

RESEARCH ARTICLE

Hybrid seed incompatibility in *Capsella* is connected to chromatin condensation defects in the endosperm

Katarzyna Dziasek¹ , Lauriane Simon¹ , Clément Lafon-Placette^{1,2} , Benjamin Laenen³ , Cecilia Wärdig¹ , Juan Santos-González¹ , Tanja Slotte³ , Claudia Köhler^{1*} 

1 Department of Plant Biology, Uppsala Biocenter, Swedish University of Agricultural Sciences, Linnean Center of Plant Biology, Uppsala, Sweden, **2** Present address: Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic, **3** Department of Ecology, Environment and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden

 These authors contributed equally to this work.

* Claudia.kohler@slu.se



 OPEN ACCESS

Citation: Dziasek K, Simon L, Lafon-Placette C, Laenen B, Wärdig C, Santos-González J, et al. (2021) Hybrid seed incompatibility in *Capsella* is connected to chromatin condensation defects in the endosperm. *PLoS Genet* 17(2): e1009370. <https://doi.org/10.1371/journal.pgen.1009370>

Editor: Xiaoqi Feng, John Innes Centre, UNITED KINGDOM

Received: September 4, 2020

Accepted: January 15, 2021

Published: February 11, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1009370>

Copyright: © 2021 Dziasek et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data available from NCBI database. For the copy number analysis, we used data available under ERR636163, ERR636164, SRR8394203, SRR8394204 for *Cr*,

Abstract

Hybridization of closely related plant species is frequently connected to endosperm arrest and seed failure, for reasons that remain to be identified. In this study, we investigated the molecular events accompanying seed failure in hybrids of the closely related species pair *Capsella rubella* and *C. grandiflora*. Mapping of QTL for the underlying cause of hybrid incompatibility in *Capsella* identified three QTL that were close to pericentromeric regions. We investigated whether there are specific changes in heterochromatin associated with interspecific hybridizations and found a strong reduction of chromatin condensation in the endosperm, connected with a strong loss of CHG and CHH methylation and random loss of a single chromosome. Consistent with reduced DNA methylation in the hybrid endosperm, we found a disproportionate deregulation of genes located close to pericentromeric regions, suggesting that reduced DNA methylation allows access of transcription factors to targets located in heterochromatic regions. Since the identified QTL were also associated with pericentromeric regions, we propose that relaxation of heterochromatin in response to interspecies hybridization exposes and activates loci leading to hybrid seed failure.

Author summary

Seed failure in response to interspecific hybridizations is a well-known reproductive barrier preventing interbreeding of closely related species and thus maintaining species boundaries. This reproductive barrier is established in the endosperm, a nourishing tissue supporting embryo growth. In this study, we discovered that the endosperm of interspecific hybrids between the recently diverged species *Capsella rubella* and *C. grandiflora* suffers from mitotic abnormalities and random chromosome loss. We found that the endosperm has reduced levels of DNA methylation and chromatin condensation, likely accounting for the chromosome loss. Importantly, we found that genes located in

and SRR5988314, SRR5988315, SRR5988316, SRR5988317 for *Cg*. Whole genome resequencing data from the *Cg* × *Cr* F1 and the *Cr* parent are available under PRJEB9020. Endosperm genomic DNA from *Cr*, *Cg*, and *Cr* × *Cg* and bisulfite sequencing data are available under PRJNA647289. *Capsella* endosperm RNA-seq data are available under GSE67359. *Arabidopsis* endosperm RNA-seq data are available under GSE84122.

Funding: This research was supported by grants from the Swedish Research Council VR (to CK, grant #2017-04119), a grant from the Knut and Alice Wallenberg Foundation (to CK, grant #2018-0206), and support from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (to CK). The work of BL and TS was supported by grants from the Science for Life Laboratory and the Swedish Research Council (grant #621-2010-5508 and #621-2013-4320). Computational work was enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2016-07213. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

pericentromeric regions were preferentially deregulated, suggesting that reduced DNA methylation exposes transcription factor binding sites in pericentromeric regions, leading to hyperactivation of genes and seed arrest. In support of the relevance of pericentromeric regions for hybrid seed arrest, we identified three QTL connected with the phenotype that were all located in pericentromeric regions. These results link epigenetic changes in hybrid endosperm with distinct genetic loci underpinning hybrid seed failure.

Introduction

The endosperm is a nourishing tissue supporting embryo growth, similar to the placenta in mammals. In most flowering plants the endosperm is a triploid tissue, derived after fertilization of the diploid central cell by one of the haploid sperm cells [1]. The endosperm is sensitive to parental genome dosage and the ratio of two maternal to one paternal genome copies is required for normal development [2,3]. In most flowering plants the endosperm initially develops as a coenocyte, where nuclear divisions are not followed by cell wall formation [4]. The transition to cellularization is essential for viable seed formation [5]; however, the precise mechanism underlying this transition and the reason for its requirement remain to be identified. Hybridizations of plants that differ in ploidy disrupt the parental genome balance and affect endosperm cellularization, leading to seed arrest [6–9]. Similarly, hybridization between closely related species frequently leads to defects in endosperm development [6,8,10–13], suggesting similar underlying mechanisms. In support of this notion, interspecies hybridization barriers can be overcome by ploidy manipulations [7,14,15]. Despite the widespread occurrence of hybrid seed failure among flowering plants, the underlying molecular mechanisms are poorly understood.

In the *Capsella* genus, the selfing species *C. rubella* (referred to as *Cr*) separated less than 200,000 years ago from the obligate outcrosser *C. grandiflora* (referred to as *Cg*) [16–18]. Both species have the same ploidy level, but are separated by a strong endosperm-based hybridization barrier in both directions of hybridization [19,20]. Nevertheless, hybrid seed phenotypes differ depending on the direction of hybridization; when *Cg* maternal plants are pollinated with *Cr* pollen (referred to as *Cg* × *Cr*), the endosperm cellularizes precociously, giving rise to very small seeds. In the reciprocal cross (*Cr* × *Cg*), endosperm cellularization is delayed and seeds collapse [19]. Very similar reciprocal phenotypes result from interploidy crosses of *Arabidopsis*; seeds derived from crosses of tetraploid maternal plants with diploid pollen donors ($4x \times 2x$) resemble *Cg* × *Cr* seeds, while crosses of diploid maternal plants with tetraploid pollen donors ($2x \times 4x$) resemble *Cr* × *Cg* seeds [19,21]. Interploidy seed failure of $2x \times 4x$ crosses is causally connected to increased expression of imprinted paternally expressed genes (PEGs) [22–25]. The PEG *PHERES1* (*PHE1*) encodes an AGAMOUS-LIKE (AGL) transcription factor that is also highly upregulated in triploid seeds [26,27]. *PHE1* acts upstream of many PEGs, likely accounting for increased expression of PEGs and other direct *PHE1* targets in triploid seeds, leading to seed arrest [25]. Many AGLs, including *PHE1* and *PHE1* orthologs, are also upregulated in *Arabidopsis* and *Capsella* *Cr* × *Cg* interspecies hybrids [19,28], correlating with the similar phenotypes of interploidy ($2x \times 4x$) and interspecies hybrid seeds [14,19,21,28]. The endosperm derived from interploidy ($2x \times 4x$) crosses has reduced levels of CHH methylation (H corresponds to A, T, or C), a hallmark of the RNA-directed DNA methylation (RdDM) pathway. The RdDM pathway establishes DNA methylation in all sequence contexts and is guided by 24-nt small RNAs (sRNAs) [29–31]. Activity of the RdDM pathway is low in the early endosperm and increases during endosperm development [32,33]. Maternal tissues

surrounding the female gametophyte form 24-nt sRNAs that accumulate in the endosperm [34,35] and may guide *de novo* methylation in the endosperm.

To elucidate the molecular cause of seed abortion in *Cr* × *Cg* interspecies hybrids, we performed a QTL analysis and identified three phenotype-associated QTL that were all localized in pericentromeric regions. Hybrid endosperm had strongly reduced CHG and CHH methylation, associated with chromatin decondensation, mitotic abnormalities and random chromosome loss. Deregulated genes in hybrids were preferentially localized in pericentromeric regions and enriched for orthologs of PHE1 targets, suggesting that increased expression of AGLs in hybrid endosperm ectopically activates target genes in hypomethylated pericentromeric regions.

