

Investigation of allele-specific expression in citrus hybrids reveals the association between siRNA-mediated *de novo* methylation and high expression in citrus genomes

Khadidiatou Diop¹ , Marie Bonnin^{1,2,3} , Anaïs Gibert¹, Christel Llauro¹, Yann Froelicher^{3,4}, Barbara Hufnagel^{3,5,*} , Nathalie Picault^{1,*}  and Frédéric Pontvianne^{1,6,*} 

¹CNRS/UPVD, Plant Genome and Development (LGDP), Perpignan 66100, France

²CNRS, Equipe d'Adaptation des Végétaux aux Changements Globaux, Projet Ressources Naturelles, UMR 6134 SPE, Univ de Corse, Corsica, France

³UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

⁴CIRAD, UMR AGAP Institut, F-20230 San Giuliano, France

⁵CIRAD, UMR AGAP Institut, F-97170 Petit-Bourg, Guadeloupe, France

⁶iGReD, Université Clermont Auvergne, CNRS, INSERM, 63001 Clermont-Ferrand, France

*Corresponding authors. E-mail: barbara.hufnagel@cirad.fr; nathalie.picault@univ-perp.fr; frederic.pontvianne@cnrs.fr

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 harbouring 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 *Citrus reticulata* (mandarin) and *Citrus australasica* (finger lime). Haplotype-resolved subgenomes assembly identified hundreds of allele-specifically expressed genes. Asymmetric reprogramming of methylation was observed, in particular an increase in CHH in *C. australasica* haplotype. Surprisingly, CHH methylation, usually associated with gene silencing, was correlated here with increased expression, but also 24 nt small RNA populations at their promoter regions. Similar analyses of the parental lines and other *Citrus* species 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 is associated with higher transcriptional activity. 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 ~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—*Citrus medica*, *Citrus maxima*, *Citrus reticulata*, *Citrus micrantha*, and *Citrus cavaleriei*—whose interspecific hybridizations, followed by clonal propagation *via* apomixis and grafting, gave rise to the multitude of cultivated varieties [2–4]. 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 *Citrus 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, USA, 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

Received: 22 July 2025. Accepted: 23 April 2026. Published: 4 May 2026. Corrected and Typeset: 1 October 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the Nanjing Agricultural University.

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 reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

efforts are increasingly turning towards 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 [12–14]. 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, ~4 million years ago [4], making it a particularly informative model to investigate the genomic consequences of wide hybridization, which integrates both authentic posthybridization regulatory changes and the persistence of conserved intragenus patterns.

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 vigour, 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 a hybrid-specific transcriptional program which is not simply an addition of each parental transcriptional program [18]. In *Arabidopsis thaliana*, rice, maize, and 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 hybrids 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, particularly diseases such as the aforementioned HLB. The aim of the present study is therefore not to identify resistance mechanisms *per se*, but to establish a haplotype-resolved framework to characterize allele-specific gene expression, DNA methylation, and small RNA accumulation in a complex interspecific citrus hybrid. 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, *C. reticulata* (mandarin) and *C. (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 not only of the genomes but also of the associated small RNAs that enable them to be deposited in certain contexts. In conclusion, this study shows that these pathways are associated with the establishment of differential allelic expression within the hybrid and that *de novo* methylation seems to play a particular role in citrus.

Results

Assembly of an interspecific citrus hybrid with a 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 [4]. 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 182× 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 as short-read sequencing. This strategy allowed the realization of accurate trio-binning and independent reconstruction of both haplotypes (Fig. 1A). Each assembled haplotypes consisted of only nine contigs, corresponding to the nine 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 and 39.9 Mb, respectively; L50 = 4 for both) and exhibited consistent GC content (~34%).

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 (Fig. 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 (Fig. 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 interchromosomal misplacements and the consistent anchoring of large

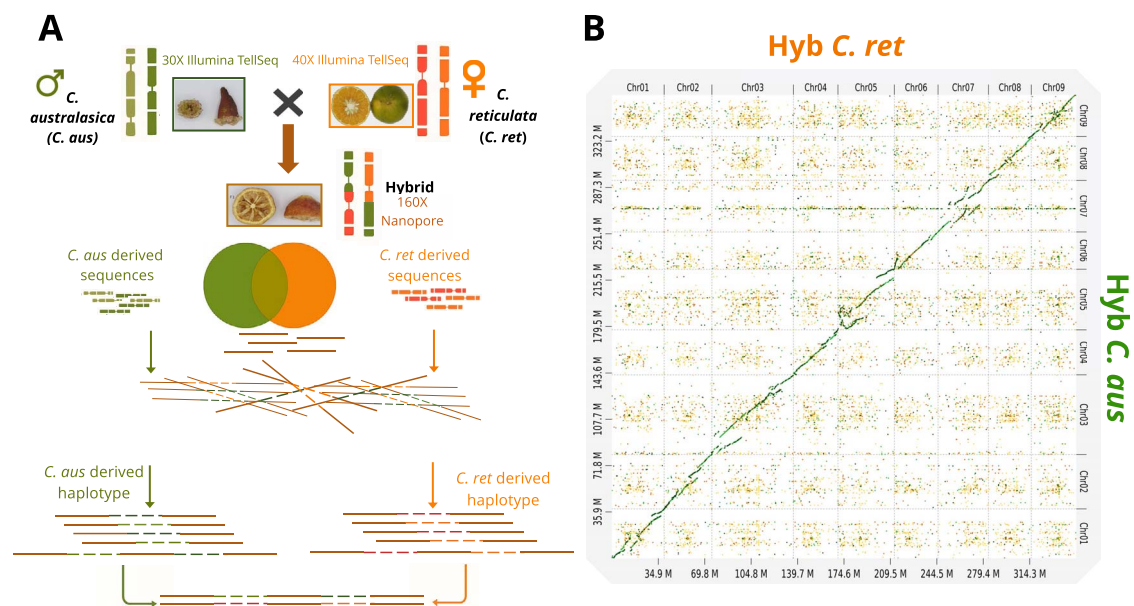


Figure 1 Haplotype-resolved genome assembly and structural variation. (A) Overview of the trio-binning genome assembly strategy used to generate haplotype-resolved assemblies of the citrus hybrid. Long reads from the hybrid were partitioned based on parental short-read data to generate separate assemblies corresponding to the maternal (*C. reticulata*, right panel) and paternal (*C. australasica*, left panel) 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.

