Role of methylation and siRNA on differential allelic expression in a hybrid of distantly related citrus species.

Diop Khadidiatou[1], Gibert Anaïs[1], Llauro Christel[1], Froelicher Yann[2], Hufnagel Barbara[3]*, Picault Nathalie[1]*, Pontvianne Frédéric[1]*

[1]CNRS/UPVD, Plant Genome and Development (LGDP), Perpignan, 66100, France
[2]CIRAD, UMR AGAP Institut, F- 20230 San Giuliano, France.
UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France
[3]CIRAD, UMR AGAP Institut, F-97170 Petit-Bourg, Guadeloupe, France.
UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France.

*co-corresponding authors
frederic.pontvianne@cnrs.fr, nathalie.picault@univ-perp.fr, barbara.hufnagel@cirad.fr,

Authors Orcid ID:
Khadidiatou Diop: 0009-0001-2255-7489
Frederic Pontvianne: 0000-0002-2913-4104
Nathalie Picault: 0000-0003-2770-9688
Barbara Hufnagel: 0000-0002-3515-2122

Abstract

DNA methylation plays a central role in the regulation of gene expression. In plants, methylation occurs in the CG, CHG and CHH contexts, *via* distinct DNA methyltransferases including MET1, CMT3 and the RNA-directed DNA Methylation (RdDM) pathway *via* DRM2. In interspecific hybrids, these epigenetic mechanisms are confronted to a mixed small RNA population and two subgenomes harboring specific methylation patterns, therefore generating unique expression profiles. The aim of this work was to understand these regulations by analysing gene expression, DNA methylation and small RNAs in a Citrus hybrid resulting from the cross between *C. reticulata* (mandarin) and *C. australasica* (caviar lemon). Haplotype-resolved subgenomes assembly identified hundreds of alleles-specifically expressed genes. Asymmetric reprogramming of methylation was observed, in particular an increase in the CHH context in *C. australasica* haplotype. Surprisingly, CHH methylation, usually associated with gene silencing, was correlated here with increased expression, but also with abundant 24nt small RNA populations at their promoter regions. Similar analyses of the parental lines suggest the correlation between CHH methylation-enriched promoter and high expression level is not due to the hybridization, but seem to be generally true for all citrus. These observations suggest that, in citrus fruit, RdDM could activate transcription. This

work also provides a full pipeline to analyse the expression profiles and DNA methylation in complex hybrids, which could be crucial for anticipating varieties resistant to diseases and the current threats affecting citriculture such as the Huanglongbing disease.

INTRODUCTION

Citrus fruits (*Citrus spp*.) are among the most economically and culturally significant fruit crops worldwide, with global production estimated at approximately 140–146 million tons annually [1]. Their exceptional phenotypic diversity in fruit size, colour, flavour, essential oil composition, and nutritional value originates from a limited number of sexually compatible ancestral taxa - *C. medica*, *C. maxima*, *C. reticulata*, *C. micrantha*, and *C. cavaleriei* - whose interspecific hybridisations, followed by clonal propagation *via* apomixis and grafting, gave rise to the multitude of cultivated varieties ([2];[3];[4]). While most commercial cultivars derive from combinations of these Asian ancestral taxa, recent phylogenomic analyses have led to the integration of formerly distinct genera such as Microcitrus and Eremocitrus into the genus *Citrus* ([2];[4]). These Oceanian lineages, including *C. australasica*, are genetically more distant from the Asian citrus gene pool, yet remain largely unused in modern breeding programs despite their potential to introduce novel adaptive traits [5].

Modern citriculture now faces an alarming convergence of challenges that threaten both productivity and sustainability. Among these, Huanglongbing (HLB), caused by *Candidatus Liberibacter spp*. and vectored by psyllids, has devastated citrus production in major growing regions, including Brazil, the United States, and China ([6];[7]). Alongside HLB, producers must cope with the intensification of abiotic stresses such as drought, salinity, and thermal extremes, as well as an increasing burden of viral and fungal pathogens ([8];[9]). These pressures are exacerbated by the narrow genetic base of most cultivated varieties, which limits the availability of resilient alleles within the traditional gene pool. As a result, breeding efforts are increasingly turning toward exotic and underutilized germplasm to overcome these constraints and ensure long-term sustainability of citrus production ([10];[11]).

In this context, wide interspecific hybridization within the expanded *Citrus* genus offers a promising avenue for enhancing the adaptive potential of cultivated citrus. Crosses involving phylogenetically distant species such as *C. australasica* (formerly *Microcitrus*), *C. glauca* (formerly *Eremocitrus*), or *C. trifoliata* (formerly *Poncirus*) now all taxonomically integrated within *Citrus* have yielded germplasm with unique traits, including drought tolerance, cold hardiness, and partial HLB tolerance ([12];[13];[14]). Despite their potential, these lineages remain largely untapped in commercial breeding programs. Notably, *C. australasica* diverged from the main Asian citrus lineages during a distinct radiation event in the early Pliocene, approximately 4 million years ago [2], making it a particularly informative model to investigate the genomic consequences of wide hybridization.

Chromatin modifications, and especially DNA methylation, can impact gene expression *via* different pathways. In plants, DNA methylation occurs in three cytosine sequence contexts: CG, CHG, and CHH (where H = A, T, or C). CG methylation is maintained primarily by METHYLTRANSFERASE 1 (MET1), while CHG methylation depends on CHROMOMETHYLASE 3 (CMT3). CHH methylation is guided by both the RNA-directed DNA methylation (RdDM) pathway, involving DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), and the CMT2 pathway. These distinct but partially overlapping mechanisms ensure the establishment and maintenance of methylation across the genome, usually with a repressive role although it can vary depending on the context [15]. Together with other chromatin modifications such as histone post-translational modifications, DNA methylation patterns shape the epigenome. In the case of hybrids, it is not just two different genomes that are associated, but two different epigenomes, thus carrying both genetic and epigenetic variation.

Interspecific hybridization, is a common process in plants which enables the apparition of new phenotypes and species [16]. In breeding, this technique is widely used especially when hybrid vigor, also known as heterosis, is displayed in the F1 sibling. Heterosis is usually characterized in F1 by an increased biomass production and stress tolerance [17]. Recent sequencing approaches revealed that hybridization could be considered as a genomic and an epigenomic shock that, *in fine*, lead to the establishment of an hybrid specific transcriptional program which is not simply an addition of each parental transcriptional program [18]. In *Arabidopsis thaliana*, in rice, in maize and in cotton, DNA methylation and siRNA populations are altered in the hybrid. Changes to the hybrid methylome can occur in a way that differential methylation patterns inherited from the parent could be compensated to resemble each other, potentially *via* the action of small RNA [19]. However, this observation needs more evidence.

Creating hybrid resistant to pathogens, with a fruit pomology reaching the standard for citriculture is necessary to save a sector in great danger due to the various threats, in particular to diseases such the aforementioned HLB. The aim of this work is therefore to understand and, in the long term, to be able to anticipate gene expression profiles in a complex hybridization context where many layers of genetic and epigenetic regulation intermingle to give a particular expression profile. To better understand these regulatory pathways, we analysed the expression profiles in a hybrid between two species separated by more than 4 million years of evolution, *Citrus reticulata* (mandarin) and *Citrus* (formerly *Microcitrus*) *australasica* (finger lime). In addition, to determine the impact of modifications to epigenetic signatures such as DNA methylation on this expression profile, we analysed the overall methylation of the genomes, but also of the associated small RNAs that enable them to be deposited in certain contexts. In the end, this study shows that these pathways participate in the establishment of differential allelic expression within the hybrid, but also that methylation seems to play a particular role within citrus.

RESULTS

Assembly of an inter-specific citrus hybrid with an haplotype resolution

Although no direct genome comparison was performed prior to assembly, the two parental lines (*C. reticulata* and *C. australasica*) are known from phylogenetic studies to have diverged more than 4 million years ago [2]. This long evolutionary separation suggests substantial sequence divergence between their genomes, likely including structural variation and high heterozygosity in their hybrid offspring. We therefore performed haplotype-resolved assembly rather than a conventional consensus assembly as previously published as trio-binning [20]. We generated long-read sequencing using Oxford Nanopore Technologies of the hybrid individual with a 160X genome coverage using leaf tissue. To enable the assignment of hybrid long reads to their respective parental origins prior to assembly, we also obtained parental-specific k-mers from Illumina Tell-Seq libraries generated for each parent, as well short-read sequencing. This strategy allowed the realization of accurate trio-binning and independent reconstruction of both haplotypes (Figure 1A). Each assembled haplotypes consisted of only nine contigs, corresponding to the 9 chromosomes with a total assembly size of 349 Mb for *C. reticulata* and 359 Mb for *C. australasica*. The assemblies were highly contiguous (N50 = 38.0 Mb and 39.9 Mb, respectively; L50 = 4 for both) and exhibited consistent GC content (~34%).