Results

QTL associated with *Cr* × *Cg* incompatibility are localized in pericentromeric regions

To identify the genetic loci involved in hybrid seed incompatibility between *Cr* and *Cg*, we generated an F2 *Cg/Cr* population and crossed 480 F2 individuals as pollen donors to *Cr* maternal plants. We scored the seed abortion rate of the resulting hybrid seeds and genotyped the F2 individuals using a double digest restriction-site associated DNA (ddRAD) approach. This information was used for mapping quantitative trait loci (QTL) associated with hybrid seed abortion (Fig 1A). We detected three significant QTL located on chromosomes 2, 3 and 7

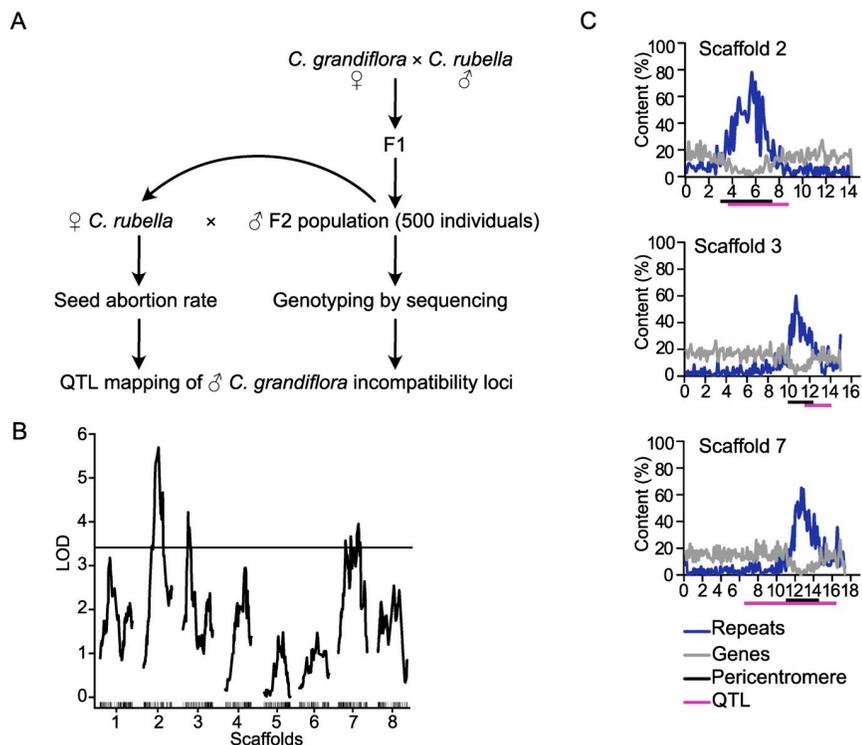


Fig 1. QTL associated with *Cr* × *Cg* incompatibility are localized in pericentromeric regions. A) Scheme of experimental design for QTL mapping. B) LOD scores for *C. grandiflora* QTL associated with abortion of *Cr* × *Cg* seeds. The plot shows the QTL LOD profile across the eight linkage groups corresponding to main pseudochromosomal scaffolds of the *C. rubella* genome assembly. The horizontal line represents the genome-wide significance threshold (= 0.05, estimated by 1000 permutations). C) Plots of gene (grey line) and repeat (blue line) content on three scaffolds containing identified QTL (from top to bottom: scaffolds 2, 3 and 7). The magenta bars represent QTL regions and the black bars indicate pericentromeric regions as previously defined [36].

<https://doi.org/10.1371/journal.pgen.1009370.g001>

(Fig 1B). All identified regions were very broad, ranging from positions 3.6 to 8.8 Mb on scaffold 2, 11.5 to 14.1 Mb on scaffold 3, and 6.4 to 16.4 Mb on scaffold 7. The strong QTL on scaffold 2 overlaps with a previously identified QTL based on a *Cg/Cr* RIL population, while we did not identify the weak QTL previously found on scaffold 8 [19]. Strikingly, all three QTL span centromeric and pericentromeric regions [36], suggesting a role of centromeric/pericentromeric regions in hybrid incompatibility in *Capsella*.

Hybrid endosperm shows mitotic abnormalities and random chromosome loss

Previous work revealed that interspecific hybrids frequently encounter chromosome loss, generally resulting in the complete elimination of one parental genome [37–39]. This uniparental genome elimination is caused by the loss of the centromere-specific histone H3 variant CENH3 from one of the parental genomes in the hybrid [40]. We hypothesized that the presence of the three QTL in centromeric/pericentromeric regions may be connected to chromosome loss in *Capsella* hybrid endosperm. To test this hypothesis, we counted the number of chromocenters in nuclei of hybrid and parental endosperms. We found indeed that the hybrid endosperm had at least one chromocenter less compared to the parental species (Fig 2A and 2B). The shape of the chromocenters was strikingly different between parents and hybrid; while *Cr* and *Cg* interphase endosperm nuclei had clearly formed chromocenters, in the hybrid endosperm the shape of the chromocenters was diffuse, indicating that the chromatin was less condensed (Fig 2B). Close inspection of metaphase plates in hybrid endosperm nuclei revealed the presence of abnormally shaped spindles and misaligned chromosomes, which were not observed in the endosperm of the *Cr* parental species (Fig 2C).

The chromocenters in embryos were highly dispersed, making it not possible to count the number of chromocenters directly in the embryos of the hybrid seeds. In order to test whether the observed defects were endosperm specific, we rescued hybrid *Cr* × *Cg* embryos and performed chromosome spreading from root nuclei of the obtained hybrid plants. Neither the number of chromocenters (Fig 2D) nor the chromatin condensation differed between parental and hybrid root tissue (Fig 2E), revealing that chromosome loss was restricted to the endosperm.

We next addressed the question whether there was a specific chromosome that was lost in the hybrid endosperm, or whether the loss occurred randomly. We manually dissected endosperm from hybrid and the non-hybrid seeds, isolated DNA and performed high-throughput sequencing. We then determined the coverage over all 8 scaffolds of the *Cr* genome (Fig 2F), but did not identify a significant coverage reduction in any scaffold, as expected if a specific chromosome would be lost. We confirmed this observation by plotting the proportion of *Cg* to total SNPs (single nucleotide polymorphisms) in all scaffolds. Since *Cg* was the paternal parent in the hybrid endosperm, this ratio is expected to be 33.3%. The obtained ratio was close to 40% in all scaffolds and did not significantly differ between the scaffolds, consistent with the similar coverage over all scaffolds (Fig 2F). Together, this data strongly suggest that the hybrid *Cr* × *Cg* endosperm suffers from random chromosome loss and that chromosome loss is specific for the hybrid endosperm and does not occur in the hybrid embryo.

Characterization of centromeric repeats of *Cr* and *Cg*

The outcome of the QTL mapping together with the fact that the hybrid endosperm encountered random chromosome loss, prompted us to investigate whether the centromeric regions of *Cr* and *Cg* differ. Fluorescence in situ hybridization (FISH) signals obtained with probes of the *Cr* centromeric consensus repeats colocalized with the chromocenters in *Cr* and *Cg* (Fig

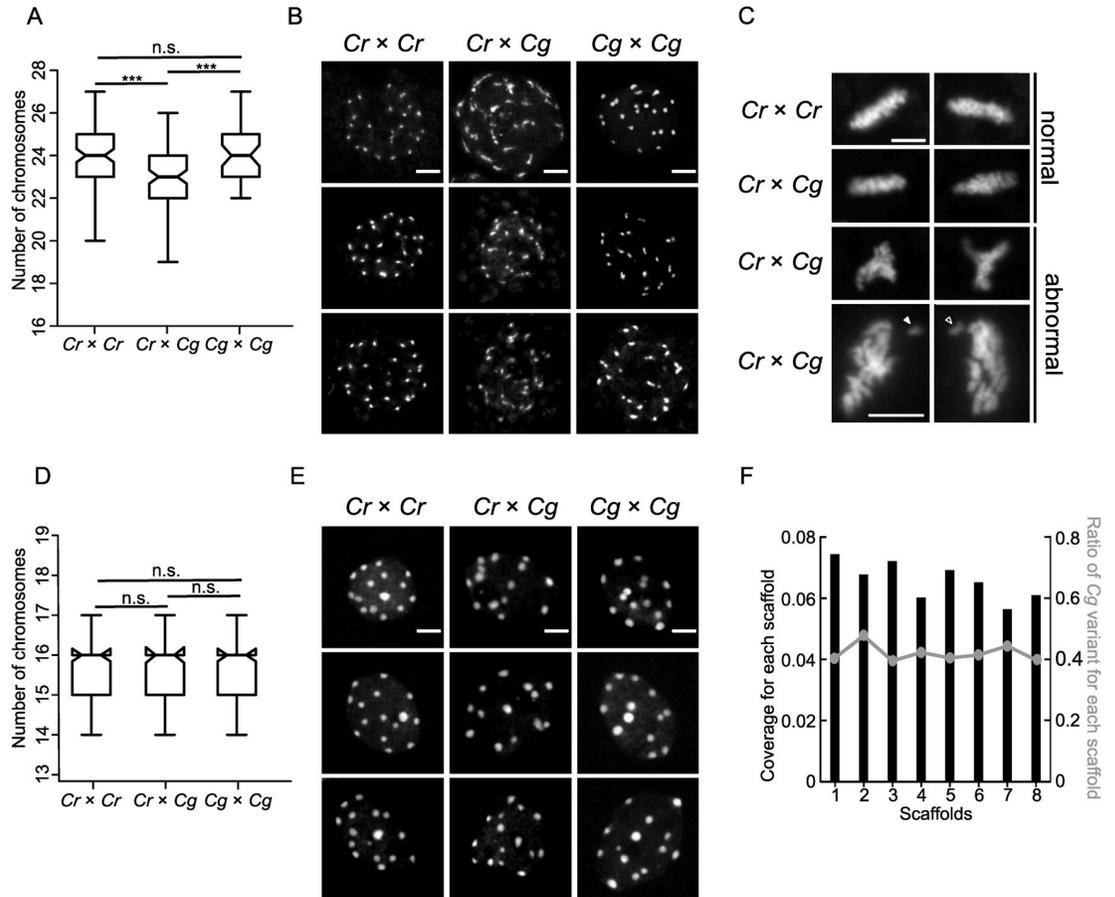


Fig 2. Hybrid endosperm shows random chromosome loss and mitotic abnormalities. A) Boxplots showing number of chromosomes in nuclei of *Cr* × *Cr* (n = 101), *Cr* × *Cg* (n = 101) and *Cg* × *Cg* (n = 53) 4 DAP seeds. Boxes show medians and the interquartile range, and error bars show the full range excluding outliers. Asterisks indicate significant differences calculated by Wilcoxon test (** p-value < 0.001). B) DAPI stained chromocenters from endosperm nuclei of *Cr* × *Cr*, *Cr* × *Cg* and *Cg* × *Cg*. Scale bar, 5 μm. C) DAPI stained pictures of metaphase plates of *Cr* × *Cr* and *Cr* × *Cg* endosperm nuclei from 4 DAP seeds. Arrows indicate lagging chromosomes. Scale bar, 5 μm. D) Box plots showing number of chromosomes in root nuclei of *Cr* × *Cr*, *Cr* × *Cg* and *Cg* × *Cg* seedlings. For each genotype 100 nuclei were analyzed. Boxes show medians and the interquartile range, and error bars show the full range excluding outliers. Differences between genotypes are not significant (Wilcoxon test). E) DAPI stained chromocenters from root nuclei of *Cr* × *Cr*, *Cr* × *Cg* and *Cg* × *Cg* seedlings. Scale bar, 3 μm. F) Histogram representing the coverage of each scaffold in hybrid nuclei. The line represents the ratio of *Cg* to total reads in each scaffold in hybrid seeds.