Table 1 Summary statistics for the contigs of the hybrid diploid genome assembled with a haplotype resolution.

Statistic	<i>C. reticulata</i> haplotype	<i>C. australasica</i> 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
BUSCO completeness (%)	98.4	98.3
Unique genes (%)	78.0	73.8
Duplicated genes (%)	19.7	24.5

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 (e.g. Fig. 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 BUSCO present as unique or duplicated gene models (see Table 1).

Deciphering the differential allelic expression in the intergeneric 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 subgenomes, 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 showed differential bias in transcript accumulation, and counted 1758 genes in which the contribution of the allele derived from *C. australasica*

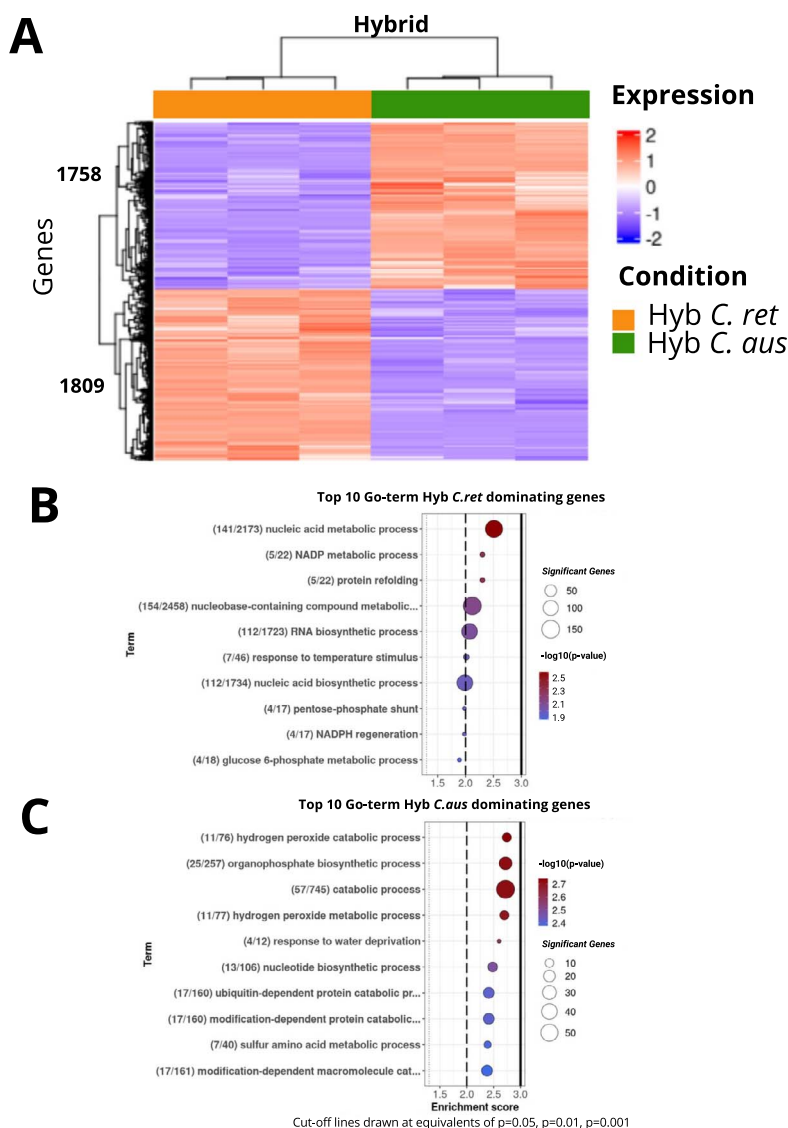


Figure 2 Allele-specific expression and functional enrichment in the hybrid. (A) Heatmap showing allele-specific expression between the two haplotypes in the hybrid. Expression values are variance-stabilized counts (VST) obtained from DESeq2 and centred gene-wise (z-score). Rows represent genes; columns represent biological replicates of the two hybrid haplotypes. The colour gradient reflects standardized expression levels from higher expression to 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).

was at least three times greater, and 1809 genes in which the contribution of the allele derived from *C. reticulata* was three times greater within the hybrid (Fig. 2A). Some transcript accumulation profiles were confirmed using RT-qPCR (Fig. S2 and Table S3).

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 peroxide, as well as in organophosphate biosynthetic processes (Fig. 2B and C and Fig. S3).

Such dominance at the GO term scale implies a biased 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 (Fig. S4). 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 (Fig. S4).

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 150 bp 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 (Fig. 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.

At the chromosomal scale, genome-wide methylation distribution showed mCG and mCHG highest in pericentromeric regions, correlating with repetitive element abundance (Fig. S5) [26]. Methylated CHH was generally lower but showed local peaks more pronounced in Hyb *C. aus* than Hyb *C. ret*, supporting enriched RdDM activity targeting the *C. australasica* haplotype.

We then validated the methylation pattern observed using WGBS in the hybrid by analysing five different promoter regions that either accumulate or do not accumulate cytosine methylation using BS-PCR followed by Sanger sequencing (Fig. S6 and Table S3). Next, we analysed DNA methylation patterns across gene bodies and their flanking regions using metaplot analyses. In CHG and CHH contexts, methylation dropped sharply within gene bodies, with elevated levels in flanking regions (Fig. 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 hybrids. 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 (Fig. 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 associated with epigenetic interactions and small RNA pathways [29].

We then examined the relationship between this methylation, whatever its context, and 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 region (1 kb upstream of the transcription start site, TSS) 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 (Fig. 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 (Fig. 3C). To assess the robustness of these results to threshold choice, we performed additional analyses using alternative methylation thresholds, which yielded consistent patterns (Fig. S7 and Table S4). 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 (Fig. S8). 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 intergenic 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 siRNA 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 (Fig. 4A) [31]. Overall, we can see that the two haplotypes participate equally in the production of these small RNAs in quantitative terms (Fig. 4B).