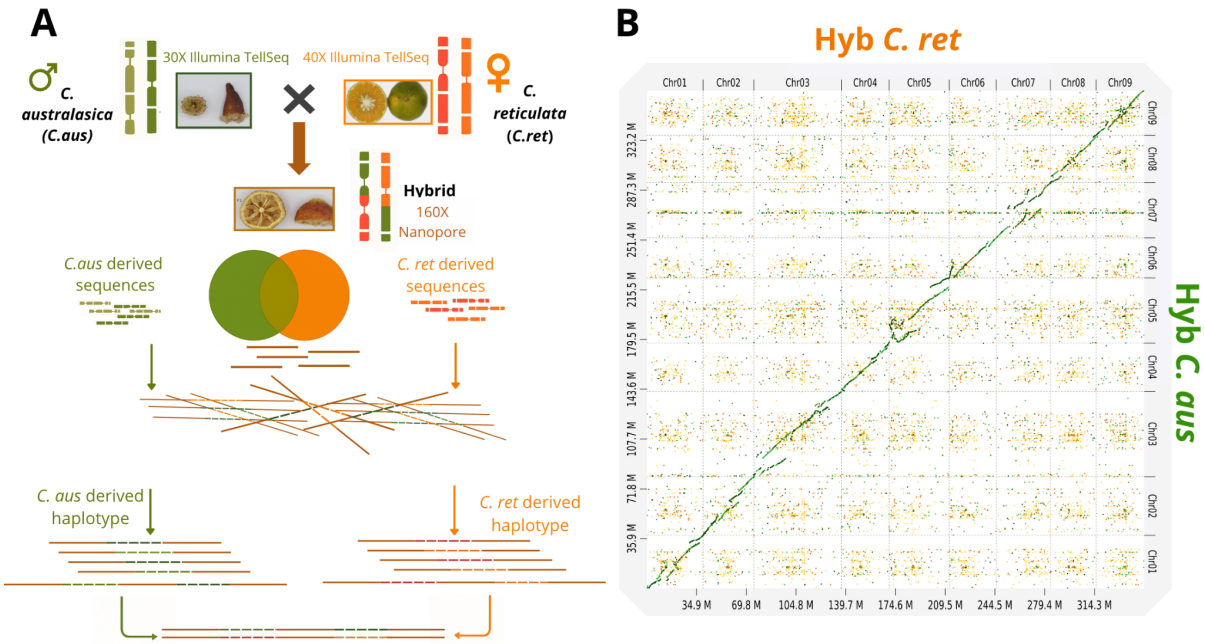

Figure 1. Haplotype-resolved genome assembly and structural variation.

*(**A**) Overview of the trio-binning genome assembly strategy used to generate haplotype-resolved assemblies of the citrus hybrid. Long reads from the hybrid were partitioned based on parental short-read data to generate separate assemblies corresponding to the maternal*

*(C. reticulata, orange) and paternal (C. australasica, green) haplotypes. (**B**) Dot-plot alignment between the maternal (Hyb C. ret) and paternal (Hyb C. aus) genome assemblies showing overall synteny and structural rearrangements. Diagonal lines indicate homologous regions; off-diagonal signals reflect structural variations between haplotypes.*

To assess the assembly quality of both haplotypes, we anchored their pseudomolecules to a consensus *Citrus* genetic map [21], constructed from seven *Citrus* species, including *C. australis* (an Oceanian species closely related to *C. australasica*) and *C. reticulata*. This anchoring was performed using the locOnRef tool from the Scaffhunter toolbox [22] allowing for the evaluation of synteny and collinearity between our assemblies and the reference map (Figure S1). The anchoring of each haplotype revealed a high degree of collinearity, with most chromosomes showing a clear one-to-one correspondence between physical and genetic coordinates. Minor discrepancies, visible as local discontinuities or small inversions in the Circos plots (Figure S1), likely reflect structural variants or orientation inconsistencies. These are expected when the available reference genomes differ from the actual parental genotypes used in the cross. Such local differences may also result from the trio-binning strategy, which preserves haplotypic divergence and can highlight micro-rearrangements between evolutionarily distant parental lineages. The absence of widespread inter-chromosomal misplacements and the consistent anchoring of large chromosome-scale scaffolds across linkage groups strongly support the structural accuracy of the haplotype-resolved assemblies. These results confirm that the assembly strategy, despite the unavailability of the actual parental genome sequences, produced high-quality haplotypes with conserved chromosome architecture and minimal technical artifacts.

The two haplotypes' genomes are highly syntenic, although several structural variants can be detected as shown by a dot plot analyses with the presence of numerous structural variations, such as an inversion in the middle of the chromosome 1, a translocation on chromosome 3 and a large duplication on chromosome 5, for example (Figure 1B). Gene prediction *via* BUSCO analysis revealed high completeness for both haplotypes (98.4% for *C. reticulata* and 98.3% for *C. australasica*), with the majority of BUSCOs present as unique or duplicated gene models (see Table 1).

Statistic	***C. reticulata* Haplotype**	***C. australasica* Haplotype**
Genome size (Mb)	349	359
Number of contigs	9	9
N50 (Mb)	38.0	39.9
L50	4	4
%GC	34.13	34.10

Statistic	*C. reticulata* Haplotype	*C. australasica* Haplotype
Genome size (Mb)	349	359
BUSCO completeness (%)	98.4	98.3
Unique genes (%)	78.0	73.8
Duplicated genes (%)	19.7	24.5

***Table1**: Summary statistics for the contigs of the hybrid diploid genome assembled with a haplotype resolution*

Deciphering the differential allelic expression in the inter-generic hybrid

Previous studies carried out on different combinations of hybrid plants have systematically revealed the presence of differential allelic expression to a greater or lesser extent within the hybrid compared with the parents, often for several hundred genes. To determine the situation within the hybrid between *C. australasica* and *C. reticulata*, we carried out a transcriptomic analysis in biological triplicate for the hybrid and each parent using leaf tissue. Within the hybrid, the transcripts were mapped onto the haplotype genomes in order to obtain an allelic resolution of the expression level of each of the genes.

To analyse differential gene expression between the two sub-genomes, we identified a common gene pool. Gene models were transferred from the reference *CITRE* [23] to each haplotype assembly using Liftoff [24]. Liftoff is a stand-alone lift-over tool that takes as input the reference genome sequence (FASTA), the target genome sequence (FASTA), and the reference gene annotation (GFF/GTF) even if assembled genomes accumulate structural variation and divergence. It uses Minimap2 [25] to map each reference gene's sequence onto the target assembly, and for each gene selects the set of exon alignments that maximizes sequence identity while preserving the exon–intron structure. Using this strategy, 95.0% of the (*CITRE*, ([23]) reference genes (33,055 out of 34,798) were successfully transferred to the maternal haplotype, and 89.0% (30,980 out of 34,798) were annotated on the paternal haplotype. Among these, 30,980 orthologous genes were found in both haplotypes and represent our pool of common genes.

We were thus able to determine which genes accumulated transcripts differentially, and counted 1758 genes in which the contribution of the allele derived from *C. australasica* was at least 3 times greater, and 1809 genes in which the contribution of the allele derived from *C. reticulata* was 3 times greater within the hybrid (Figure 2A).

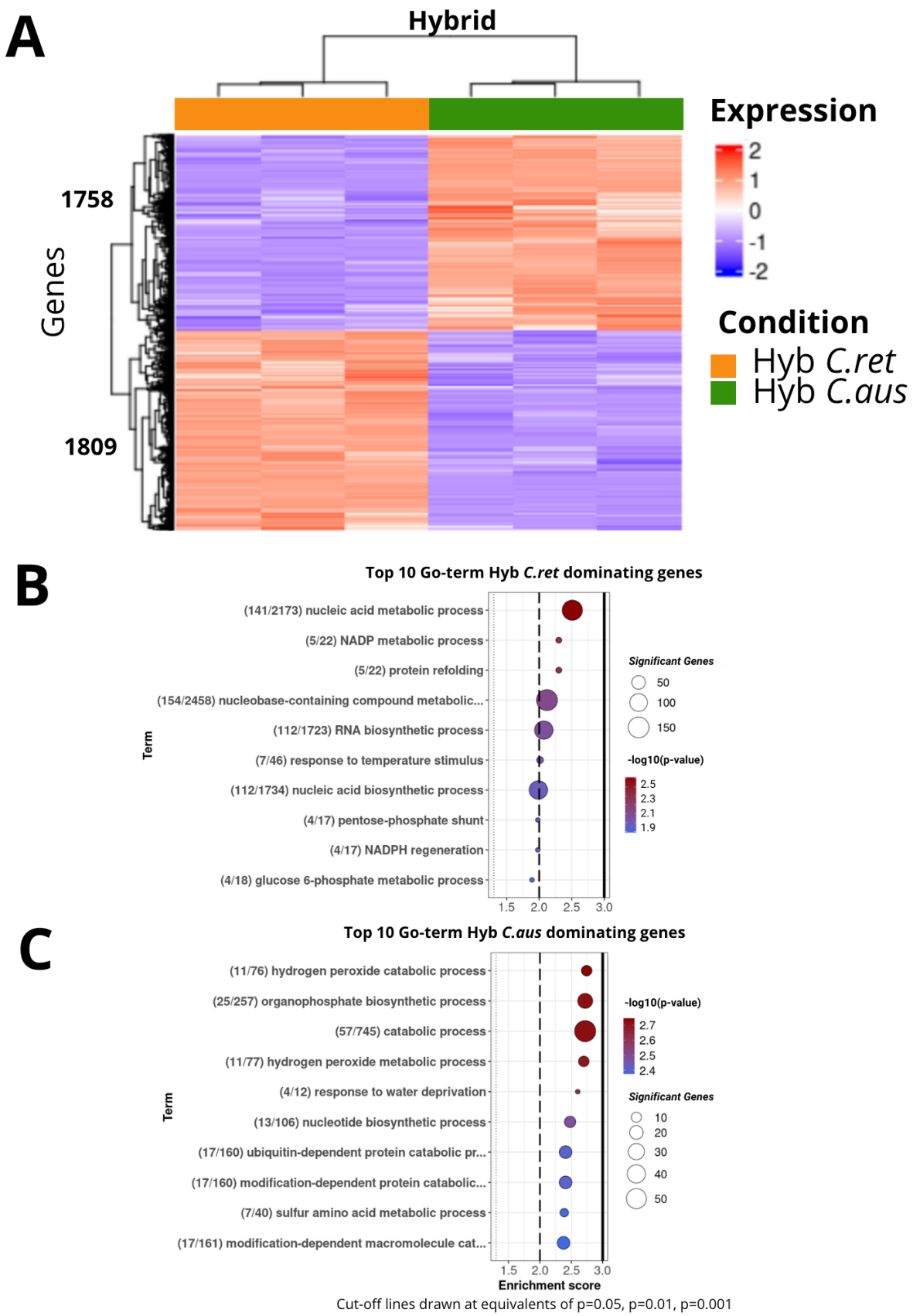

Figure 2. Allele-specific expression and functional enrichment in the hybrid.