<https://doi.org/10.1371/journal.pgen.1009370.g002>

3A), indicating that centromeric repeats of *Cr* and *Cg* were similar and positioned in the centromeric region.

We tested whether there are species-specific SNPs for *Cr* or *Cg* in the centromeric region. However, in the published centromeric *Cr* sequence [41] we did not identify specific mutations present only in one of the two species, the SNP profiles for *Cr* and *Cg* centromeric repeats were almost identical (S1 Fig).

To address the question whether the number of centromeric repeats differ between *Cr* and *Cg*, we blasted the consensus sequence of the *Cr* centromeric repeats [41] to *Cg* scaffolds and found 88 scaffolds with multiple hits organized in tandem repeats, revealing that the *Cg* genome contains centromeric repeats similar to *Cr*. Next, we used the *Cr* centromeric consensus sequence as reference and mapped the publicly available genomic data of four different *Cg* and *Cr* populations to the centromeric repeat sequence. On average, 8.9% of the *Cg* reads could be mapped to centromeric repeats, while only an average of 6.7% of the *Cr* reads,

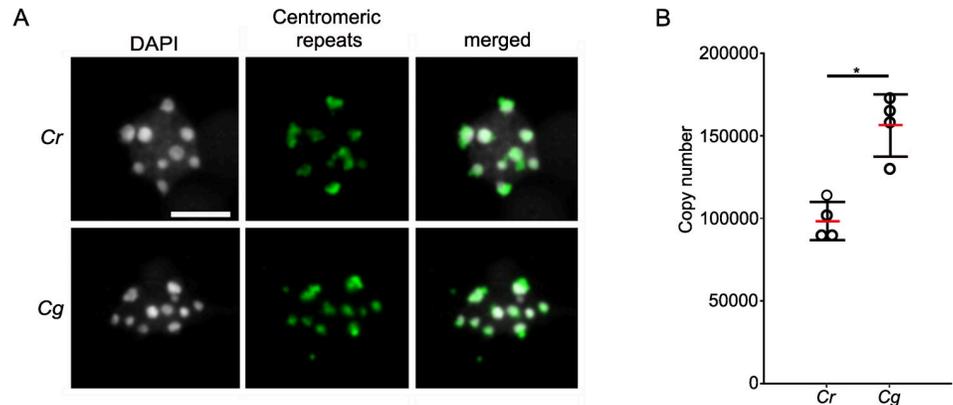


Fig 3. Characterization of centromeric repeats of *Cr* and *Cg*. A) FISH with centromeric repeat probes on *Cr* and *Cg* leaf nuclei. Scale bar, 5 μ m. B) Column scatter plot showing copy number of centromeric repeats in *Cr* and *Cg* genomes. Red bar represents the mean value. Asterisks indicate significant differences calculated by Wilcoxon test (* p-value < 0.05).

<https://doi.org/10.1371/journal.pgen.1009370.g003>

revealing that centromeric repeats are more abundant in the *Cg* compared to the *Cr* genome. To further test this, we determined the centromeric repeat copy number in *Cr* and *Cg* by comparing the coverage of the centromeric repeats with the coverage of single copy genes. We found indeed about 1.5 times more copies in *Cg* (~150 000) than in *Cr* (~100 000) (Fig 3B).

Together, these results show that centromeric repeats from *Cr* and *Cg* are highly similar in sequence, but that the *Cg* genome contains substantially more copies of centromeric repeats than the *Cr* genome.

Hybrid endosperm shows chromatin decondensation associated with loss of DNA methylation on transposable elements

The centromere-specific histone variant CENH3 is crucial for proper chromocenter formation [42]. Unlike typical histones, CENH3 is a fast evolving protein and CENH3 differences between species were reported to cause chromosome elimination [40,43,44]. Alignment of the protein sequences of CENH3 from *Cr* (Carubv10010436m) and *Cg* (Cagra.1968s0080.1) revealed that they differed in two amino acids, while the canonical H3 histones in these two species were identical (S2 Fig). In order to test whether there were global changes in CENH3 loading in the hybrid endosperm nuclei in comparison to non-hybrid nuclei, we generated an antibody against *Capsella* CENH3 and performed immunolocalization using this antibody. However, we did not observe obvious differences in CENH3 localization between parental *Cr* and hybrid endosperm (Fig 4A), indicating that CENH3 is properly loaded in the hybrid endosperm. We therefore considered it unlikely that failure of CENH3 loading accounts for the observed chromosome loss.

Chromocenters of hybrid endosperm were less condensed compared to the parental endosperm (Fig 2B). Since loss of DNA methylation is frequently accompanied with reduced chromatin condensation and heterochromatin loss [45,46], we tested whether reduced DNA methylation in hybrid endosperm may account for reduced chromocenter condensation, leading to chromosome loss. We performed bisulfite sequencing of DNA isolated from parental and hybrid endosperm at 4 DAP. We observed a strong decrease of CHG and CHH methylation on transposable elements (TEs) in hybrid endosperm in comparison to both parental species (Fig 4B and 4C); while genes had an intermediate level of DNA methylation in all sequence contexts compared to the parental species (S3 Fig).

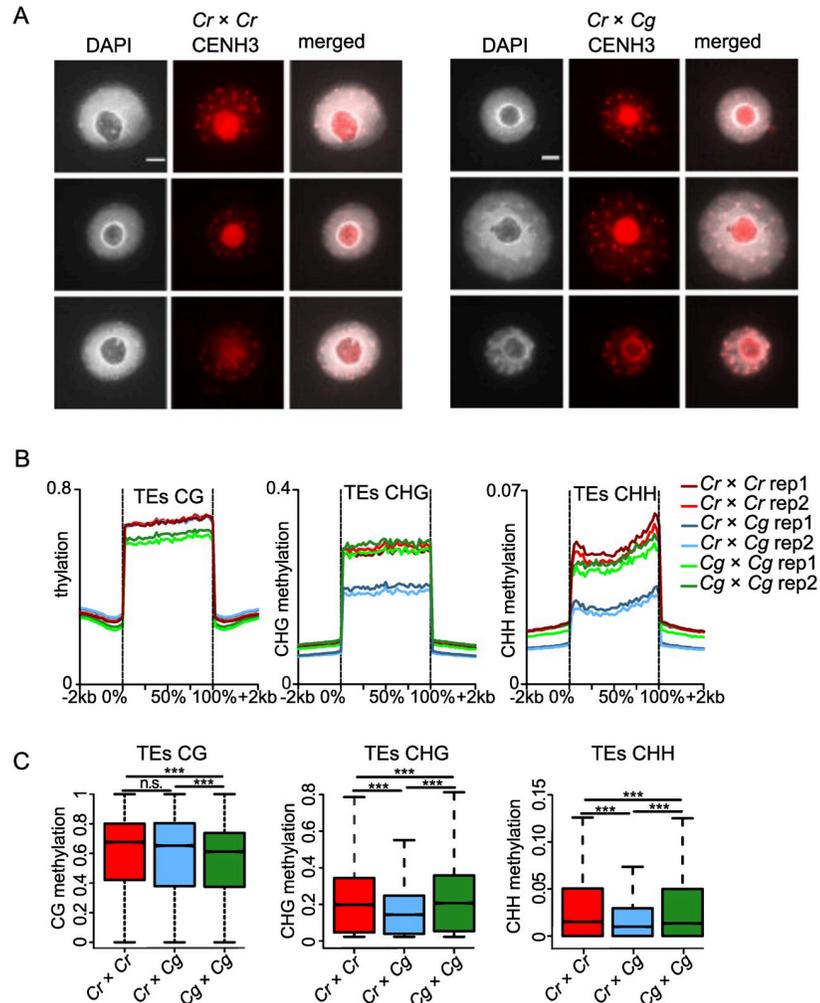


Fig 4. Endosperm chromatin decondensation associates with loss of DNA methylation but not CENH3 occupancy. A) Immunostaining of CENH3 in the endosperm of *Cr* × *Cr* and *Cr* × *Cg* 4 DAP seeds and chromatin staining with DAPI. Scale bar, 3 μm. B) Methylation level of CG, CHG and CHH of transposable elements (TEs) in *Cr*, *Cr* × *Cg* and *Cg* endosperm of 4 DAP seeds. C) Boxplots showing the methylation level of transposable elements (TEs) in *Cr* × *Cr* and *Cr* × *Cg* and *Cr* × *Cg* endosperm of 4 DAP seeds. Boxes show medians and the interquartile range, and error bars show the full range excluding outliers. Asterisks indicate statistically significant differences calculated by Wilcoxon test (***) p-value < 0.001, * p-value < 0.05).