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 subgenomes, while the other two-thirds originate either from *C. reticulata*- or *C. australasica*-derived haplotype (Fig. S9). 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-nt 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 may show higher expression levels.

We therefore focused on the siRNA clusters targeting only the promoter regions (1 kb upstream of the TSS). 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 (Fig. 4B).

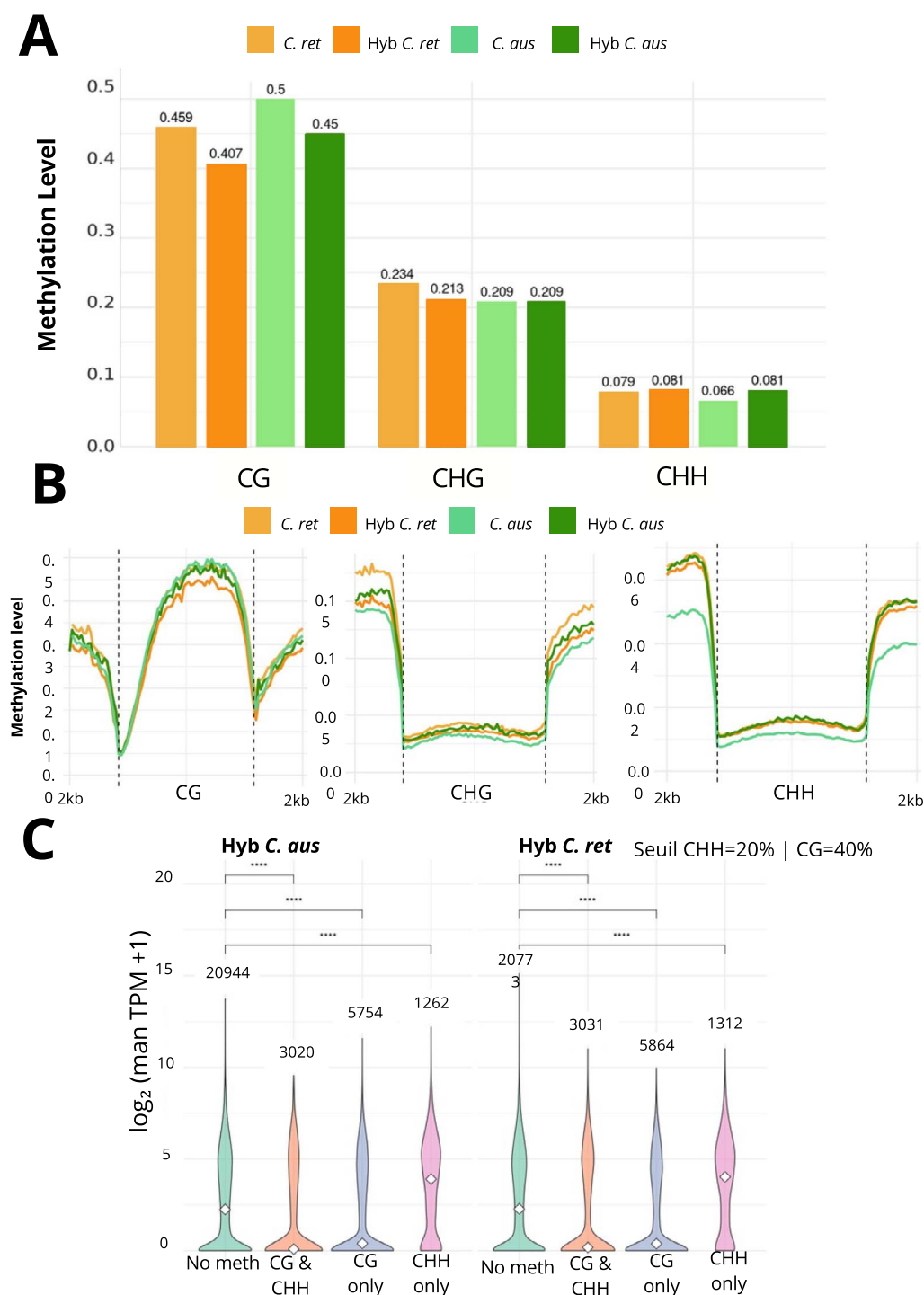


Figure 3 DNA methylation profiles and correlation with gene expression. (A) Global DNA methylation levels (in %) in CG, CHG, and CHH contexts across parental haplotypes (*C. reticulata*, *C. australasica*) and their corresponding haplotypes within the hybrid. Values indicate mean methylation levels across 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 ($\log_2(\text{mean TPM} + 1)$) 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 $\geq 20\%$, CG $\geq 40\%$). The number of genes in each category is indicated above each violin, and median expression values are shown within each distribution. Differences between categories were assessed using pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg correction. Effect sizes are reported as median shifts (Δ_{med}) relative to the 'No methylation' category, and are displayed above the comparisons.

Among these haplotype-specific siRNA clusters, only half are also expressed in the parental lines (Fig. 4C), highlighting a potential asymmetric inheritance. 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