*(**A**) Heatmap showing allele-specific expression between the two haplotypes in the hybrid. Expression values are variance-stabilized counts (VST) obtained from DESeq2 and centred*

*gene-wise (z-score). Rows represent genes; columns represent biological replicates of the two hybrid haplotypes. The color gradient reflects standardized expression levels: red indicates higher expression, blue indicates lower expression (unit: log_2-transformed normalized expression). (**B**) Bubble plot of enriched GO Biological Process terms among genes upregulated on the maternal haplotype. Bubble size represents gene count; colour reflects enrichment p-value. (**C**) Bubble plot of enriched GO terms among genes upregulated on the paternal haplotype, plotted as in panel (**B**).*

We then determined whether certain classes of genes were particularly affected by this differential expression by determining changes in transcript accumulation according to gene ontology (GO term). Several groups of genes were found to be significantly differentially expressed within the hybrid. The *C. reticulata* haplotype is dominant for genes involved in metabolic processes for nucleic acid and NADP, while the *C. australasica* haplotype is dominant for genes implicated in catabolic processes particularly hydrogen peroxyde, as well as in organophosphate biosynthetic processes (Figure 2B and 2C and Figure S2).

Such dominance at the GO term scale implies a bias implication of one of the parents in the hybrid in the regulation of certain metabolic and physiological pathways that could explain partially the phenotype observed in the hybrid.

Finally, we identified 289 and 268 genes in the hybrid that are exclusively expressed either by the *C. australasica* or the *C. reticulata* haplotypes respectively (Figure S3). Again, these numbers are very similar, demonstrating a somewhat overall balance between the two haplotypes. In both cases, the majority of these genes are only expressed in the corresponding parental lines (Figure S3).

DNA methylation landscape in hybrid haplotypes reveals context- and parent-specific reprogramming

To determine whether some of this differential allelic expression may be the result of changes operating at the epigenetic level, we analysed cytosine methylation globally for the hybrid and each parent using leaf tissue. Whole-genome Bisulfite sequencing (WGBS) approaches were used and sequenced using 150bp paired-reads with depth of coverage equivalent to 20 times the average citrus genome size in biological triplicate.

Overall levels of DNA methylation revealed distinct patterns across the three cytosine contexts (Figure 3A). Methylated CG (mCG) were the most abundant, followed by methylated CHG and methylated CHH (mCHG and mCHH). Both hybrid haplotypes displayed a significant reduction in CG methylation relative to their respective parents, with Hyb *C. ret* showing a 5.2% decrease and Hyb *C. aus* a 5% decrease compared to *C. reticulata* and *C. australasica*, respectively. In contrast, mCHG levels were relatively stable across hybrids and parents.

Notably, mCHH showed a haplotype-specific reprogramming: while Hyb *C. ret* maintained mCHH levels comparable to *C. reticulata*, Hyb *C. aus* exhibited a clear increase compared to *C. australasica*, suggesting targeted activation of RdDM in the paternal haplotype.

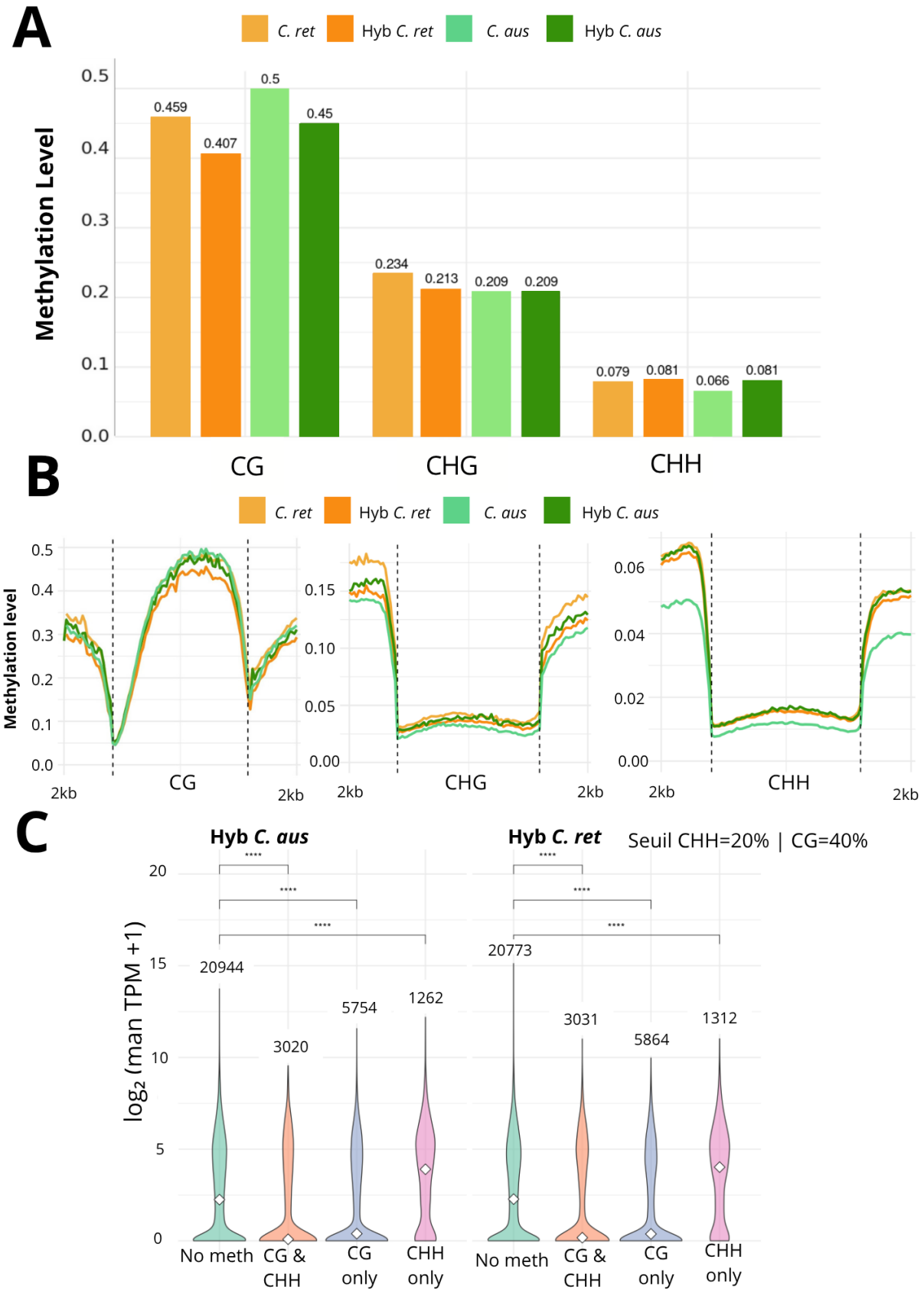

Figure 3. DNA methylation profiles and correlation with gene expression.
(A) *Global DNA methylation levels (in %) in CG, CHG, and CHH contexts across parental haplotypes (C. reticulata, C. australasica) and their corresponding haplotypes within the hybrid. Values indicate mean methylation levels across all annotated genes.* ***(B)*** *DNA methylation metaprofiles (in %) across gene bodies and ±2 kb flanking regions in each cytosine context. Curves represent average methylation profiles for each genotype and context.* ***(C)*** *Violin plots showing gene expression levels (log_2(mean TPM + 1)) according to promoter methylation status in CG and CHH contexts in the two haplotypes of the hybrid. Promoters are categorized into four classes: "No methylation", "CG & CHH", "CG only", and "CHH only" based on thresholds (CHH ≥ 20%, CG ≥ 40%). Statistical differences in expression across categories were assessed using a Kruskal–Wallis test ($p < 0.05$), followed by pairwise Dunn's post-hoc comparisons with Benjamini–Hochberg correction. The number of genes in each category is indicated above each violin. Asterisks denote significance levels from post-hoc comparisons: * = padj < 0.05, ** = padj < 0.01, *** = padj < 0.001, **** = padj < 1e-4.*

At the chromosomal scale, genome-wide methylation distribution showed mCG and mCHG highest in pericentromeric regions, correlating with repetitive element abundance (Figure S4) [26]. Methylated CHH was generally lower but showed local peaks more pronounced in Hyb *C. aus* than Hyb *C. ret*, supporting reinforced RdDM activity targeting the *C. australasica* haplotype.

