**7**:730–8. https://doi.org/10.1038/s41477-021-00922-0.
115. Huang Z, Guo X, Qin J. *et al.* Accurate RNA velocity estimation based on multibatch network reveals complex lineage in batch scRNA-seq data. *BMC Biol* 2024;**22**:290. https://doi.org/10.1186/s12915-024-02085-8.
116. Yadav VK, Santos-González J, Köhler C. INT-hi-C reveals distinct chromatin architecture in endosperm and leaf tissues of *Arabidopsis*. *Nucleic Acids Res* 2021;**49**:4371–85. https://doi.org/10.1093/nar/gkab191.
117. Consortium PCA, Jha SG, Borowsky AT. *et al.* Vision, challenges and opportunities for a plant cell atlas. *eLife* 2021;**10**:e66877. https://doi.org/10.7554/eLife.66877.
118. Henkhaus N, Bartlett M, Gang D. *et al.* Plant science decadal vision 2020-2030: reimagining the potential of plants for a healthy and sustainable future. *Plant. Direct* 2020;**4**:e00252. https://doi.org/10.1002/pld3.252.
119. Ma L, Liu J, Sun W. *et al.* scMFG: a single-cell multi-omics integration method based on feature grouping. *BMC Genomics* 2025;**26**:132. https://doi.org/10.1186/s12864-025-11319-0.
120. Guo X, Huang Z, Ju F. *et al.* Highly accurate estimation of cell type abundance in bulk tissues based on single-cell reference and domain adaptive matching. *Adv Sci* 2024;**11**:2306329. https://doi.org/10.1002/advs.202306329.
121. Xu C, Jackson SA. Machine learning and complex biological data. *Genome Biol* 2019;**20**:76. https://doi.org/10.1186/s13059-019-1689-0.
122. Karami K, Zerehdaran S, Javadmanesh A. *et al.* Attribute selection and model evaluation for the maternal and paternal imprinted genes in bovine (*Bos Taurus*) using supervised machine learning algorithms. *J Anim Breed Genet* 2019;**136**:205–16. https://doi.org/10.1111/jbg.12379.
123. Haig D, Westoby M. Parent-specific gene expression and the triploid endosperm. *American Naturalist* 1989;**134**:147–55. https://doi.org/10.1086/284971.
124. Wolf JB, Hager R. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol* 2006;**4**:e380. https://doi.org/10.1371/journal.pbio.0040380.
125. Raissig MT, Bemer M, Baroux C. *et al.* Genomic imprinting in the Arabidopsis embryo is partly regulated by PRC2. *PLoS Genet* 2013;**9**:e1003862. https://doi.org/10.1371/journal.pgen.1003862.
126. Ashbrook DG, Reinmar H. Empirical testing of hypotheses about the evolution of genomic imprinting in mammals. *Front Neuroanat* 2013;**7**:6. https://doi.org/10.3389/fnana.2013.00006.
127. Villar CBR, Erilova A, Makarevich G. *et al.* Control of *PHERES1* imprinting in Arabidopsis by direct tandem repeats. *Mol Plant* 2009;**2**:654–60. https://doi.org/10.1093/mp/ssp014.
128. Haun WJ, Danilevskaya ON, Meeley RB. *et al.* Disruption of imprinting by mutator transposon insertions in the 5′ proximal regions of the *Zea mays Mez1* locus. *Genetics* 2009;**181**:1229–37. https://doi.org/10.1534/genetics.108.093666.
129. Li T, Yin L, Stoll CE. *et al.* Conserved noncoding sequences and de novo *mutator* insertion alleles are imprinted in maize. *Plant Physiol* 2022;**191**:299–316. https://doi.org/10.1093/plphys/kiac459.

Briefings in Functional Genomics, 2026, **25**, elaf025
https://doi.org/10.1093/bfgp/elaf025
Review Paper