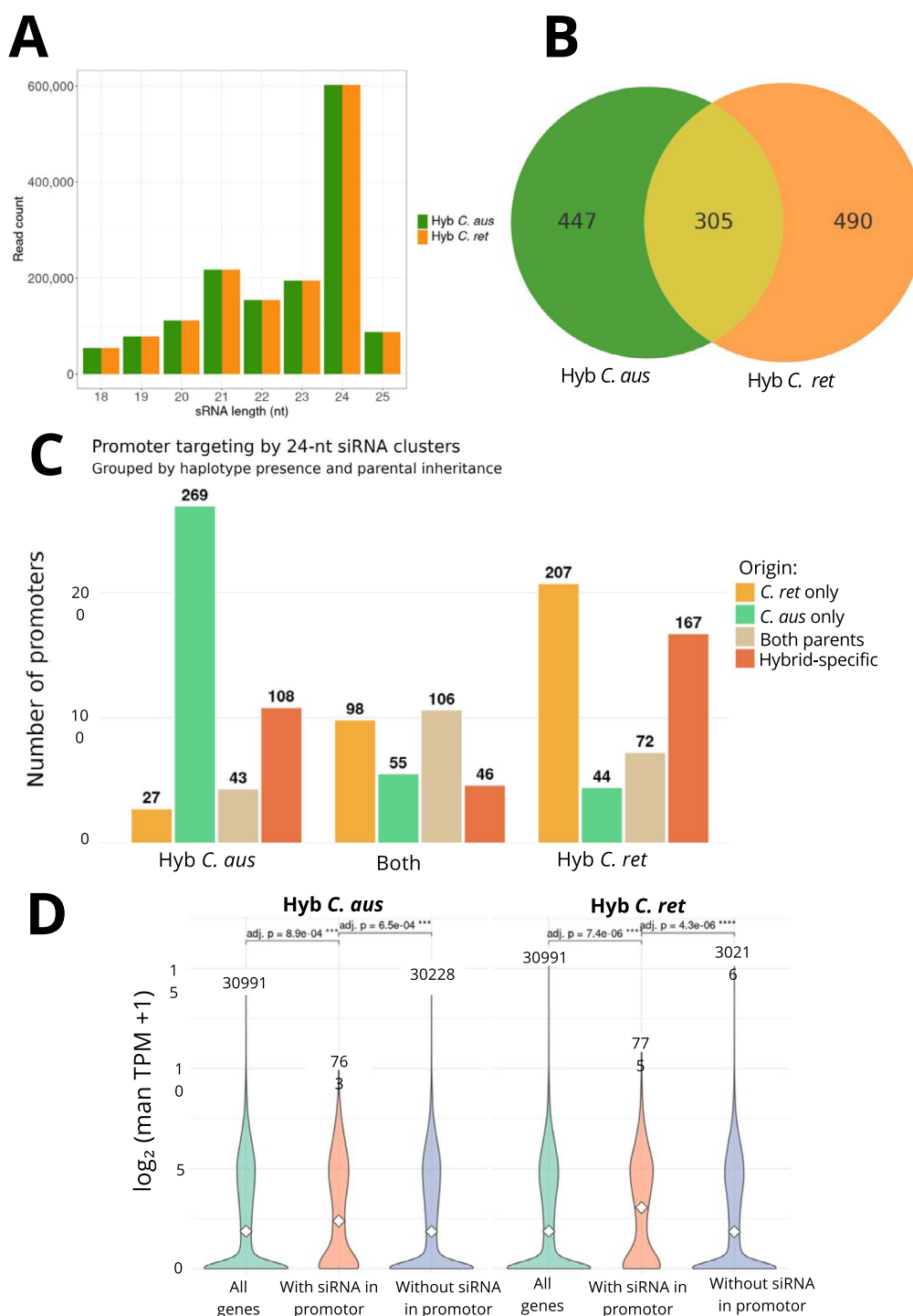


Figure 4 Haplotype-specific distribution and effects of 24-nt siRNA clusters. (A) Size distribution of mapped small RNAs in each hybrid haplotype. (B) Venn diagram showing the overlap of 24-nt siRNA clusters between the two haplotypes of the hybrid. The intersection represents clusters targeting the 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, *C. australasica* only, both parents, or are hybrid-specific. (D) Violin plots showing gene expression levels ($\log_2(\text{TPM} + 1)$) depending on the presence or absence of 24-nt siRNA clusters in promoters, for each hybrid haplotype. The number of genes in each category is indicated above each violin, and median expression values are shown within each distribution. Differences between groups were assessed using Wilcoxon rank-sum tests with Benjamini–Hochberg correction. Effect sizes are reported as median shifts (Δ_{med}) between groups and are displayed on the plots.

promoter (Fig. 4D). Altogether, these results demonstrate not only that the contribution of siRNA is essentially haplotype-dependent within the hybrid but also that the genes associated

with siRNA-targeted promoters tend to exhibit higher transcript levels consistent with a positive association between RdDM-related features and gene expression within the hybrid.