We next analysed DNA methylation patterns in a 1 kb window surrounding gene bodies. In CHG and CHH contexts, methylation dropped sharply within gene bodies, with elevated levels in flanking regions (Figure 3B). mCHH around genes was markedly higher in *C. australasica* haplotype than in *C. reticulata* haplotype, reinforcing the idea of RdDM-driven reprogramming is more pronounced on the *C. australasica*-derived haplotype. Levels of mCG between the hybrid haplotypes were minor at the genic level, but a modest decrease in the *C. reticulata* haplotype was noted in promoter-proximal regions, whereas gene-body mCG remained relatively stable and high in both hybrid sub-genomes. These patterns reflect the complex locus- and context-specific DNA methylation reprogramming reported during cellular differentiation and hybridization ([27];[28]) Direct comparisons between each hybrid haplotype and its respective parent confirmed an asymmetric pattern of methylation reprogramming (Figure 3B). Hyb *C. aus* showed increased mCHH compared to *C. australasica*, particularly in gene-flanking regions, while mCG was largely conserved. Hyb *C. ret* exhibited slightly elevated mCG relative to *C. reticulata*, with stable mCHH levels. This supports models where epigenetic remodelling in hybrids is haplotype-specific and influenced by parental origin, consistent with findings that DNA methylation remodelling in hybrids is often driven by epigenetic interactions and small RNA pathways [29].

We then looked at the impact of this methylation, whatever its context, on gene expression within the hybrid, with the idea of evaluating this type of epigenetic modification on the establishment of the differential allelic expression previously observed. For each haplotype, we analysed the relative expression of the genes when their promoter (- 2kb before TIS to ATG) was either unmethylated, or methylated at CG (at least 40%), or methylated at CHH (at least 20%), or both. As expected, genes whose promoters accumulate only mCG or both mCG and mCHH are strongly correlated with a low level of expression (Figure 3C), whatever the haplotype, compared to non-methylated promoters. In contrast, promoter regions only accumulating mCHH associate significantly with genes with a higher expression level than those without mCG only, both mCG and mCHH or in comparison with those that are unmethylated (Figure 3C). We also confirmed this result by analysing numerous genes where indeed high gene expression was not incompatible with the presence of small clusters and mCHH accumulation in their promoter (Figure S5). Taken together, our data reveal the presence of significant changes at the methylation levels, with contrast impact on gene expression depending on the cytosine context.

Small RNA population and impact in the hybrid

In plants, mCHH accumulation in promoter regions is the result of the action of RdDM and RNA polymerases IV and V, which are specific to plants and enable the generation of small 24 nt RNA (hereafter named siRNA). Associated with an Argonaute protein (AGO4 in *A. thaliana*), these siRNAs are then able to target sites for CHH methylation *via* the action of DRM2 [30]. Therefore, we expect that at least some of the promoters accumulating CHH methylation will also accumulate siRNA. To test this hypothesis, we performed small RNA-seq analyses for the hybrid and each parent using leaf tissue in triplicates.
Small RNA were grouped according to their size, showing the presence of two major types of small RNA: those of 21-nt, corresponding to microRNA, and those of 24-nt, which are more abundant and correspond to siRNA (Figure 4A) [31]. Overall, we can see that the two haplotypes participate equally in the production of these smallRNAs in quantitative terms (Figure 4B).

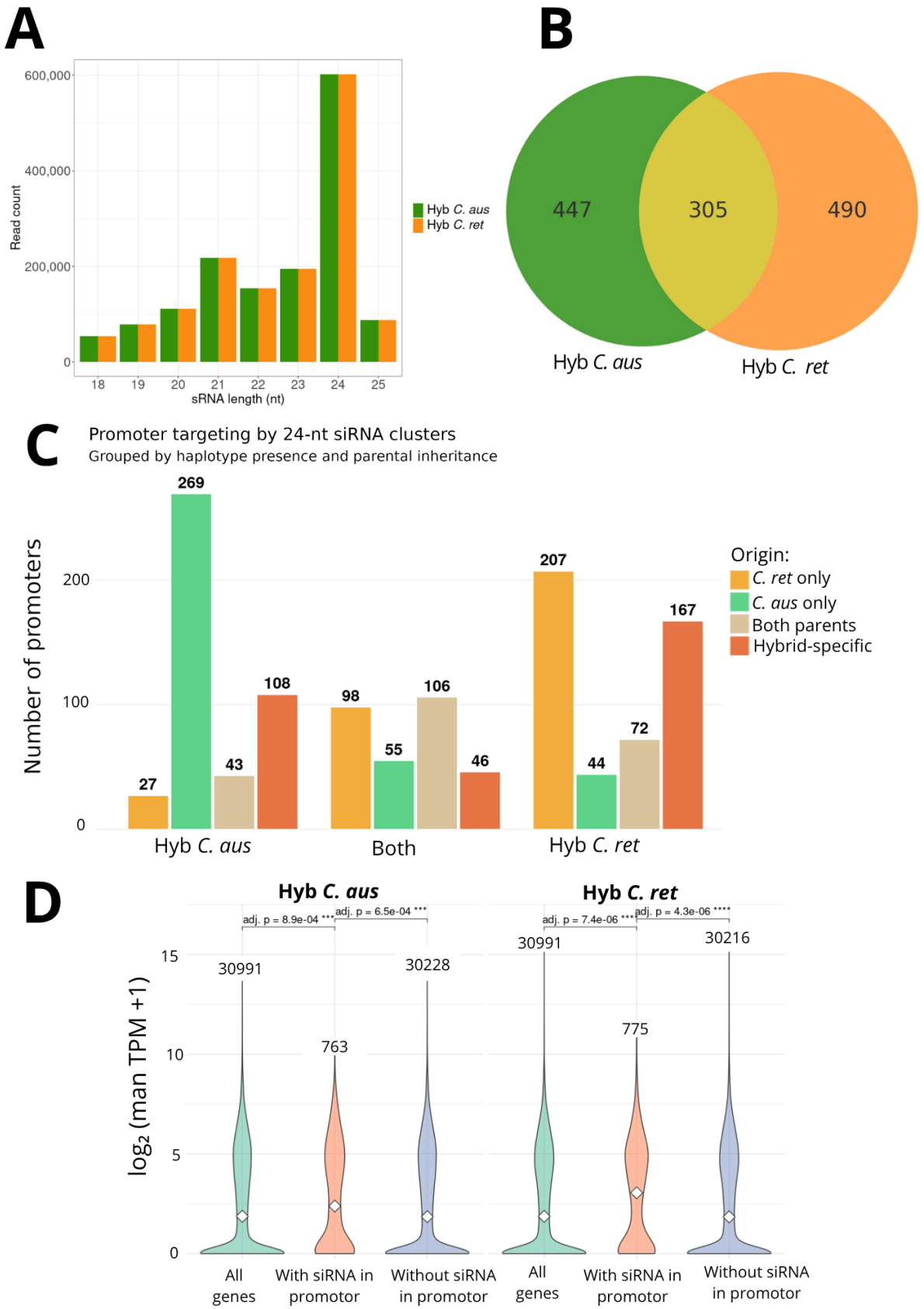

Figure 4. Haplotype-specific distribution and effects of 24-nt siRNA clusters.

(A) *Size distribution of mapped small RNAs in each hybrid haplotype.* ***(B)*** *Venn diagram showing the overlap of 24-nt siRNA clusters between the two haplotypes of the hybrid. The*

intersection represents clusters targeting the same promoters in both haplotypes. ***(C)*** *Classification of gene promoters targeted by 24-nt siRNA clusters in the hybrid, based on parental origin. Promoters are grouped according to whether they are targeted by siRNA clusters inherited from C. reticulata only (orange), C. australasica only (green), both parents (beige), or are hybrid-specific (red).* ***(D)*** *Violin plots showing gene expression levels (log_2(TPM+1)) depending on the presence or absence of 24-nt siRNA clusters in promoters, for each hybrid haplotype. Genes with siRNAs in their promoters are significantly less expressed than those without (Wilcoxon test, adjusted p-values indicated).*

We then focused on the siRNA, responsible for targeting genomic regions in CHH methylation *via* the RdDM pathway. To analyse whether these siRNA target the same regions, we mapped them on the haplotype-resolved hybrid genomes. A third of the individual siRNA originates from both sub-genomes, while the other two-thirds originate either from *C. reticulata* or *C. australasica*-derived haplotype (Figure S6). Because of the important sequence variation existing between the two haplotypes, we decided to question their functional impact once grouped into clusters when their coverage peaks were located within a 200-nucleotide window, as defined by ShortStack. Promoters targeted solely by CHH methylation are expected to be targeted by siRNA which then participate in the recruitment of DRM2 at promoter regions, then genes whose promoters are targeted by siRNA should also be more transcribed overall. We therefore focused on siRNA clusters targeting promoter regions.