<https://doi.org/10.1371/journal.pgen.1009370.g004>

Deregulated genes in hybrid endosperm are preferentially localized in pericentromeric regions

Around half of the TEs losing DNA methylation were localized in pericentromeric regions (S1 Table), which was significantly more than expected by chance ($p < 0.0001$, Chi-square test). We therefore asked whether deregulated genes are also preferentially located in pericentromeric regions. We analyzed previously published transcriptome data of parental and hybrid seeds [19] to test the spatial distribution of genes that were overexpressed in the hybrid relative to the *Cr* parent. For this purpose, we divided each scaffold into deciles containing equal numbers of genes and plotted the number of upregulated genes in each group in all scaffolds (Fig 5A). Strikingly, upregulated genes were preferentially localized in pericentromeric regions (determined in [36] in all scaffolds ($p < 0.0001$, binomial test), suggesting that loss of DNA methylation and chromatin decondensation cause preferential activation of genes in

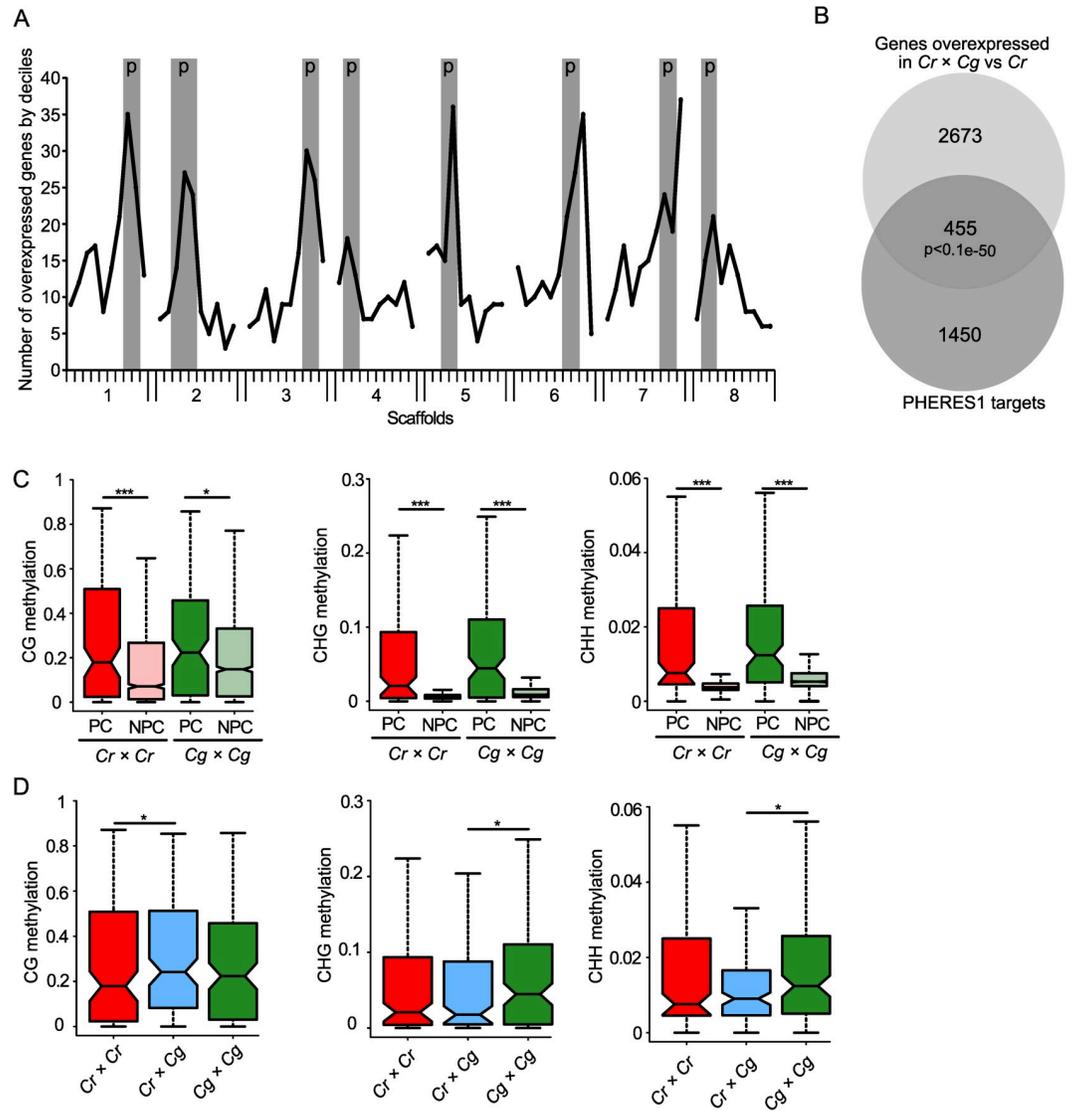


Fig 5. Deregulated genes in hybrid endosperm are preferentially localized in pericentromeric regions. A) Number of genes overexpressed in *Cr* × *Cg* in comparison to *Cr* per decile of genes on each scaffold. Pericentromeric regions (p) are highlighted in grey for each scaffold. B) Venn diagram showing the overlap between significantly overexpressed genes in hybrid seeds compared to *Cr* × *Cr* and PHE1 targets. C) Boxplots showing average methylation levels of the promoter regions (upstream 500 bp of the transcriptional start site) of PHE1 target orthologs in pericentromeric (PC) and non-pericentromeric (NPC) regions in *Cr* × *Cr* and *Cg* × *Cg* endosperm of 4 DAP seeds. D) Boxplots showing methylation levels of the promoter regions of PHE1 target orthologs in pericentromeric regions in *Cr* × *Cr*, *Cr* × *Cg*, and *Cg* × *Cg* endosperm of 4 DAP seeds. Boxes show medians and the interquartile range, and error bars show the full range excluding outliers. Asterisks indicate significant differences calculated by Wilcoxon test (***) p-value < 0.001, ** p-value < 0.01, * p-value < 0.05).

<https://doi.org/10.1371/journal.pgen.1009370.g005>

pericentromeric regions. We found a similar pattern, while less pronounced, when analyzing genes being deregulated in hybrid seeds in comparison to both parents. Also here, there was a preferential localization of upregulated genes in pericentromeric regions (S4A Fig) ($p < 0.0001$, binomial test). The pattern for downregulated genes looked strikingly different, where with the exception of scaffold 1 no enrichment in pericentromeric regions was detected (S4B Fig).

We previously found that AGAMOUS-LIKE (AGL) MADS-box transcription factors are strongly upregulated in *Capsella* hybrid endosperm [19]. Similarly, AGL transcription factors

are also highly upregulated in triploid *Arabidopsis* seeds [5,26,28,47] and the AGL transcription factor PHERES1 (PHE1) was shown to causally account for triploid seed arrest [25]. The *Capsella* ortholog of *PHE1* (Carubv10020903m.g, referred to as *CrPHE1*) is also highly upregulated *Capsella* hybrid seeds [19]. We therefore tested whether deregulated genes in hybrid *Capsella* seeds were enriched for orthologs of PHE1 target genes. We found that out of the 3128 upregulated genes in the hybrid compared to *Cr*, 455 were orthologs of PHE1 targets (Fig 5B), which is a significant overlap ($p < 1.e-50$, hypergeometric test). Interestingly, there were also significantly more pericentromeric PHE1 targets (26 out of 146 (= 17.8%)) overexpressed in hybrid seeds compared to non-pericentromeric PHE1 targets (213 out of 1759 (= 12.1%)) ($p < 0.05$, hypergeometric test). The overlap of PHE1 targets with deregulated genes was also significant when testing deregulated genes in the hybrid in comparison to both parents (S4C Fig; $p < 0.05$, hypergeometric test).

Previous work suggested that PHE1 binding is prevented by DNA methylation [25], suggesting that increased expression of PHE1 targets in pericentromeric regions of hybrid endosperm is a consequence of reduced DNA methylation in those regions. We compared the DNA methylation levels in the promoter region of PHE1 target orthologs in pericentromeric regions and non-pericentromeric region and found indeed that targets in pericentromeric regions had higher levels of DNA methylation in all sequence contexts compared to non-pericentromeric targets (Fig 5C). While CG methylation of pericentromeric targets in the hybrid endosperm was slightly increased compared to parental methylation, CHG and CHH methylation was significantly lower in the hybrid compared to parental *Cg* endosperm (Fig 5D). Importantly, the PHE1 target motifs predominantly contain cytosines in CHH context (Batista et al., 2019) [25]; supporting the idea that loss of CHH methylation in hybrid endosperm may expose binding sites for *CrPHE1* and cause increased expression of *CrPHE1* targets. Among the deregulated orthologs of PHE1 targets were *AGL40* on scaffold 7 and *AGL95* on scaffold 3, located within the identified QTL. Based on yeast-two-hybrid interaction data, *AGL40* encodes for a direct interaction partner of PHE1; while *AGL95* is a paralog of PHE1 [48]. Both genes were also highly upregulated in triploid seeds (S5 Fig), as expected for direct PHE1 target genes. There was a pronounced loss of CHH and CHG methylation in the promoter and coding region of *AGL40* in the hybrid endosperm compared to both parents (S5 Fig). *AGL95* also had reduced DNA methylation in the hybrid endosperm compared to the *Cg* parent (S5 Fig). Moreover, other PHE1 targets that were deregulated in *Capsella* hybrids were similarly deregulated in triploid seeds (S6 Fig).

Together, this data support the idea that loss of DNA methylation in hybrid endosperm exposes binding sites for *CrPHE1* and potentially other type I AGLs and the resulting increased expression of those targets causally connects to failure in endosperm cellularization and seed arrest.