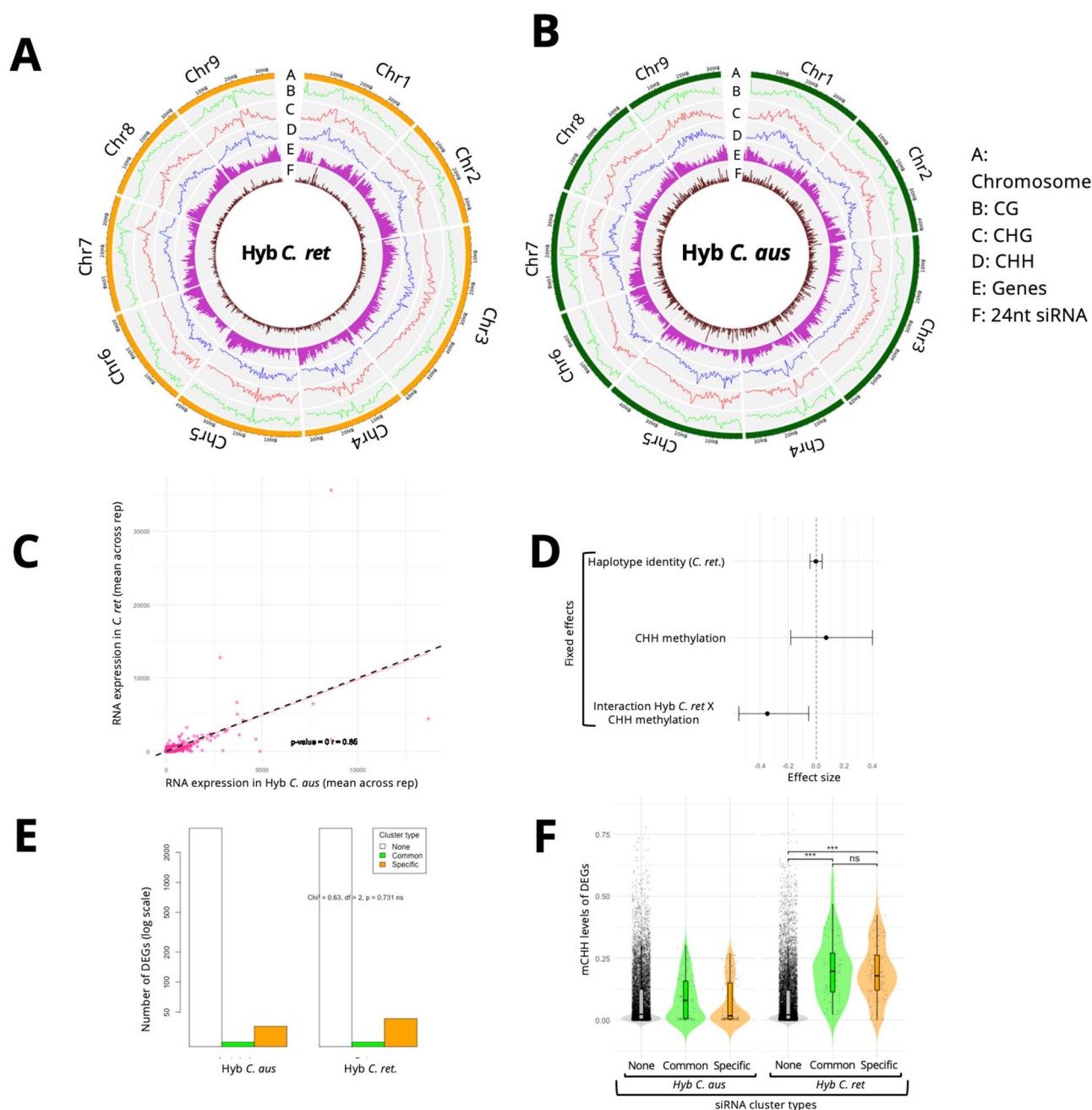


Figure 5 Genome-wide epigenomic patterns and overlap with expression. (A, B) Circos plots representing genome-wide profiles of DNA methylation, gene density, and 24-nt siRNA clusters across the nine chromosomes of the two haplotypes of the hybrid: (A) *C. reticulata*-derived haplotype (left panel) and (B) *C. australasica*-derived haplotype (right panel). Tracks from outer to inner rings: A: chromosomes; B–D: DNA methylation levels in CG (B), CHG (C), and CHH (D) contexts; E: gene density; F: density of 24-nt siRNA clusters. (C) Spearman's rank correlation in gene expression between haplotypes. (D) Effect sizes ($\pm 95\%$ CI) from a linear mixed model evaluating the association, of haplotype, mCHH levels, and their interaction on DEG expression (TPM). Contrasts are treatment-coded with *C. australasica* haplotype as the reference. CIs overlapping zero indicate nonsignificance. (E) DEG distribution 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: *** $P \leq 0.001$, ** $P \leq 0.01$, * $P \leq 0.05$; ns, not significant.

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 (Fig. 5A and B). We then performed several statistical analyses of gene expression within the hybrid, quantified as TPM.

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 (Fig. 5C), suggesting that most genes exhibit similar expression levels across the two haplotypes. We then tested whether the 3567 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 transcript identity was included as a random effect to reflect the grouping structure of the data. This approach captures underlying clustering or coexpression patterns between genes that could confound the interpretation of haplotype and methylation effects. The model revealed no overall differences in gene expression of DEG between haplotypes, confirming that neither haplotype systematically dominates the other in terms of expression of DEG (top; Fig. 5D). It also showed that mCHH alone had no consistent correlation with gene expression on DEG (middle; Fig. 5D). However, a significant interaction between haplotype identity and mCHH levels was detected (bottom; Fig. 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 associations of mCHH with gene expression of DEG is haplotype-dependent.

We then investigated the impact of hybridization on siRNA clusters on promoters of DEG. We categorized 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 DEG per category and haplotype, and used a Pearson's χ^2 test of independence to assess differences in distribution. The distribution of siRNA cluster types did not significantly differ between haplotypes ($\chi^2 = 0.63$, $df = 2$, $P = .73$). Most DEG lacked a detectable cluster in both haplotypes, and the proportions of 'Common' and 'Specific' clusters were similar across haplotypes (Fig. 5E). These results suggest that the presence or specificity of siRNA clusters among differentially expressed genes (DEGs) 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 DEG. Since methylation data did not follow a normal distribution, we used nonparametric 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 ($\chi^2(2) = 2.35$, $P = .31$). No pairwise comparisons were significant after correction. In contrast, *C. reticulata* haplotype showed a significant effect of siRNA cluster type on mCHH ($\chi^2(2) = 232.25$, $P < 2.2 \times 10^{-16}$). *Post hoc* tests revealed that DEG 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 ($P = .67$ 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 (Fig. 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 is 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 centres, 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 (TEs). The strong colocalization of mCHH and siRNA 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* $\times$ *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 (Fig. 5A and B), although a potential, we identified a haplotype-bias impact in the hybrid for DEG (Fig. 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 previously shown, analysis of expression data with either cytosine methylation or the presence of siRNA revealed that genes were more highly expressed when their promoters were associated with mCHH accumulation (Fig. 3) or were targeted by siRNA (Fig. 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 genes with detectable expression (Fig. 6A) from genes with no detectable expression (Fig. 6B). Remarkably, genes whose promoters accumulate these two characteristics are 10 times more likely to be associated detectable expression than among genes with no detectable expression. 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 detectable expression (Fig. 6A) and in 77% of cases (24 promoters out of 31) for genes with no detectable expression (Fig. 6B). We additionally applied a more stringent expression threshold (TPM > 1), which reduced the number of genes classified as expressed as expected, but did not alter the overall trend (Fig. S10).

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 (Fig. S11). 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. A closer examination of siRNA–CHH methylation modules shows that both CHH methylation and siRNA accumulate specifically within the 1 kb region upstream of the TSS, mostly at expressed genes (Fig. S12).