We therefore focused on the siRNA clusters targeting only the promoter regions (1kb before ATG). Once again, the majority of siRNA groups were produced by only one of the two haplotypes within the hybrid. Only 25% were produced simultaneously by both haplotypes in the hybrid (Figure 4B). Finally, we analysed the global expression of the genes targeted or not by siRNA clusters in their promoter and demonstrated that those with targeted promoters accumulate more transcribed than those non-targeted in their promoter. Altogether, these results demonstrate not only that the contribution of siRNAs is essentially haplotype-dependent within the hybrid, but also that the targets of siRNAs tend to be more transcribed, supporting the hypothesis of a potential involvement of RdDM in transcriptional activation within the hybrid.

Impact of RdDM on differentially expressed genes in the hybrid

To explore the impact of hybridization on the interrelation between gene expression, DNA methylation, and siRNA production, we integrated our epigenomic data at the haplotype level (Figure 5A–B). We then performed several statistical analyses of gene expression within the hybrid, quantified as TPM.

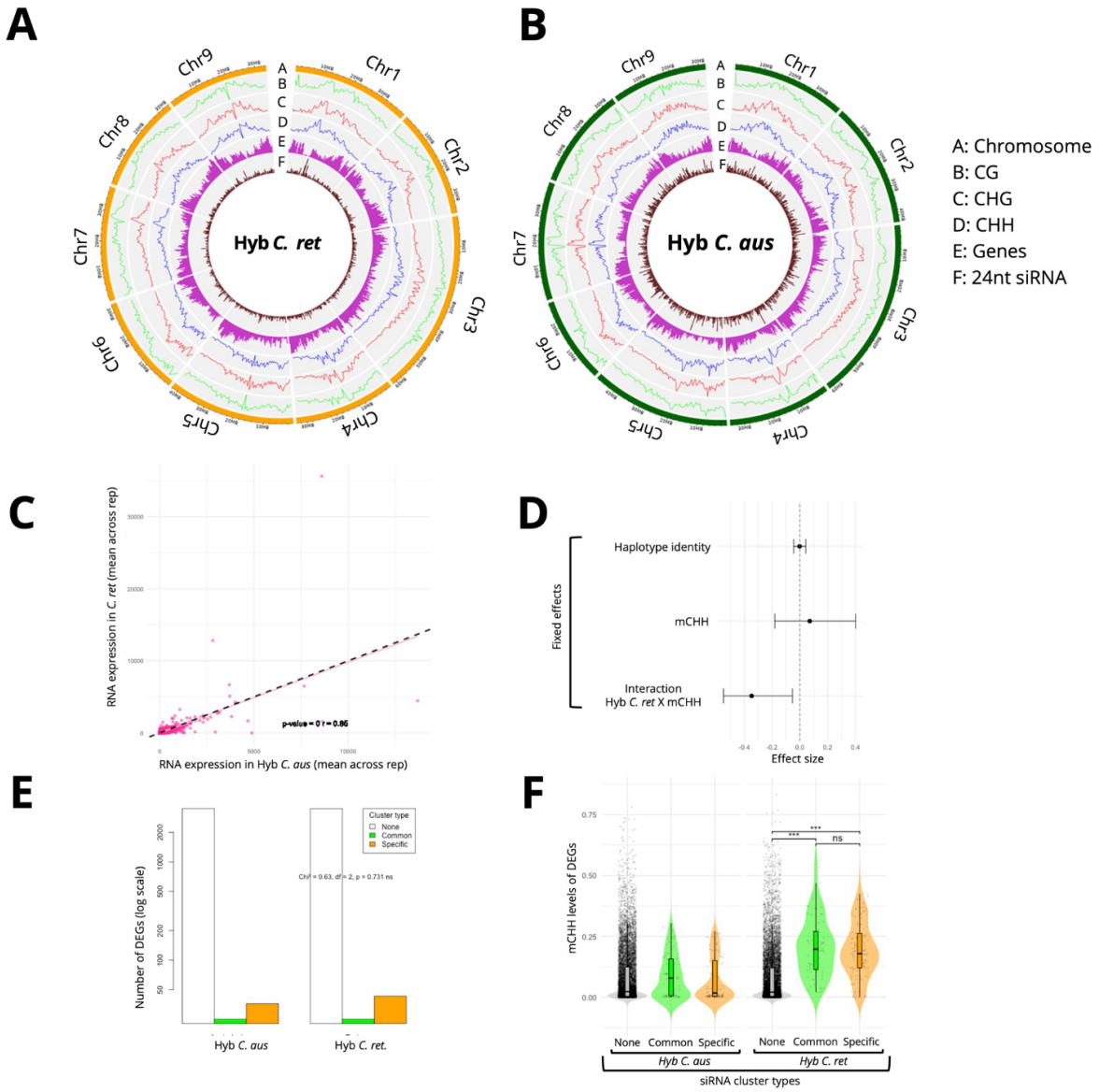

Figure 5. Genome-wide epigenomic patterns and overlap with expression.

*(**A–B**) Circos plots representing genome-wide profiles of DNA methylation, gene density, and 24-nt siRNA clusters across the nine chromosomes of the two haplotypes of the hybrid: (**A**) C. reticulata-derived haplotype (orange) and (**B**) C. australasica-derived haplotype (green). Tracks from outer to inner rings: A : Chromosomes; B–D : DNA methylation levels in CG (red), CHG (blue), and CHH (purple) contexts; E : Gene density; F : Density of 24-nt siRNA clusters. (**C**) Spearman's rank correlation in gene expression between haplotypes. (**D**) Effect sizes (±95% CI) from a linear mixed model evaluating the impact of haplotype, mCHH levels, and their interaction on DEGs expression (TPM). Contrasts are treatment-coded with C. australasica haplotype as the reference. CIs overlapping zero indicate non-significance. (**E**) Distribution of differentially expressed genes by siRNA cluster types and haplotypes (None = no cluster, Common = shared, Specific = haplotype-specific). (**F**) mCHH levels across small RNA cluster types by haplotype. Dunn's test with Bonferroni correction: Significance: *** p ≤ 0.001, ** p ≤ 0.01, * p ≤ 0.05, ns = not significant.*

First, we tested whether gene expression levels were correlated between haplotypes using Spearman's rank correlation in R. We observed a strong and highly significant correlation (Spearman's $\rho = 0.85$, $p < 2.2 \times 10^{-16}$), closely following a 1:1 relationship (Figure 5C), suggesting that most genes exhibit similar expression levels across the two haplotypes. We then tested whether the 3,567 DEG within the hybrid were influenced by haplotype identity, mCHH levels, or the interaction between the two. We used a linear mixed-effects model with haplotype identity, mCHH and the interaction included as fixed effects, while the first principal component (PC1) from a PCA of gene expression profiles was included as a random effect. This approach captures underlying clustering or co-expression patterns between genes that could confound the interpretation of haplotype and methylation effects. The model revealed no overall differences in gene expression of DEGs between haplotypes, confirming that neither haplotype systematically dominates the other in terms of expression of DEGs (Top; Figure 5D). It also showed that mCHH alone had no consistent effect on gene expression on DEGs (Middle; Figure 5D). However, a significant interaction between haplotype identity and mCHH levels was detected (Bottom; Figure 5D): a marked reduction in gene expression was specifically observed when mCHH occurred on the *C. reticulata* haplotype compared to *C. australasica* haplotype. This suggests that the effect of mCHH on gene expression of DEGs is haplotype-dependent.

We then investigated the impact of hybridization on siRNA clusters on promoters of DEGs. We categorised clusters in three classes: "Specific" (detected in only one haplotype), "Common" (shared between both), and "None" (genes with no detected cluster in their promoter). We counted the number of DEGs per category and haplotype, and used a Pearson's Chi-squared test of independence to assess differences in distribution. The distribution of siRNA cluster types did not significantly differ between haplotypes ($Chi^2 = 0.63$, df = 2, $p = 0.73$). Most DEGs lacked a detectable cluster in both haplotypes, and the proportions of "Common" and "Specific" clusters were similar across haplotypes (Figure 5E). These results suggest that the presence or specificity of siRNA clusters among differentially expressed genes is similar across haplotypes.

Finally, to test whether mCHH levels differed on promoter regions targeted by siRNA clusters according to their origin, we performed another test, still focusing on DEGs. Since methylation data did not follow a normal distribution, we used non-parametric Kruskal-Wallis rank sum tests to compare methylation levels across cluster types ("None", "Specific", "Common") separately for each haplotype. In the *C. australasica* haplotype, mCHH levels did not differ significantly between cluster types ($\chi^2(2) = 2.35$, $p = 0.31$). No pairwise comparisons were significant after correction. In contrast, *C. reticulata* haplotype showed a significant effect of siRNA cluster type on mCHH ($\chi^2(2) = 232.25$, $p < 2.2\times10^{-16}$). Post hoc tests revealed that DEGs with no associated siRNA cluster had significantly lower methylation than those with either "common" or "specific" clusters ($p < 0.001$ for both comparisons). Moreover, there was no significant difference between "common" and "specific" clusters ($p = 0.67$ after Bonferroni correction). These results suggest that in *C. reticulata* haplotype, the presence of a siRNA

cluster—regardless of whether it is shared or specific—is associated with an increase in mCHH, whereas in *C. australasica* haplotype, mCHH is not significantly affected by cluster occurrence (Figure 5F).

Association of RdDM with expressed genes is not linked to hybridization

Our strategy revealed that while for the majority of genes, the parental contribution to transcript accumulation within the hybrid was balanced, around 10% of common genes were dominated by the contribution of just one of the two parents. This dominance is even more marked for siRNA, where only a third are expressed by both haplotypes.