Discussion

Understanding the molecular cause for hybrid incompatibility is a major goal of evolutionary biology. It is also of high relevance for plant breeding, since it may facilitate the generation of new hybrid varieties. In this study, we provide insights into hybrid incompatibility of two closely related *Capsella* species. We found that *Cr* × *Cg* hybrid seeds undergo endosperm-specific chromatin decondensation, leading to random chromosome loss. Chromatin decondensation is likely a consequence of reduced DNA methylation in the endosperm.

Hypomethylation in pericentromeric regions exposes binding sites for the AGL transcription factor *CrPHE1*, leading to hyperactivation of potential *CrPHE1* targets.

Pericentromeric regions play a role in hybrid incompatibility

To uncover the genetic elements involved in *Cr* × *Cg* incompatibility, we performed QTL mapping using an F2 *Cg/Cr* population. In a previous study we followed a similar approach using *Cg/Cr* recombinant inbred lines (RILs) [19]. However, the majority of RILs did not trigger seed abortion when crossed with *Cr*, indicating that the alleles responsible for incompatibility are purged from the RIL population. To overcome this problem, in this study we used a *Cg/Cr* F2 population and identified three *Cg* loci that contribute to *Cr* × *Cg* incompatibility, consistent with our previous genetic prediction [19]. One of the QTLs on scaffold 2 was also detected in our previous study [19]. Interestingly, all three QTL were localized in pericentromeric regions, suggesting a particular role of pericentromeric regions in hybrid incompatibility. Pericentromeric heterochromatic regions were previously shown to be relevant for hybrid incompatibility in *Drosophila*. Hybrid incompatibility genes *Lhr* (*Lethal hybrid rescue*) and *Hmr* (*Hybrid male rescue*) encode for heterochromatin proteins and localize to centromeric heterochromatin [49–51]. Furthermore, the incompatibility locus *Zhr* (*Zygotic hybrid rescue*) is a species-specific heterochromatic repeat present on the X-chromosome of *D. melanogaster* but absent in *D. simulans* [52]. The paternal *D. melanogaster* X chromosome, containing the *Zhr* repeats, fails to segregate in a *D. simulans* maternal background. Our finding that the three *Cr* × *Cg* incompatibility QTL mapped to pericentromeric regions and that hybrid endosperm underwent chromosome loss, prompted us to investigate the role of pericentromeric regions in hybrid incompatibility. Previous work revealed that hybridization of *Arabidopsis* lines expressing species-specific variants of CENH3 to wild-type individuals causes severe chromosome segregation errors [44]. These abnormalities occurred in embryo and endosperm, contrasting to the endosperm-specific chromosome loss that we observed in *Cr* × *Cg* hybrids. Adding the fact that we did not observe any differences in CENH3 loading on mitotic chromosomes in hybrid endosperm, we consider it unlikely that the two amino acid differences between *Cr* and *Cg* CENH3 account for the observed chromosome loss. Our data rather point that differences in pericentromeric repeat number between *Cr* and *Cg* connects to hybrid seed failure. We found that length and sequence composition of *Cg* centromeric repeats was similar as previously described for *Cr* [41]. However, centromeres of *Cg* contained about 50% more repeats than *Cr*, revealing species-specific differences in centromere size between *Cr* and *Cg*. The precise size of the *Cg* genome is unknown, but flow cytometry analysis predicts that it is about 10% larger than *Cr* [53]. Based on our work, this difference is likely to be contributed by the increased number of centromeric repeats of *Cg*.

Possible consequences of differential centromeric repeat numbers in *Cr* and *Cg*

Maternal tissues are the main source of 24-nt sRNAs accumulating in the endosperm [34,35] and likely guide *de novo* methylation in the endosperm. Since sperm DNA is highly depleted of CHH methylation [54,55], increased paternal genome dosage or, possibly, increased numbers of centromeric repeats, may lead to remethylation failure, if maternal 24-nt siRNAs are rate-limiting. Consistent with this scenario, CHH methylation is strongly depleted in the endosperm of 2x × 4x hybrid endosperm [56,57], similar to what we observed in this study. Centromeric repeats in *Arabidopsis* are also present in pericentromeric regions where they are heavily methylated [58]; therefore, increased numbers of centromeric repeats may impact on methylation levels outside of the centromere. Thus, non-matching dosage of maternal 24-nt sRNAs and paternal genome copies or repeats may lead to hypomethylation and consequently decondensation and random chromosome loss in the endosperm. The consistent phenotypes observed in *Cr* × *Cg* hybrid seeds make it however unlikely that random chromosome loss is

causally connected to endosperm failure. Both, $Cr \times Cg$ and $2x \times 4x$ *Arabidopsis* hybrid seeds show similar phenotypic defects, most prominently cellularization defects in the endosperm [19,21]. Increased expression of the AGL transcription factor *PHE1* is causally responsible for $2x \times 4x$ *Arabidopsis* hybrid seed defects [25] and *CrPHE1* and related AGLs are similarly highly upregulated in $Cr \times Cg$ hybrid seeds [19]. Consistent with the idea that increased expression of *CrPHE1* and related AGLs are connected to $Cr \times Cg$ hybrid seed failure, we found a high overrepresentation of *PHE1* target orthologs among deregulated genes in *Capsella* hybrids. Importantly, *PHE1* binding was shown to be negatively impacted by DNA methylation [25]. Since pericentromeric regions lose DNA methylation in hybrids, this may explain preferential activation of *CrPHE1* targets in hypomethylated pericentromeric regions.

In summary, in this study we uncovered the molecular defects occurring in $Cr \times Cg$ hybrid endosperm and its likely genetic cause. We report that hybrid endosperm has reduced levels of CHG and CHH methylation, likely causing reduced chromatin condensation and random chromosome loss. We speculate that the cause for the hypomethylation is an imbalance of maternal *Cr* 24-nt sRNAs and *Cg* centromeric repeats. Increased expression of *CrPHE1* and related AGLs hyperactivate *CrPHE1* targets preferentially in hypomethylated pericentromeric regions, causing a phenotypic mimic to interploidy hybrid seeds in *Arabidopsis*.

Materials and methods

Plant material and growth conditions

In this study, we used the *Cr* accessions *Cr48.21* and *Cr1g* and the *Cg* accessions *Cg89.3*, *Cg81*, *Cg89.16* and *Cg94*. Seeds were surface sterilized and sown on agar plates containing $\frac{1}{2}$ Murashige and Skoog (MS) medium and 1% sucrose. After stratification for 2 days in the dark at 4°C, seedlings were grown in a growth room under long-day photoperiod (16 h light and 8 h darkness) at 22°C light and 20°C darkness temperature and a light intensity of 110 μ E. Seedlings were transferred to pots and plants were grown in a growth chamber at 60% humidity and daily cycles of 16 h light at 21°C and 8 h darkness at 18°C and a light intensity of 150 μ E.

DNA isolation and preparation of ddRAD-Seq libraries

Mature leaves of $Cg \times Cr$ F_2 individuals (480 in total) were collected and flash frozen in liquid nitrogen. Genomic DNA was isolated from leaves using a CTAB extraction protocol [59]. We used a double-digest RAD-sequencing (ddRAD-seq) protocol [60] modified according to [61]. Briefly, about 500 ng of DNA per sample was successively digested with the restriction enzymes *EcoRI* and *Taq*^oI. The resulting fragments were ligated with restriction site-specific barcoded adapters and size-selected (to ~550 bp) using AMPure beads (Beckman). The adapters were labeled with biotin, which allowed to perform an additional selection of adapter-ligated fragments using Dynabeads M-270 Streptavidin (Invitrogen). In total, 5 dual indexing 96-plex ddRAD-seq libraries were prepared of the F_2 samples. The F_1 and *C. rubella* parental samples were previously sequenced [62]. ddRAD-seq libraries were processed with 125bb paired-end sequencing on a total of five lanes (one library per lane) of Illumina HiSeq2500 system at the SNP&SEQ Technology Platform of SciLifeLab, Uppsala, Sweden.

ddRAD-Seq read processing, variant calling and filtering

We obtained on average over 2.2 million reads per F_2 individual. Short-read data were demultiplexed and trimmed from barcode sequences using *ipyrad* [63] and we detected and trimmed sequencing adapters using *trimmomatic* v0.36 [64]. Trimmed reads were mapped to the v1.0 *Cr* reference genome assembly [18] using *BWA-MEM* [65]. We called variants and genotypes

using GATK 3.8–0 [66] HaplotypeCaller after Base Quality Score Recalibration (BQSR) using a set of known SNPs [67]. We filtered the resulting vcf file using VCFtools [68] to retain only biallelic SNPs that were genotyped in at least 95% of the samples with a read depth between 8 and 200 and a mapping quality of at least 50. SNPs in repetitive regions (identified using RepeatMasker as in [62]) were further removed using bedtools v. 2.26.0 [69]. After filtering, we retained a total of 13,326 SNPs.

Linkage map construction

We constructed a linkage map based on our SNP data in R/QTL [70]. For efficient linkage map construction, we first thinned our SNP set to retain 948 equally spaced SNPs using mapthin [71]. We inferred the parental origin of all SNP alleles based on whole-genome resequencing data from the *Cr* parent and the F1 [62]. We discarded individuals that were genotyped for less than 95% of markers and filtered markers for segregation distortion as recommended in the R/QTL manual. The final map was constructed based on 623 markers genotyped in 383 F2s. Briefly, we partitioned SNPs into linkage groups using a maximum recombination fraction of 0.35 and a minimum LOD score of 8, and ordered markers on each linkage group. The resulting linkage map had a total length of 529 cM and eight linkage groups with more than one SNP marker, in good agreement with the expected haploid chromosome number ($n = 8$) of *Cr* and *Cg*.