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* (Fig. 6C, Table S5) and in *C. australasica* (Fig. 6D and F, Table S5). *In fine*, in both *C. reticulata* and

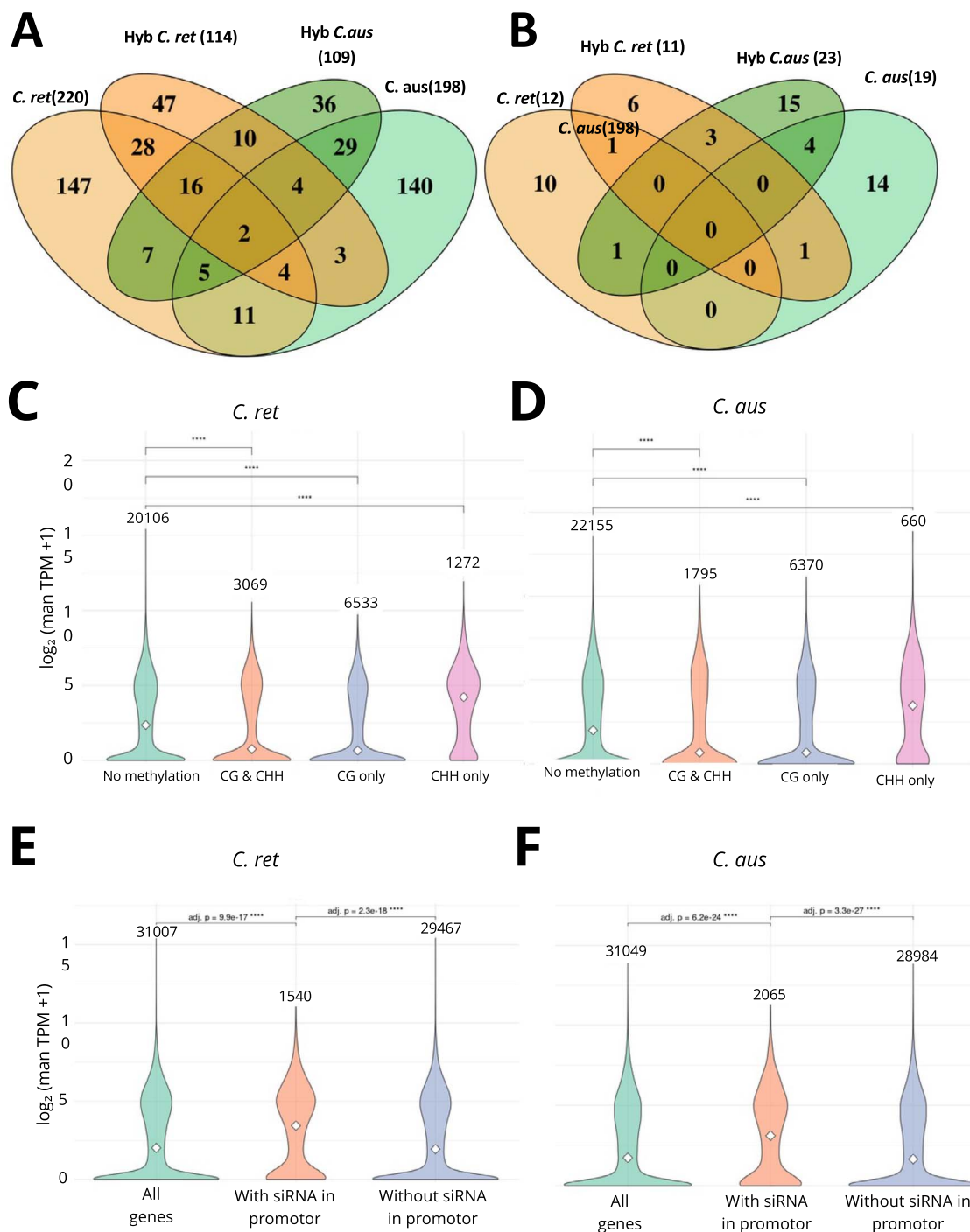


Figure 6 Epigenetic regulation of gene expression in the parental genomes. (A, B) Venn diagrams showing the overlap of genes whose promoters are targeted by 24-nt siRNA clusters across parental and hybrid haplotypes. (A) Genes with detectable expression (TPM > 0) with promoter siRNA clusters. (B) Genes with no detectable expression (TPM = 0) with promoter siRNA clusters. Each subset indicates the number of genes whose promoters are targeted by siRNA in one or more haplotypes. (C, D) Violin plots showing gene expression levels ($\log_2(\text{mean TPM} + 1)$) 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. The number of genes in each category is indicated above each violin, and median expression values are shown within each distribution. Differences between categories were assessed using pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg correction. Effect sizes are reported as median shifts (Δmed) relative to the 'No methylation' category and are displayed above the comparisons. (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*. Differences were assessed using Wilcoxon rank-sum tests with Benjamini-Hochberg correction. Effect sizes are reported as median shifts (Δmed) between groups and are displayed on the plots.