Genome-wide analyses showed that cytosine methylation, across CG, CHG, and CHH contexts, accumulates in gene-poor regions near the chromosomal centers, likely corresponding to centromeric or pericentromeric regions, despite centromeres not being clearly annotated in Citrus [32]. Both haplotypes display consistent patterns of mCHH and siRNA accumulation in these regions, indicative of canonical RdDM activity, potentially targeting transposable elements. The strong co-localization of mCHH and siRNAs in pericentromeric regions further supports their coordinated role in heterochromatin maintenance, although siRNA-independent mCHH accumulation may also be involved ([33] ; [34]). Despite the interspecies nature of the hybrid (*C. reticulata* × *C. australasica*), the epigenomic landscapes of both haplotypes show remarkable global symmetry. This conserved chromosomal organization suggests that genome compartmentalization and RdDM activity are stable features across divergent Citrus lineages and may contribute to the overall transcriptional stability observed in the hybrid (Figure 5A–B), although we identified a potential haplotype-bias impact in the hybrid for DEG (Figure 5F).

To assess whether RdDM is linked to transcriptional activity, we analysed gene expression levels in relation to both mCHH accumulation and siRNA presence. As described above, analysis of expression data with either cytosine methylation or the presence of siRNA revealed that genes were more highly expressed when they either accumulated mCHH (Figure 3) or were targeted by siRNA (Figure 4). To see if these two aspects are linked, we analysed all the genes targeted by siRNA and that also accumulate at least 20% mCHH in their promoter region, separating the expressed genes (Figure 6A) from the genes that are little or no expressed (Figure 6B). Remarkably, genes whose promoters accumulate these two characteristics are 10 times more likely to be associated with expressed than unexpressed genes. However, these targeted genes are only targeted within the hybrid in at least half the cases: in 49% of cases (93 promoters out of 191) when the genes are associated with expression (Figure 6A) and in 77% of cases (24 promoters out of 31) when the genes are not or only slightly expressed (Figure 6B).

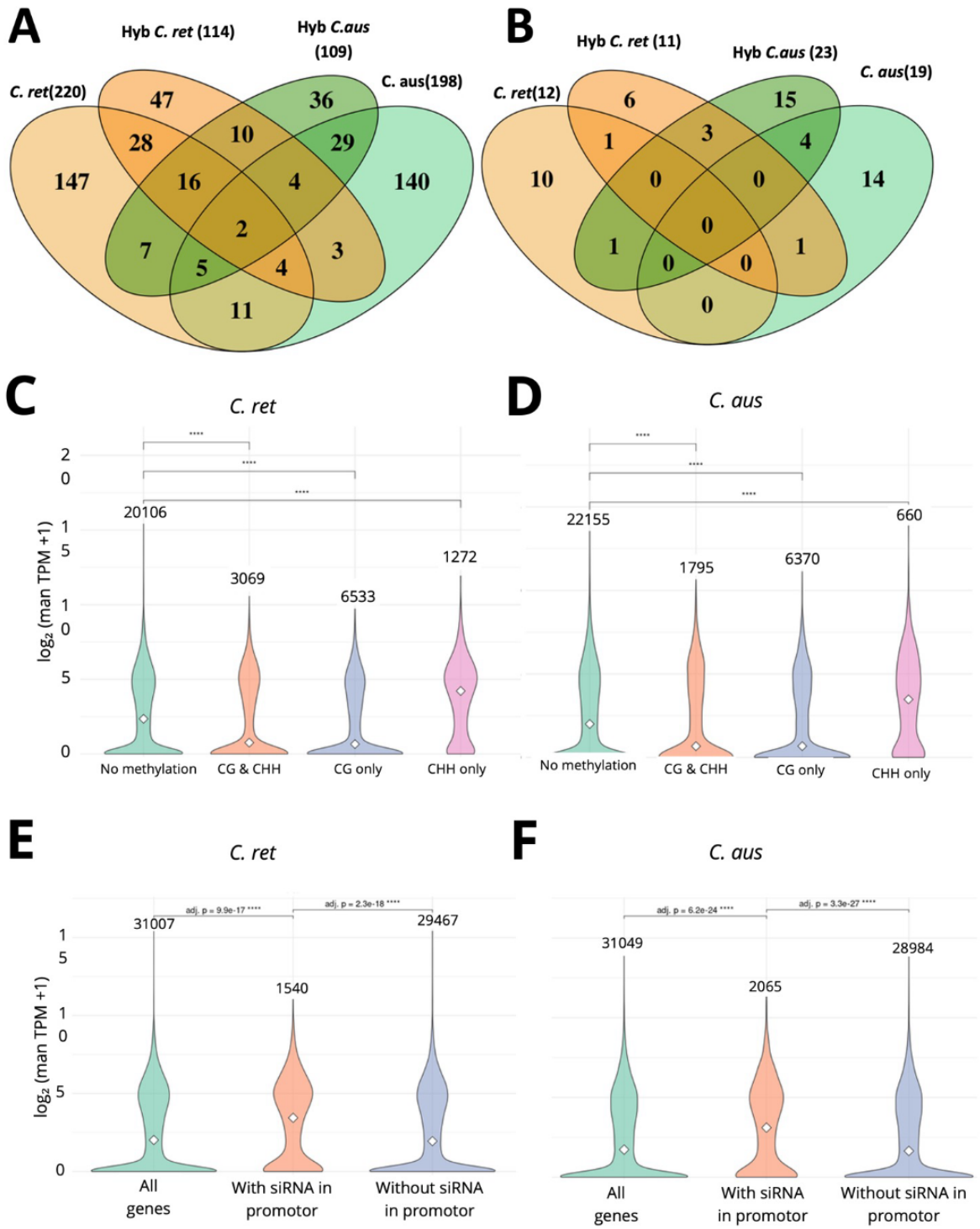

Figure 6. Epigenetic regulation of gene expression in the parental genomes.

*(**A**–**B**) Venn diagrams showing the overlap of genes whose promoters are targeted by 24-nt siRNA clusters across parental and hybrid haplotypes. (**A**) Expressed genes (TPM > 1) with promoter siRNA clusters. (**B**) Non-expressed genes (TPM ≤ 1) with promoter siRNA clusters.*

Each subset indicates the number of genes whose promoters are targeted by siRNAs in one or more haplotypes. ***(C–D)*** *Violin plots showing gene expression levels (log_2(mean TPM + 1)) in parental genomes:* ***(C)*** *C. reticulata,* ***(D)*** *C. australasica. Genes are grouped according to promoter methylation status: No methylation, CG only, CHH only, and CG & CHH, using thresholds of 40 % CG and 20 % CHH methylation. Numbers above each plot indicate the number of genes in each category. Asterisks indicate significant differences (Wilcoxon test, **** = p < 1e-4).* ***(E–F)*** *Gene expression comparison based on the presence or absence of 24-nt siRNA clusters in promoters for:* ***(E)*** *C. reticulata,* ***(F)*** *C. australasica. Wilcoxon rank-sum tests were used to assess statistical differences; adjusted p-values are indicated (Benjamini–Hochberg correction).*

To test whether RdDM targets are also associating with higher expression in the parental lines, we analysed the transcriptomic, DNA methylation and small RNA data in the parents, using the haplotype resolved reference genome of the hybrid (Figure S7). Interestingly, mCHH is also linked to high gene expression in the parental lines, which goes to show that this atypical correlation is not linked to hybridization *per se*, but would rather be a more general feature in *Citrus* genus.

To confirm this hypothesis, we perform the same analyses made on the hybrid haplotype individually on methylation and siRNA accumulation in the parental lines, and indeed confirm a similar association of expressed genes with siRNA cluster or with only mCHH both in *C. reticulata* (Figure 6C and E) and in *C. australasica* (Figure 6D and F). *In fine*, in both *C. reticulata* and *C. australasica*, genes with promoter accumulating mCHH and/or siRNA clusters were more likely to be expressed than genes without these marks, reinforcing the conclusion that this epigenetic pattern is conserved across species. Venn diagram analysis further showed partial overlap between mCHH and siRNA-targeted genes in each parent (Figure S8), suggesting that these two layers of RdDM regulation may act independently or additively.

Together, these results support a model in which RdDM-associated targeted genes are positively associated with gene expression in *Citrus*, and this association is conserved across divergent species.

DISCUSSION

In this study, we investigated the transcriptional and epigenetic consequences of an interspecific hybridization event between two phylogenetically distant *Citrus* species separated by over 4 million years of divergence: *C. reticulata* and *C. australasica*. By integrating high-quality haplotype-resolved genome assemblies with transcriptome, DNA methylation, and small RNA datasets, we provide new insights into the regulatory dynamics occurring in such distant hybrid. To analyse the citrus hybrid and its parent, we developed a pipeline adapted to the integration of genomic and epigenomic data with a haplotype

resolution using the strategy previously published in [20] and that resolves allelic variation before assembly. This trio-binning approach enabled the separation of two highly divergent haplotypes and produced chromosome-scale assemblies with high contiguity and collinearity (Figure1 and S1). The observed structural variants, including small inversions and local discontinuities, may reflect true polymorphisms between the actual parental genomes or arise from hybrid-specific chromosomal rearrangements. While the assemblies already provide a robust foundation for genomic comparisons, future polishing steps, for example through the integration of Hi-C data, could further improve scaffold orientation and help distinguish between biological rearrangements and minor assembly inconsistencies. More importantly, the haplotype-resolved assembly provided the basis for identifying differentially expressed genes with allelic resolution in the hybrid.