QTL mapping

The phenotyping was performed by crossing each F2 individual as pollen donor to the *Cr*48.21 accession as maternal plant, and for each *Cr* × F2 cross, 5 siliques were harvested, amounting to about 60 seeds per cross. The rate of aborted seeds (shriveled and dark) was measured for each cross and was used as continuous trait for the QTL analysis. The QTL analysis was done using interval mapping with the expectation maximization algorithm in R/QTL [70]. The background control parameter was set to a standard model with backward regression to select any possible QTL with standard five control markers. The QTL analysis gave three significant peaks that were located in three different chromosomes. The genome-wide significance threshold ($\alpha = 0.05$) was obtained based on 1000 permutations. In the three significant QTLs, the presence of *Cg* alleles correlated with seed abortion.

Embryo rescue

Embryo rescue was performed as described in [19].

Chromosome spreading in root tips

Seedlings were grown on ½ MS agar plates. After 10 days, the root tips (about 1 cm) were cut with a razor blade. Roots were treated with colchicine (100 μM), 8-hydroxyquinoline (2.5 mM) and oryzalin (100 μM) for 2h at room temperature (RT) to block cell division and then fixed in ethanol:acetic acid (3:1) for 3h at RT. Fixed tissue was rinsed with citrate buffer for 10 min, transferred to 500 μl enzyme mix (0.3% w/v cytohelicase, 0.3% w/v pectolyase, 0.3% w/v cellulase) and incubated at 37°C for 1h. Then, the enzyme mix was removed and the tissue was washed with citrate buffer for 30 min. Root tips were incubated in a drop of 45% acetic acid for a few minutes and then transferred on a clean microscope slide. Tissue fragments were teased apart with a fine needle and gently squashed with the cover slip. Slides were frozen in liquid nitrogen; cover slips were removed and the squashed tissue was air dried. Slides were

mounted with Vectashield mounting medium with 4',6-Diamidino-2'-phenylindole dihydrochloride (DAPI) (BioNordika AB).

Immunostaining

10 siliques at 4 days after pollination (DAP) for each genotype were harvested and fixed in cold 4% formaldehyde in Tris buffer (10 mM Tris-HCl pH 7.5, 10 mM NaEDTA, 100 mM NaCl) for 20 min and washed for 2×10 min with cold Tris buffer. Seeds were isolated from the siliques and chopped with a razor blade in 100 µl LB01 buffer (15 mM Tris-HCl pH 7.5, 2 mM NaEDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl and 0.1% Triton X-100). The cell slurry was filtered through a 30 µm falcon cell strainer. 5 µl of nuclei suspension was mixed with 10 µl of sorting buffer (100 mM Tris-HCl pH 7.5, 50 mM KCl, 2 mM MgCl₂, 0.05% TWEEN-20 and 5% sucrose), spread on a polylysine slide and air dried for 2 h. Slides were postfixed in 2% formaldehyde in PBS for 5 min and washed with water. Slides were covered with 1X PBS +0.5% Triton X-100 and kept in a moist chamber for 45 min at RT, then washed 3 times with 1X PBS for 5 min. Slides were denatured by adding 30 µl of deionized water and heated on a preheated plate at 80°C for 8 min, then cooled down by dipping in 1XPBS. Slides were incubated with primary antibody (diluted 1:100 in 5% BSA, 0.05% TWEEN-20 in 1X PBS) for 1h at RT and then overnight at 4°C, then washed the 3 times in 1XPBS 5min at RT and incubated with the secondary antibody (1:200, Abcam ab175471) for 2-3h at RT. Slides were washed 3 times with PBS for 5 min at RT and mounted with Vectashield mounting medium with DAPI (BioNordika AB). The experiment has been performed in three independent biological replicates.

Endosperm nuclei spreading

4 DAP seeds were harvested and incubated overnight in a mix of 2.5 mM 8-hydroxyquinoline, 100 µM oryzalin, 100 µM colchicine, then fixed for at least 5h in fixative ethanol:acetic acid (3:1) at 4°C. Seeds were washed with 10 mM citrate buffer and incubated for 5h in an enzyme mix containing 0.3% cytohellicase, 0.3% pectolyase, and 0.3% cellulase in 10mM citrate buffer. After digestion, 5~10 seeds were put on a slide, squashed with a needle, spread with acetic acid (60%) and fixed with 3:1 ethanol:acetic acid on the slide. Slides were mounted with Vectashield mounting medium with DAPI (BioNordika AB). The experiment has been performed in three independent biological replicates.

Fluorescence in situ hybridization

For fluorescence in situ hybridization (FISH), three week old leaves were fixed in ethanol-acetic acid (3:1) and FISH was performed as described [72]. Centromeric repeat probes were amplified by PCR using Biotin-11-dUTP (ThermoFisher) with primer TCTAGCACTTGTA ATCAATCAAATTC and AGAAGTGAGAAGAAAGACTTG. To detect the probe, an anti-Biotin antibody (FITC) (ab53469, Invitrogen) was used at a concentration of 1/1000.

Copy number estimation

To estimate the copy number of centromeric repeats through next-generation-sequencing (NGS), we divided the average coverage over the repeats by the average coverage over three reference genes (Carubv10009577m, Carubv10011569m, Carubv10006001m for *Cr* and their orthologues Cagra.3392s0015, Cagra.0804s0001, Cagra.3807s0042 for *Cg*). For each individual dataset analyzed in this study, we mapped the reads to a single reference consisting of the *Cr*

centromeric repeat consensus sequence [41] using BWA-MEM (v0.7.8) [65]. We retrieved per-base read depth with the function Depthofcoverage from GATK (v3.5) [66].

Identification of polymorphisms

To determine polymorphisms in centromeric repeats, we mapped the reads with BWA-MEM [65] to the consensus centromeric repeat of *Cr* and used previously published scripts (https://gist.github.com/laurianesimon/a9fc44aa83305c576e914710cae75f87#file-listsnpmodules_v2; https://gist.github.com/laurianesimon/a9fc44aa83305c576e914710cae75f87#file-countpolymorphisms_v4 [73]) and extracted and quantified polymorphisms for each position of the centromeric repeat from all mapped reads. We used whole genome sequencing data of leaves SRR5988314 and ERR636164 for *Cg* and *Cr* polymorphisms, respectively.

Whole genome and bisulfite sequencing

Endosperm from 4DAP seeds was manually dissected in extraction buffer from the DNAeasy kit (Qiagen). We followed the protocol for DNA extraction as recommended in the manual. Isolated DNA was sent directly to Novogene (Hongkong, China) for whole genome or bisulfite sequencing on a HiSeqX in 150-bp paired-end mode. Two replicates of 4DAP endosperm bisulfite sequencing were performed for each cross: *Cr* × *Cg*, *Cr* × *Cr*, *Cg* × *Cg*.

RNAseq analyses

For RNA analysis, reads were mapped to the *Arabidopsis* or the *Capsella* reference genomes, using TopHat v2.1.0 [74]. Gene expression was normalized to reads per kilobase per million mapped reads (RPKM) using GFOLD [75]. Expression level for each condition was calculated using the mean of the expression values in both replicates. Differentially regulated genes across the replicates were detected using the rank product method, as implemented in the Bioconductor RankProd Package [76].

Variant frequency calling

Whole genome sequencing data from 4DAP manually dissected endosperm from *Cr* × *Cr* and *Cr* × *Cg* were mapped to the *Cr* reference genome. Variants were called with bcftools call/mpileup [77] (<http://samtools.github.io/bcftools/bcftools.html>). Only variants specifically present in the *Cr* × *Cg* data were considered to be *Cg* variants. Means of the ratio between *Cg* variant/base coverage were calculated for each chromosome.

CENH3 antibody design

Custom antibodies for detection of CENH3 from *Capsella* were generated by BioNordika AB. Two peptides were used for immunization of rabbits: NH₂- C+DFDLARRLGGKGRPW-COOH and NH₂- C+QASQKKKPYRYPGT-CONH₂. Antibodies were subjected to affinity purification and validated by ELISA. They were coupled to KLH (keyhole limpet hemocyanin) carrier protein and delivered in standard buffer (PBS 1x, 0.01% thimerosal and 0.1% BSA).

Quantitative data

Raw data for all plots in the manuscript are shown in [S1 Data](#).

Supporting information

S1 Fig. Centromeric repeats in *Cr* and *Cg* genomes are identical. Frequency of SNPs along centromeric repeats in *Cr* and *Cg* genomes.
(EPS)

S2 Fig. Protein alignment of *Cr* and *Cg* CENH3 variants and H3 variants. Protein alignment of *Cr* and *Cg* CENH3 variants, H3.1 (Carubv10003405m and Cagra.4395s0104 are identical to each other) and H3.3 (Carubv10021126m and Cagra.0799s0032.1 are identical to each other). Red arrows indicate amino acid differences between the proteins. Grey areas mark the sequences used as peptides for antibody production.
(EPS)

S3 Fig. Methylation level of genes in *Cr*, *Cg* and *Cr* × *Cg* endosperm of 4 DAP seeds. A) Metagene plots showing methylation level of genes in *Cr*, *Cg* and *Cr* × *Cg* endosperm of 4 DAP seeds. B) Boxplots showing the methylation level of genes in *Cr*, *Cr* × *Cg* and *Cg* endosperm of 4 DAP seeds. Boxes show medians and the interquartile range, and error bars show the full range excluding outliers. Asterisks indicate significant differences calculated by Wilcoxon test (***p*-value < 0.001).
(EPS)