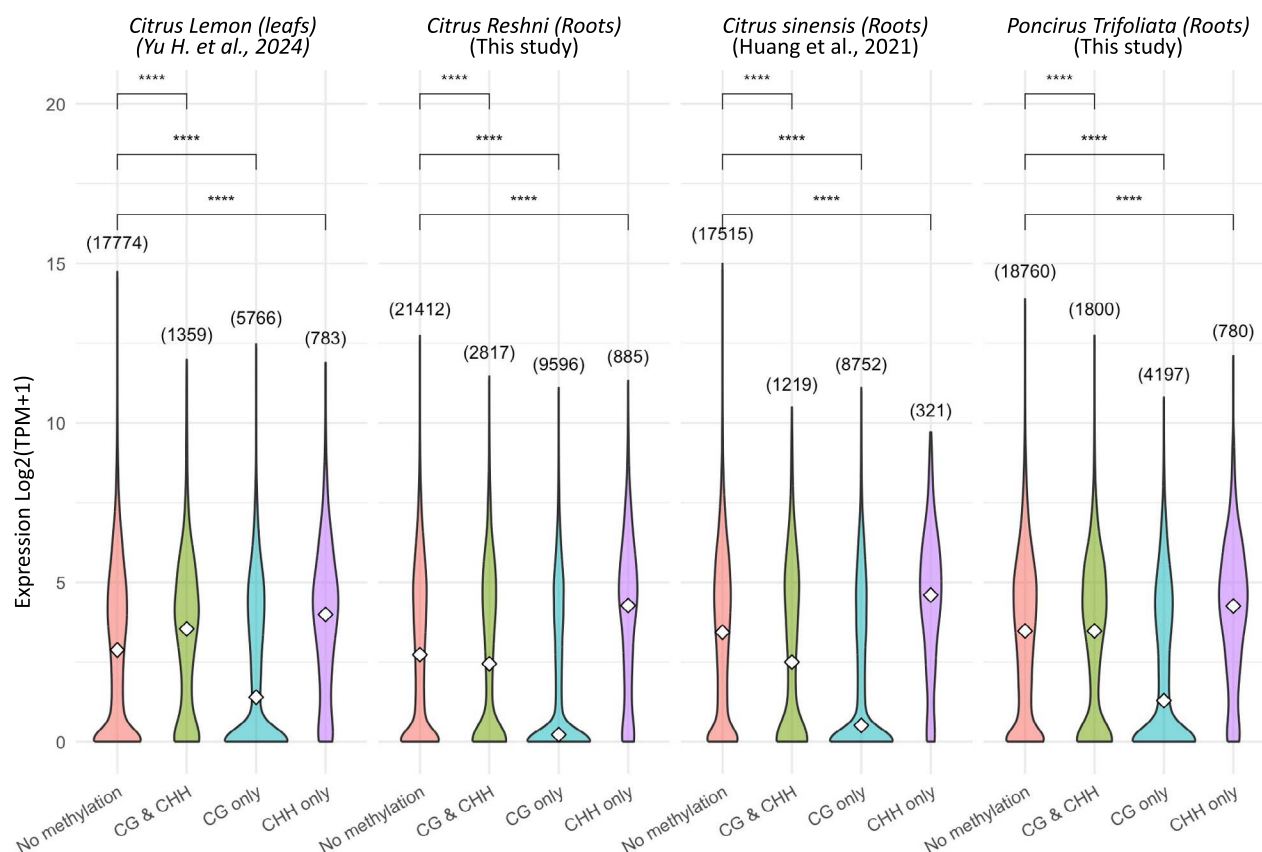


Figure 7 Correlation between DNA methylation and expression profiles in root and shoot tissues for *Citrus* and related species. Violin plots showing gene expression levels ($\log_2(\text{mean TPM} + 1)$) according to promoter methylation status in CG and CHH contexts in *Citrus lemon*, *Citrus reshni*, *Citrus sinensis*, and *Poncirus trifoliata*. Promoters are categorized into four classes: 'No methylation', 'CG & CHH', 'CG only', and 'CHH only' based on thresholds (CHH $\geq 20\%$, CG $\geq 40\%$). The number of genes in each category is indicated above each violin, and median expression values are shown within each distribution. Differences between categories were assessed using pairwise Wilcoxon rank-sum tests with Benjamini-Hochberg correction. Effect sizes are reported as median shifts (Δ_{med}) relative to the 'No methylation' category and are displayed above the comparisons.

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 (Fig. S13), suggesting that these two layers of RdDM regulation may act independently or additively.

To assess whether this profile is shared across citrus species, we took advantage of the published RNA-seq and WGBS-seq data in citrus lemon [35]. Again, a significant higher expression is also observed for the 783 genes that accumulate CHH methylation in their promoter (Fig. 7, left panel). Moreover, as only leaf tissue has been analysed so far, we also analysed the correlation between gene expression and accumulation of CHH methylation only in the promoter for other citrus and related species, this time using root tissues (Fig. 7). WGBS and RNA-seq analyses, from *C. reshni* (Cleopatra mandarin, a member of *C. reticulata*) and *Poncirus trifoliata*, were performed for this study. In addition, we also reanalysed advantage of the methylation and expression data previously published for *C. sinensis* [36]. In these three species, we were able to confirm a significant increase in gene expression when CHH-only methylation accumulates in gene promoters (1 kb upstream of the TSS) (Fig. 7, Table S6).

Altogether, 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. If the results suggest stable associations across datasets, they do not strictly exclude the influence of stage- or environment-specific factors.

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 (Fig. 1 and Fig. 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, e.g. 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 DEGs with allelic resolution in the hybrid. Because all analyses were conducted under nonstressed conditions, in the absence of both biotic and abiotic challenges, the regulatory patterns described here should be interpreted as a basal, haplotype-resolved framework rather than as direct evidence of stress- or disease-resistance mechanisms.

Hundreds of expressed genes accumulate transcripts mainly from only one of the parental-derived genomes, which implicate a certain dominance at the gene scale (Fig. 2, Figs S3 and S4). 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. These haplotype-specific patterns should be further evaluated in additional hybrid combinations to assess their generality across the *Citrus* genus.

At the methylation level, the global epigenomic landscapes were remarkably conserved between haplotypes, closer inspection of Fig. 5A and 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 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 is significantly associated with DEG derived from the *C. reticulata* haplotype (Fig. 5D and Fig. S15). This haplotype-specific mCHH methylation is accompanied by a specific accumulation of siRNA in the *C. reticulata* haplotype promoter of DEG (Fig. 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 (Fig. 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. In addition, we cannot entirely

rule out the potential impact of stage- or environment-specific factors. 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 [37], which is the reverse of our observation here. Further transcriptional and epigenetic analyses on future generations will show whether this imbalance persists.