Hundreds of expressed genes accumulate transcripts mainly from only one of the parental-derived genomes, which implicate a certain dominance at the gene scale (Figure 2, S2 and S3). Although some metabolic or catabolic pathways are enriched among the dominant genes, we do not know at this stage the impact of this imbalanced expression. However, we can note that this imbalance seems to be somehow equilibrated between the two parent-derived genomes since a similar number of DEG are found in the haplotype-resolved hybrid transcriptome.

At the methylation level, the global epigenomic landscapes were remarkably conserved between haplotypes, closer inspection of Figure 5A–B reveals subtle but potentially meaningful differences in local siRNA accumulation. Discrete siRNA peaks specific to one haplotype, particularly on chromosomes 6 and 9, may reflect species-specific bursts of transposable element (TE) activity or structural rearrangements. These features could mark regions of differential genome organization between *C. reticulata* and *C. australasica,* but definitive interpretation will require high-quality TE annotation of each phased haplotype.

Quantitatively, if a global decrease in mCG and mCHG is observed in the hybrid compared to the parent, mCHH is slightly higher, suggesting a particular role or deregulation of the RdDM pathway. Interestingly, mCHH only significantly affects DEG derived from the *C. reticulata* haplotype (Figure 5D and S9). This haplotype-specific mCHH methylation is accompanied by a specific accumulation of siRNA in the *C. reticulata* haplotype promoter of DEG (Figure 5F). The reason why only one haplotype seems to be sensitive to RdDM is not clear. mCHH accumulation in the hybrid seems to be additive, reaching the highest level of the parent as shown in (Figure 3B). In consequence, we detected an increase in the level of mCHH of around 50% throughout the genome of the *C. australasica* haplotype compared to the parental line. This increase, in one generation, is considerable and could probably explain at least partially why mCHH do not have as much impact on gene transcriptional regulation in the *C. australasica* haplotype than in the *C. reticulata* haplotype, within the hybrid. One explanation could be that because of the sequence variation between the two haplotypes, mCHH readers

may act preferentially on their own haplotype genome and a 50% increase is probably due to deregulation and/or an imbalance that cannot be corrected in this generation, saturating mCHH reader belonging to the *C. reticulata* haplotype. Previous studies in *A. thaliana* hybrids demonstrated that altered DNA methylation and siRNA patterns exist in the hybrid, potentially contributing to the hybrid vigour phenotype [19]. However, in that case, an increased mCG and a decreased mCHH has been observed relative to the parents [35], which is the reverse of our observation here. Further transcriptional and epigenetic analyses on future generations will show whether this imbalance persists.

Our data suggest that the RdDM pathway is much more implicated in transcriptional activation, than repression in *Citrus*. Indeed, we identified hundreds of highly expressed genes in the hybrid, but also in the parental lines that accumulate both mCHH to a level higher than 20%, and that are also targeted by siRNA at their promoter. This number of targets is probably higher since it cannot be ruled out that the lack of detection of siRNAs clusters is not the result of transcriptional variation due to the transient nature of their production. These characteristics are typical of RdDM targets. In contrast, only a few low-expressed genes have at least one of these characteristics in the three different genomic citrus contexts. Generally, RdDM is strongly associated with gene transcriptional repression in *A. thaliana*, maize, tomato and other plant species ([36];[15]). However, such a paradigm has been recently challenged by recent studies. In rice, the presence of mCHH on transgenes and endogenous sequences is not always linked to gene repression [37]. In *Poncirus trifoliata*, a close relative of the citrus genus usually used as rootstock, highly expressed genes were more likely to accumulate mCHH sequence contexts in their promoter regions in this species [38]. The molecular mechanisms that explain the potential involvement of RdDM either in transcriptional repression or activation have been revealed in *A. thaliana*, *via* the identification of different readers of methylated CHH that belong to SUVH histone methyltransferases [39]. When mCHH is associated with repression, SUVH4, 5 and 6 are recruited. But when mCHH is associated with high expression, SUVH1 and SUVH3 are implicated instead [40]. These homologs also exist in citrus and are expressed in the parental and the hybrid plants. Analyses of additional citrus methylome especially in the CHH context but also on other plant species may reveal a bias in Citrus and potentially abroad where RdDM generates transcriptional activation rather than silencing.

CONCLUSIONS

Our study reveals the impact of DNA methylation and siRNA on the establishment of DEG in a complex inter-specific citrus hybrid. However, additional mechanisms can explain the presence of so many DEG in the hybrid. For example, structural variation including duplication, translocation and inversion contributes to the generated genomic diversity and gene copy number changes that can lead to phenotypic plasticity, including in crops [41].

However, their accurate detection requires a near perfect assembly, usually assisted with chromatin-chromatin interaction maps (Hi-C) approaches which is not our case here.
Mobile genetic elements such as transposable elements (TE) can also affect gene expression in two ways: either *via* creating structural variation *via* their neo-insertion, or by modulating gene expression in their vicinity ([42],[43], [44] [45]). Their presence in promoter regions can therefore impact siRNA production and inheritance in hybrid context [46] . However, in addition to a near perfect assembly, a fine-tuned TE annotation is required in that case and could be the subject of a future study on this hybrid. Finally, mono-allelic expression of certain genes in this complex hybrid may be the consequences of the evolutionary process and the accumulation of genetic and epigenetic modifications in the promoter regions, affecting transcription factors association. Indeed, distant by 4 million years of evolution, the co-evolutionary process between the transcription factors and their targeted regulatory elements in gene promoters may not be fully conserved between the two species.

In any case, this study proposes an innovative and effective analysis strategy for analysing the molecular consequences of hybridization between distantly related species within the citrus genus. We hope that this strategy will be used in the near future to demonstrate (*i*) the reproducibility of DEG profiles and the impact of DNA methylation and siRNA production in several independent F1 hybrids between these two distantly related species, *C. reticulata* and *C. australasica*, and (*ii*) the ability to combine genomic and epigenomic approaches on young citrus hybrids to anticipate their phenotype during their reproductive stage in terms of their ability to resist pathogens such as Huanglongbing [47], but also to produce fruit that meets the standards for marketing.

MATERIEL AND METHODS

Plant material

Hybrid F1 result from a cross between *Citrus reticulata* Blanco cv Fortune mandarin (female parent) and *Citrus australasica* cv Caviar Bachès (Male parent), located at the INRAE CIRAD station in San Giuliano, France.

Genome Assembly and Annotation

Nanopore reads from the hybrid were assigned to parental haplotypes using a trio-binning approach. K-mers specific to each parent were extracted from Tell-Seq data using Meryl. Reads were assembled separately for each haplotype using Canu v2.2 [48] with options *genomeSize=430m*. Initial polishing of the assemblies was performed using Medaka [49], followed by a second round with Pilon [50] using the mapped Tell-Seq reads and the options *--fix snps,indels --minmq 3*.

Gene annotation transfer

Gene models were transferred from the reference genomes using Liftoff [24] with the parameters *--copies --polish --exclude_partial,* ensuring accurate mapping of complete annotations.

Linear mixed-effects model

To account for non-independence in the data, we used a linear mixed-effects model, including the first principal component (PC1) from a PCA of gene expression profiles as a random effect. Replicates were also tested as a random effect but did not improve model fit, and were therefore excluded. To address non-linearity, gene expression values were log-transformed (log + 1), and model coefficients and confidence intervals were subsequently back-transformed for interpretation. Model assumptions, including normality of residuals, homoscedasticity and the absence of outlier, were assessed using diagnostic plots (e.g., Q-Q plots, Variance ratio test, etc.) and found to be reasonably satisfied. Models were fitted in R using the lmer [51] function from the *lme4* package.
To compare methylation levels across cluster types, we used non-parametric Kruskal-Wallis rank sum tests. When the global test was significant, post hoc pairwise comparisons were performed using Dunn's test with Bonferroni correction.

Additionnal supporting information can be found in the supplemental material and methods document.

ACKNOWLEDGMENTS
The authors would like to thank Patrick Ollitrault, Raphaël Morillon, François Luro, Anne Roulin, Flavia Thiebaut, Marie-Christine Carpentier, Moaine Elbaidouri, Gaëtan Droc, and Gauthier Sarah for fruitful discussions. This work received support from the sequencing and Bioinformatic platforms “Bio-environnement” at the University of Perpignan, from MESO@LR-Platform at the University of Montpellier and technical support from the bioinformatics group of the UMR AGAP Institute, member of the South Green Bioinformatics Platform. The authors are also grateful to Michel and Benedicte Bachès for sharing leaves to analyse the parental line *C. australasica.*

CONTRIBUTIONS
A. G. performed the statistical modelling presented in FIG 5C-F. Y. F. created the citrus hybrid. K. D. extracted the RNA and DNA from the parent and the hybrid with the help of N. P. and F.P. C.L. performed nanopore sequencing. F.P., N.P. and B. H. conceived and designed all experiments. K. D. performed all bioinformatic analyses with some help of M-C.C and N.P. and B.H. F.P. wrote the article with substantial contributions of K.D., B.H and A. G.. N.P. edited the article. N.P. and F.P. acquired main funding. All authors approved submission. All authors have read and agreed to the published version of the manuscript.