S4 Fig. Upregulated genes in hybrid endosperm are preferentially localized in pericentromeric regions. A) Number of genes upregulated in *Cr* × *Cg* in comparison to *Cr* and *Cg* per decile of genes on each scaffold. Pericentromeric regions (p) are highlighted in grey for each scaffold B) Number of genes downregulated in *Cr* × *Cg* in comparison to *Cr* per decile of genes on each scaffold. Pericentromeric regions (p) are highlighted in grey for each scaffold. C) Venn diagram showing the overlap between significantly overexpressed genes in hybrid seeds compared to *Cr* × *Cr* and *Cg* × *Cg* and PHE1 targets (p-values were calculated using the supertest function from R package SuperExactTest [78]).
(EPS)

S5 Fig. *AGL40* and *AGL95* are overexpressed in *Capsella* hybrid endosperm and in the endosperm of *Arabidopsis* triploid seeds. A) Expression level of *AGL95* and *AGL40* in *Cr*, *Cg* and *Cr* × *Cg* endosperm. B) Expression level of *AGL95* and *AGL40* in the endosperm of diploid and triploid seeds. C) CHG and CHH methylation on *AGL40*. D) CHG and CHH methylation on *AGL95*.
(EPS)

S6 Fig. PHE1 target genes are overexpressed in *Capsella* hybrid endosperm and in the endosperm of *Arabidopsis* triploid seeds. Scatter plot showing expression of PHE1 target genes in *Cr* × *Cg* compared to *Cr* × *Cr* seeds and deregulated genes in the endosperm of *Arabidopsis* triploid versus diploid seeds. Upregulated PHE1 target genes are highlighted in red (p value was calculated using the Pearson correlation test).
(EPS)

S1 Table. Localization of TEs losing DNA methylation.
(EPS)

S1 Data. Raw data for all quantitative assays shown in the manuscript.
(XLSX)

Acknowledgments

We thank Kim Steige for helping with generating the F2 population used in this study.

Author Contributions

Conceptualization: Katarzyna Dziasek, Lauriane Simon, Clément Lafon-Placette, Tanja Slotte, Claudia Köhler.

Data curation: Benjamin Laenen, Juan Santos-González.

Formal analysis: Lauriane Simon, Benjamin Laenen, Juan Santos-González.

Funding acquisition: Claudia Köhler.

Investigation: Katarzyna Dziasek, Lauriane Simon, Clément Lafon-Placette, Cecilia Wärdig.

Methodology: Tanja Slotte, Claudia Köhler.

Project administration: Claudia Köhler.

Supervision: Claudia Köhler.

Validation: Katarzyna Dziasek, Lauriane Simon, Claudia Köhler.

Writing – original draft: Katarzyna Dziasek, Lauriane Simon, Claudia Köhler.

Writing – review & editing: Katarzyna Dziasek, Lauriane Simon, Clément Lafon-Placette, Tanja Slotte, Claudia Köhler.

References

1. Li J, Berger F. Endosperm: Food for humankind and fodder for scientific discoveries. *New Phytol.* 2012; 195: 290–305. <https://doi.org/10.1111/j.1469-8137.2012.04182.x> PMID: 22642307
2. Leblanc O, Pointe C, Hernandez M. Cell cycle progression during endosperm development in *Zea mays* depends on parental dosage effects. *Plant J.* 2002; 32: 1057–1066. <https://doi.org/10.1046/j.1365-3113x.2002.01491.x> PMID: 12492846
3. Lin BY. Ploidy barrier to endosperm development in maize. *Genetics.* 1984; 107: 103–15. PMID: 17246209
4. Costa LM, Gutiérrez-Marcos JF, Dickinson HG. More than a yolk: The short life and complex times of the plant endosperm. *Trends in Plant Science.* 2004. pp. 507–514. <https://doi.org/10.1016/j.tplants.2004.08.007> PMID: 15465686
5. Hehenberger E, Kradolfer D, Köhler C. Endosperm cellularization defines an important developmental transition for embryo development. *Development.* 2012; 139: 2031–2039. <https://doi.org/10.1242/dev.077057> PMID: 22535409
6. Brink RA, Cooper DC. The endosperm in seed development. *Bot Rev.* 1947; 132: 423–541.
7. Woodell SRJ, Valentine DH. Studies in british primulas. IX. seed incompatibility in diploid-autotetraploid crosses. *New Phytol.* 1961; 60: 282–294. <https://doi.org/10.1111/j.1469-8137.1961.tb06256.x>
8. Ramsey J, Schemske DW. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst.* 1998; 29: 467–501. <https://doi.org/10.1146/annurev.ecolsys.29.1.467>
9. Lafon-Placette C, Köhler C. Endosperm-based postzygotic hybridization barriers: Developmental mechanisms and evolutionary drivers. *Mol Ecol.* 2016. <https://doi.org/10.1111/mec.13552> PMID: 26818717
10. Ishikawa R, Ohnishi T, Kinoshita Y, Eiguchi M, Kurata N, Kinoshita T. Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 2011; 65: 798–806. <https://doi.org/10.1111/j.1365-3113X.2010.04466.x> PMID: 21251103
11. Sukno S, Ruso J, Jan CC, Melero-Vara JM, Fernández-Martínez JM. Interspecific hybridization between sunflower and wild perennial *Helianthus* species via embryo rescue. *Euphytica.* 1999; 106: 69–78. <https://doi.org/10.1023/A:1003524822284>
12. Dinu II, Hayes RJ, Kynast RG, Phillips RL, Thill CA. Novel inter-series hybrids in *Solanum*, section Petota. *Theor Appl Genet.* 2005; 110: 403–415. <https://doi.org/10.1007/s00122-004-1782-x> PMID: 15517147
13. Roy AK, Malaviya DR, Kaushal P. Generation of interspecific hybrids of *Trifolium* using embryo rescue techniques. *Methods Mol Biol.* 2011; 710: 141–151. https://doi.org/10.1007/978-1-61737-988-8_12 PMID: 21207268

14. Lafon-Placette C, Johannessen IM, Hornslien KS, Ali MF, Bjerkan KN, Bramsiepe J, et al. Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc Natl Acad Sci U S A*. 2017; 114: E1027–E1035. <https://doi.org/10.1073/pnas.1615123114> PMID: 28115687
15. Tonosaki K, Sekine D, Ohnishi T, Ono A, Furuumi H, Kurata N, et al. Overcoming the species hybridization barrier by ploidy manipulation in the genus *Oryza*. *Plant J*. 2018; 93: 534–544. <https://doi.org/10.1111/tbj.13803> PMID: 29271099
16. Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, Wright SI. Recent speciation associated with the evolution of selfing in *Capsella*. *Proc Natl Acad Sci U S A*. 2009; 106: 5241–5245. <https://doi.org/10.1073/pnas.0807679106> PMID: 19228944
17. Guo YL, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, Weigel D, et al. Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. *Proc Natl Acad Sci U S A*. 2009; 106: 5246–5251. <https://doi.org/10.1073/pnas.0808012106> PMID: 19307580
18. Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo YL, et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet*. 2013; 45: 831–835. <https://doi.org/10.1038/ng.2669> PMID: 23749190
19. Rebernick CA, Lafon-Placette C, Hatorangan MR, Slotte T, Köhler C. Non-reciprocal interspecies hybridization barriers in the *Capsella* Genus are established in the endosperm. *PLoS Genet*. 2015; 11: e1005295. <https://doi.org/10.1371/journal.pgen.1005295> PMID: 26086217
20. Lafon-Placette C, Hatorangan MR, Steige KA, Cornille A, Lascoux M, Slotte T, et al. Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. *Nat Plants*. 2018; 4: 352–357. <https://doi.org/10.1038/s41477-018-0161-6> PMID: 29808019
21. Scott RJ, Spielman M, Bailey J, Dickinson HG. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development*. 1998; 125: 3329–3341. PMID: 9693137
22. Kradolfer D, Wolff P, Jiang H, Siretskiy A, Köhler C. An imprinted gene underlies postzygotic reproductive isolation in *Arabidopsis thaliana*. *Dev Cell*. 2013; 26: 525–535. <https://doi.org/10.1016/j.devcel.2013.08.006> PMID: 24012484
23. Jiang H, Moreno-Romero J, Santos-González J, De Jaeger G, Gevaert K, Van De Slijke E, et al. Ectopic application of the repressive histone modification H3K9me2 establishes post-zygotic reproductive isolation in *Arabidopsis thaliana*. *Genes Dev*. 2017; 31: 1272–1287. <https://doi.org/10.1101/gad.299347.117> PMID: 28743695
24. Wang G, Jiang H, Del Toro de León G, Martínez G, Köhler C. Sequestration of a transposon-derived siRNA by a target mimic imprinted gene induces postzygotic reproductive isolation in *Arabidopsis*. *Dev Cell*. 2018; 46: 696–705.e4. <https://doi.org/10.1016/j.devcel.2018.07.014> PMID: 30122632
25. Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, et al. The MADS-box transcription factor Pheres1 controls imprinting in the endosperm by binding to domesticated transposons. *Elife*. 2019; 8. <https://doi.org/10.7554/eLife.50541> PMID: 31789592
26. Erilova A, Brownfield L, Exner V, Rosa M, Twell D, Scheid OM, et al. Imprinting of the Polycomb group gene MEDEA serves as a ploidy sensor in *Arabidopsis*. *PLoS Genet*. 2009; 5: e1000663. <https://doi.org/10.1371/journal.pgen.1000663> PMID: 19779546
27. Tiwari S, Spielman M, Schulz R, Oakey RJ, Kelsey G, Salazar A, et al. Transcriptional profiles underlying parent-of-origin effects in seeds of *Arabidopsis thaliana*. *BMC Plant Biol*. 2010; 10: 72. <https://doi.org/10.1186/1471-2229-10-72> PMID: 20406451
28. Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L. Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr Biol*. 2009; 19: 1128–1132. <https://doi.org/10.1016/j.cub.2009.05.068> PMID: 19559614
29. Cao X, Jacobsen SE. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr Biol*. 2002; 12: 1138–1144. [https://doi.org/10.1016/s0960-9822\(02\)00925-9](https://doi.org/10.1016/s0960-9822(02)00925-9) PMID: 12121623
30. Zilberman D, Cao X, Jacobsen SE. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*. 2003; 299: 716–719. <https://doi.org/10.1126/science.1079695> PMID: 12522258
31. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. RNA polymerase *v* transcription guides ARGONAUTE4 to chromatin. *Nat Genet*. 2009; 41: 630–634. <https://doi.org/10.1038/ng.365> PMID: 19377477
32. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *Curr Biol*. 2012; 22: 1825–1830. <https://doi.org/10.1016/j.cub.2012.07.061> PMID: 22940470

33. Moreno-Romero J, Jiang H, Santos-González J, Köhler C. Parental epigenetic asymmetry of PRC 2-mediated histone modifications in the *Arabidopsis* endosperm. *EMBO J*. 2016; 35: 1298–1311. <https://doi.org/10.15252/embj.201593534> PMID: 27113256
34. Grover JW, Kendall T, Baten A, Burgess D, Freeling M, King GJ, et al. Maternal components of RNA-directed DNA methylation are required for seed development in *Brassica rapa*. *Plant J*. 2018; 94: 575–582. <https://doi.org/10.1111/tbj.13910> PMID: 29569777
35. Grover JW, Burgess D, Kendall T, Baten A, Pokhrel S, King GJ, et al. Abundant expression of maternal siRNAs is a conserved feature of seed development. *Proc Natl Acad Sci U S A*. 2020; 117: 202001332. <https://doi.org/10.1073/pnas.2001332117> PMID: 32541052
36. Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Nueffer B, et al. Long-term balancing selection drives evolution of immunity genes in *Capsella*. *Elife*. 2019;8. <https://doi.org/10.7554/eLife.43606> PMID: 30806624
37. Kasha KJ, Kao KN. High frequency haploid production in barley (*Hordeum vulgare* L.). *Nature*. 1970; 225: 874–876. <https://doi.org/10.1038/225874a0> PMID: 16056782
38. Lange W. Crosses between *Hordeum vulgare* L. and *H. bulbosum* L. II. Elimination of chromosomes in hybrid tissues. *Euphytica*. 1971; 20: 181–194. <https://doi.org/10.1007/BF00056078>
39. Bennett MD, Finch RA, Barclay IR. The time rate and mechanism of chromosome elimination in *Hordeum* hybrids. *Chromosom*. Springer-Verlag; 1976; 54: 175–200. <https://doi.org/10.1007/BF00292839>
40. Sanei M, Pickering R, Kumke K, Nasuda S, Houben A. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc Natl Acad Sci U S A*. 2011; 108: E498–E505. <https://doi.org/10.1073/pnas.1103190108> PMID: 21746892
41. Hall SE, Luo S, Hall AE, Preuss D. Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. *Genetics*. 2005; 170: 1913–1927. <https://doi.org/10.1534/genetics.104.038208> PMID: 15937135
42. Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell*. 2002; 14: 1053–1066. <https://doi.org/10.1105/tpc.010425> PMID: 12034896
43. Henikoff S, Ahmad K, Malik HS. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science*.; 2001. pp. 1098–1102. <https://doi.org/10.1126/science.1062939> PMID: 11498581
44. Maheshwari S, Tan EH, West A, Franklin FCH, Comai L, Chan SWL. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet*. 2015; 11. e1004970. <https://doi.org/10.1371/journal.pgen.1004970> PMID: 25622028
45. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol*. 2015; 16: 519–532. <https://doi.org/10.1038/nrm4043> PMID: 26296162
46. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010. pp. 204–220. <https://doi.org/10.1038/nrg2719> PMID: 20142834
47. Jullien PE, Berger F. Parental genome dosage imbalance deregulates imprinting in *Arabidopsis*. *PLoS Genet*. 2010; 6. e1000885. <https://doi.org/10.1371/journal.pgen.1000885> PMID: 20333248
48. Pařenicová L, De Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. Molecular and phylogenetic analyses of the complete MADS-Box transcription factor family in *Arabidopsis*: New openings to the MADS world. *Plant Cell*. 2003; 15: 1538–1551. <https://doi.org/10.1105/tpc.011544> PMID: 12837945
49. Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*. 2006; 314: 1292–1295. <https://doi.org/10.1126/science.1133953> PMID: 17124320
50. Thomae AW, Schade GOM, Padeken J, Borath M, Vetter I, Kremmer E, et al. A pair of centromeric proteins mediates reproductive isolation in *Drosophila* species. *Dev Cell*. 2013; 27: 412–424. <https://doi.org/10.1016/j.devcel.2013.10.001> PMID: 24239514
51. Satyaki PRV, Cuykendall TN, Wei KHC, Brideau NJ, Kwak H, Aruna S, et al. The *Hmr* and *Lhr* hybrid incompatibility genes suppress a broad range of heterochromatic repeats. *PLoS Genet*. 2014; 10. e1004240. <https://doi.org/10.1371/journal.pgen.1004240> PMID: 24651406
52. Ferree PM, Barbash DA. Species-Specific Heterochromatin Prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. Noor MAF, editor. *PLoS Biol*. 2009; 7: e1000234. <https://doi.org/10.1371/journal.pbio.1000234> PMID: 19859525
53. Hurka H, Friesen N, German DA, Franzke A, Nueffer B. 'Missing link' species *Capsella orientalis* and *Capsella thracica* elucidate evolution of model plant genus *Capsella* (Brassicaceae). *Mol Ecol*. 2012; 21: 1223–1238. <https://doi.org/10.1111/j.1365-294X.2012.05460.x> PMID: 22288429

54. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*. 2012; 151: 194–205. <https://doi.org/10.1016/j.cell.2012.09.001> PMID: 23000270
55. Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*. 2012; 337: 1360–1364. <https://doi.org/10.1126/science.1224839> PMID: 22984074
56. Schatlowski N, Wolff P, Santos-González J, Schoft V, Siretskiy A, Scott R, et al. Hypomethylated pollen bypasses the interploidy hybridization barrier in *Arabidopsis*. *Plant Cell*. 2014; 26: 3556–3568. <https://doi.org/10.1105/tpc.114.130120> PMID: 25217506
57. Martinez G, Wolff P, Wang Z, Moreno-Romero J, Santos-González J, Conze LL, et al. Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat Genet*. 2018; 50: 193–198. <https://doi.org/10.1038/s41588-017-0033-4> PMID: 29335548
58. Zhang W, Lee HR, Koo DH, Jiang J. Epigenetic modification of centromeric chromatin: Hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell*. 2008; 20: 25–34. <https://doi.org/10.1105/tpc.107.057083> PMID: 18239133
59. Doyle J, Doyle F. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 1987; 19:11–15.
60. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*. 2012; 7: e37135. <https://doi.org/10.1371/journal.pone.0037135> PMID: 22675423
61. Liu X, Karrenberg S. Genetic architecture of traits associated with reproductive barriers in *Silene*: Coupling, sex chromosomes and variation. *Mol Ecol*. 2018; 27: 3889–3904. <https://doi.org/10.1111/mec.14562> PMID: 29577481
62. Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T. *Cis*-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. *Mol Biol Evol*. 2015; 32: 2501–2514. <https://doi.org/10.1093/molbev/msv169> PMID: 26318184
63. Eaton DAR. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*. 2014; 30: 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121> PMID: 24603985
64. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
65. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN] [Preprint]. 2013. Available from: <http://www.arxiv-vanity.com/papers/1303.3997/>.
66. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20: 1297–1303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
67. Steige KA, Laenen B, Reimegård J, Scofield DG, Slotte T. Genomic analysis reveals major determinants of *cis*-regulatory variation in *Capsella grandiflora*. *Proc Natl Acad Sci U S A*. 2017; 114: 1087–1092. <https://doi.org/10.1073/pnas.1612561114> PMID: 28096395
68. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
69. Quinlan AR, Hall IM, Chen Q, Yang L, Huang H, Miki D, et al. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278
70. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 2003; 19: 889–890. <https://doi.org/10.1093/bioinformatics/btg112> PMID: 12724300
71. Howey R, Cordell HJ. MapThin: Thinning your map files for linkage analyses! 2011. Available from: <http://www.staff.ncl.ac.uk/richard.howey/mapthin/>.
72. Bowler C, Benvenuto G, Laflamme P, Molino D, Probst A V, Tariq M, et al. Chromatin techniques for plant cells. *Plant J*. 2004; 39: 776–789. <https://doi.org/10.1111/j.1365-313X.2004.02169.x> PMID: 15315638
73. Simon L, Rabanal FA, Dubos T, Oliver C, Lauber D, Poulet A, et al. Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in *Arabidopsis thaliana*. *Nucleic Acids Res*. 2018; 46: 3019–3033. <https://doi.org/10.1093/nar/gky163> PMID: 29518237
74. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25: 1105–1111. <https://doi.org/10.1093/bioinformatics/btp120> PMID: 19289445

75. Feng J, Meyer CA, Wang Q, Liu JS, Liu XS, Zhang Y. GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics*. 2012; 28: 2782–2788. <https://doi.org/10.1093/bioinformatics/bts515> PMID: 22923299
76. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*. 2006; 22: 2825–2827. <https://doi.org/10.1093/bioinformatics/btl476> PMID: 16982708
77. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
78. Wang M, Zhao Y, Zhang B. Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep*. 2015; 5: 1–12. <https://doi.org/10.1038/srep16923> PMID: 26603754