DATA AVAILABILITY STATEMENT

Nucleotide sequencing data generated in this study have been deposited in European Nucleotide Archive (ENA) under the accession number PRJEB94133 (https://www.ebi.ac.uk/ena/browser/search/PRJEB94133) and PRJEB94142 (https://www.ebi.ac.uk/ena/browser/search/PRJEB94142).

FUNDING

KD is supported by a PhD fellowship from Allocation Doctorale Occitanie (2022-25 Save-Citrus) and the Doctoral school ED 305 from UPVD. This work was mainly supported by the UPVD (BQR-2022). This study was also supported by the Laboratoires d'Excellence (LABEX) TULIP (ANR-10-LABX-41) and by the École Universitaire de Recherche (EUR) TULIP-GS (ANR-18-EURE-0019). BH work was supported by the French National Research Agency (ANR) under Grant ANR-23-CE20-0038.

CONFLICT OF INTERESTS

The authors declare no competing interest.

REFERENCES

1 Home | Food and Agriculture Organization of the United Nations. https://www.fao.org/home/en (accessed 22 Jul2025).

2 Wu GA *et al.* Genomics of the origin and evolution of Citrus. *Nature* 2018; **554**: 311–316.

3 Ollitrault P, Curk F, Krueger R. Citrus taxonomy. In: *The Genus Citrus*. Elsevier, 2020, pp 57–81.

4 Curk F, Luro F, Hussain S, Ollitrault P. Citrus Origins. In: *Citrus Production*. CRC Press, 2022.

5 Nakandala U *et al.* Phylogenetic relationships among Australian native citrus species based upon complete chloroplast genomes and single copy nuclear genes. *Trop Plants* 2023; **2**. doi:10.48130/TP-2023-0021.

6 Costa GV da, Neves CSVJ, Bassanezi RB, Leite Junior RP, Telles TS. Economic impact of Huanglongbing on orange production. *Rev Bras Frutic* 2021; **43**: e.

7 Alquézar B, Carmona L, Bennici S, Miranda MP, Bassanezi RB, Peña L. Cultural Management of Huanglongbing: Current Status and Ongoing Research. *Phytopathology* 2022; **112**: 11–25.

8 Yesiloglu T, Yilmaz B, Incesu M, Çimen B. Challenges to the Citrus Industry. 2022, pp 23–

33.

9 Dahro B, Li C, Liu J-H. Overlapping responses to multiple abiotic stresses in citrus: from mechanism understanding to genetic improvement. *Hortic Adv* 2023; **1**: 4.

10 Al-Khayri. *Advances in Plant Breeding Strategies: Fruits*. 2018 doi:10.1007/978-3-319-91944-7.

11 Cimen B, Yesiloglu T, Yilmaz B, Incesu M. Genetic Improvement in Citrus. In: *Citrus Production*. CRC Press, 2022.

12 Peng Z *et al.* A chromosome-scale reference genome of trifoliate orange (Poncirus trifoliata) provides insights into disease resistance, cold tolerance and genome evolution in Citrus. *Plant J* 2020; **104**: 1215–1232.

13 Alves MN *et al.* Insight into resistance to 'Candidatus Liberibacter asiaticus,' associated with Huanglongbing, in Oceanian citrus genotypes. *Front Plant Sci* 2022; **13**. doi:10.3389/fpls.2022.1009350.

14 Mahmoud LM, Dutt M. Novel citrus hybrids incorporating Australian lime genetics: development of HLB-tolerant citrus rootstocks and physiological changes in 'Valencia' sweet orange scions. *Front Plant Sci* 2025; **16**. doi:10.3389/fpls.2025.1614845.

15 Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010; **11**: 204–220.

16 Mason AS, Batley J. Creating new interspecific hybrid and polyploid crops. *Trends Biotechnol* 2015; **33**: 436–441.

17 Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA. Heterosis. *Plant Cell* 2010; **22**: 2105–2112.

18 Hochholdinger F, Yu P. Molecular concepts to explain heterosis in crops. *Trends Plant Sci* 2025; **30**: 95–104.

19 Greaves IK *et al.* Epigenetic Changes in Hybrids. *Plant Physiol* 2015; **168**: 1197–1205.

20 Koren S *et al.* De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* 2018; **36**: 1174–1182.

21 Ollitrault P *et al.* Comparative genetic mapping and a consensus interspecific genetic map reveal strong synteny and collinearity within the Citrus genus. *Front Plant Sci* 2024; **15**. doi:10.3389/fpls.2024.1475965.

22 Martin G *et al.* Improvement of the banana "Musa acuminata" reference sequence using

NGS data and semi-automated bioinformatics methods. *BMC Genomics* 2016; **17**: 243.

23 Droc G *et al.* A super-pangenome for cultivated citrus reveals evolutive features during the allopatric phase of their reticulate evolution. 2024; : 2024.10.17.618847.

24 Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinforma Oxf Engl* 2021; **37**: 1639–1643.

25 Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; **34**: 3094–3100.

26 Giraud D *et al.* Construction of a comprehensive library of repeated sequences for the annotation of Citrus genomes. *BMC Genomic Data* 2025; **26**: 30.

27 Kakoulidou I, Johannes F. DNA methylation remodeling in F1 hybrids. *Plant J Cell Mol Biol* 2024; **118**: 671–681.

28 Lauss K *et al.* Parental DNA Methylation States Are Associated with Heterosis in Epigenetic Hybrids. *Plant Physiol* 2018; **176**: 1627–1645.

29 Kakoulidou I *et al.* Mapping parental DMRs predictive of local and distal methylome remodeling in epigenetic F1 hybrids. *Life Sci Alliance* 2024; **7**. doi:10.26508/lsa.202402599.

30 Matzke MA, Kanno T, Matzke AJM. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu Rev Plant Biol* 2015; **66**: 243–267.

31 Zhan J, Meyers BC. Plant Small RNAs: Their Biogenesis, Regulatory Roles, and Functions. *Annu Rev Plant Biol* 2023; **74**: 21–51.

32 Xia Q-M *et al.* Localization and characterization of Citrus centromeres by combining half-tetrad analysis and CenH3-associated sequence profiling. *Plant Cell Rep* 2020; **39**: 1609–1622.

33 Stroud H *et al.* Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 2014; **21**: 64–72.

34 He L *et al.* Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proc Natl Acad Sci* 2021; **118**: e2107320118.

35 Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES. Trans Chromosomal Methylation in Arabidopsis hybrids. *Proc Natl Acad Sci* 2012; **109**: 3570–3575.

36 Rymen B, Ferrafiat L, Blevins T. Non-coding RNA polymerases that silence transposable elements and reprogram gene expression in plants. *Transcription* 2020; **11**: 172–191.

37 Okano Y, Miki D, Shimamoto K. Small interfering RNA (siRNA) targeting of endogenous promoters induces DNA methylation, but not necessarily gene silencing, in rice. *Plant J* 2008; **53**: 65–77.

38 Wang Y *et al.* CHH methylation of genes associated with fatty acid and jasmonate biosynthesis contributes to cold tolerance in autotetraploids of Poncirus trifoliata. *J Integr Plant Biol* 2022; **64**: 2327–2343.

39 Pontvianne F, Blevins T, Pikaard CS. Arabidopsis Histone Lysine Methyltransferases. In: Kader J-C, Delseny M (eds). *Advances in Botanical Research*. Academic Press, 2010, pp 1–22.

40 Harris CJ *et al.* A DNA methylation reader complex that enhances gene transcription. *Science* 2018; **362**: 1182–1186.

41 Picart-Picolo A *et al.* Large tandem duplications affect gene expression, 3D organization, and plant-pathogen response. *Genome Res* 2020; **30**: 1583–1592.

42 Emmerson R, Catoni M. The role of mobile DNA elements in the dynamic of plants genome plasticity. *J Exp Bot* 2025. doi:10.1093/jxb/erae523.

43 Zhang P *et al.* Extrachromosomal circular DNA and structural variants highlight genome instability in Arabidopsis epigenetic mutants. *Nat Commun* 2023; **14**: 5236.

44 Thieme M, Brêchet A, Bourgeois Y, Keller B, Bucher E, Roulin AC. Experimentally heat-induced transposition increases drought tolerance in Arabidopsis thaliana. *New Phytol* 2022; **236**: 182–194.

45 Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat Commun* 2019; **10**: 5818.

46 Li Y, Varala K, Moose SP, Hudson ME. The Inheritance Pattern of 24 nt siRNA Clusters in Arabidopsis Hybrids Is Influenced by Proximity to Transposable Elements. *PLOS ONE* 2012; **7**: e47043.

47 Graham JH, Bassanezi RB, Dawson WO, Dantzler R. Management of Huanglongbing of Citrus: Lessons from São Paulo and Florida. *Annu Rev Phytopathol* 2024; **62**: 243–262.

48 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation.

Genome Res 2017; **27**: 722–736.

49 Wick R. rrwick/Unicycler. 2025.https://github.com/rrwick/Unicycler (accessed 3 Jul2025).

50 Walker BJ. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement | PLOS One. 2014.https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0112963 (accessed 2 Jul2025).

51 Bates D *et al.* lme4: Linear Mixed-Effects Models using 'Eigen' and S4. 2025.https://cran.r-project.org/web/packages/lme4/index.html (accessed 3 Jul2025).