We observe that the RdDM is associated with transcriptional activation rather than repression in *Citrus*. However, establishing causality between these molecular layers will require targeted functional validation. One hypothesis could be that the RdDM pathway is affected in citrus, with transcriptional regulation affecting DNA methylation deposition or maintenance. However, the expression profile of the major members of the pathway appears stable in the hybrid and parents (Fig. S14). Indeed, we identified hundreds of highly expressed genes in the hybrid and in the parental lines that accumulate both mCHH to a level higher than 20% and that are also targeted by siRNA at their promoters. This number of targets is probably higher since it cannot be ruled out that the lack of detection of siRNA 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 [15, 38]. 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 [39]. 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 [40]. This pattern is consistent with observations in other species, although differences in tissue context and experimental conditions require cautious interpretation. 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 [41]. 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 [42]. These homologs also exist in citrus and are expressed in the parental and the hybrid plants. Creating SUVH1 and SUVH3 methylation reader knock-out using CRISPR-cas9 should reveal, in the future, if the association is mediated by these specific histone methyltransferases. Furthermore, we could investigate the impact of CHH methylation on highly expressed genes by generating knockouts of essential RdDM core components, such as the *de novo* methyltransferase DRM2 or the catalytic subunit of RNA polymerase V. One additional possibility is that open chromatin states and high transcription may favour *de novo* CHH methylation, as previously observed in chromatin mutants in *A. thaliana* [43]. This hypothesis could be tested by changing the chromatin status of different identified promoter using methylase or demethylase and see how it affects siRNA accumulation and gene expression [44]. However, while CRISPR-Cas9 technology is feasible and established protocols exist in citrus [45], it remains technically challenging and time-intensive. As an

alternative approach, we propose to investigate the relationship between CHH methylation patterns and gene expression in citrus under stress conditions or across different developmental stages. This will allow us to assess how these genes respond when their methylation or expression profiles are experimentally altered.

Conclusions

Our study reveals the impact of DNA methylation and siRNA on the establishment of DEG in a complex interspecific 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 contribute to the generated genomic diversity and gene copy number changes that can lead to phenotypic plasticity, including in crops [46]. However, these structural interpretations remain preliminary, as their accurate detection requires a near perfect assembly, usually assisted with chromatin–chromatin interaction maps (Hi-C) as well as detailed TE annotations, approaches which is not our case in the present study. Mobile genetic elements, such as TEs, 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 [47–50]. Their presence in promoter regions can therefore impact siRNA production and inheritance in hybrid context [51]. 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 coevolutionary 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 HLB [52], and also to produce fruit that meets the standards for marketing.

Materials and methods

Plant material

Hybrid F1 result from a cross between *C. reticulata* Blanco cv. Fortune mandarin (female parent) and *C. australasica* cv. Caviar Bachès (male parent), located at the INRAE CIRAD station in San Giuliano, France. The tissue analysed for *P. trifoliata* and *C. reshni* has been previously described in Bonnin *et al.* [53]. The plant material used to analyse *Citrus lemon* and *C. sinensis* has been previously described in Yu *et al.* [35] and Huang *et al.* [36], respectively.

Long-read sequencing

Genomic DNA from the hybrid was sequenced using the Oxford Nanopore Technologies (ONT) PromethION platform at the LGDP sequencing facility in Perpignan, generating a total of 63.88 Gb of long-read data. This corresponds to $\sim 182\times$ coverage of the ~ 350 Mb genome.

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 [54] with options *genomeSize = 430 m*. Initial polishing of the assemblies was performed using Medaka [55], followed by a second round with Pilon [56] using the mapped Tell-Seq reads and the options *–fix snps,indels –minmq 3*. Statistical analyses of the assembly are available in the (Table S2).

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.

Acknowledgements

The authors would like to thank Raphaël Morillon, Patrick Ollitrault, 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, the 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 would like to thank the ‘Occitanie Region’ for funding K.D. and the ‘Collectivité de Corse’ for funding M.B. The authors are also grateful to Michel and Benedicte Bachès for sharing leaves to analyse the parental line *C. australasica*. K.D. is supported by a PhD fellowship from Allocation Doctorale Occitanie (2022–25 Save-Citrus) and the Doctoral school ED 305 from UPVD. M.B. was funded by the ‘Collectivité de Corse’ and CIRAD Montpellier UMR AGAP Team Seapag and Team MEAC at LGDP. 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). B.H. work was supported by the French National Research Agency (ANR) under grant ANR-23-CE20-0038.

Author 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. M.B. realised RNA-seq and WGBS analyses for *C. reshni* and *P. trifoliata*. 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 from M.-C.C., 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. F.P., N.P., and B.H. coordinated the research. All authors approved submission. All authors have read and agreed to the published version of the manuscript.

Data availability

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>). Additional supporting information can be found in the Supplemental Material and methods document.

Conflicts of interest statement

The authors declare no competing interest.

Supplementary material

Supplementary material is available at *Horticulture Research* online.

References

- Mukhametzyanov RR, Brusenko SV, Khezhev AM. *et al.* Changing the global production and trade of citrus fruits. In: Popkova EG, Bogoviz AV, Sergi BS, Kaurova OV, Maloletko AN, eds. *Sustainable Development of the Agrarian Economy Based on Digital Technologies and Smart Innovations*. Springer Nature Switzerland: Cham, 2024;19–24
- Curk F, Luro F, Hussain S. *et al.* Citrus origins. In: *Citrus Productions*. Boca Raton: CRC Taylor and Francis Group, Press, 2022,
- Ollitrault P, Curk F, Krueger R. Citrus taxonomy. In: *The Genus Citrus*. Elsevier, 2020,
- Wu GA, Terol J, Ibanez V. *et al.* Genomics of the origin and evolution of citrus. *Nature*. 2018;554:311–6
- Nakandala U, Furtado A, Smith MW. *et al.* Phylogenetic relationships among Australian native citrus species based upon complete chloroplast genomes and single copy nuclear genes. *Trop Plants*. 2023;2:21
- Alquézar B, Carmona L, Bennici S. *et al.* Cultural management of Huanglongbing: current status and ongoing research. *Phytopathology*. 2022;112:11–25
- Costa GV, Neves CSVJ, Bassanezi RB. *et al.* Economic impact of Huanglongbing on orange production. *Rev Bras Frutic*. 2021;43:
- 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
- Yesiloglu T, Yilmaz B, Incesu M. *et al.* Challenges to the citrus industry. In: *Citrus Productions*. Boca Raton: CRC Taylor and Francis Group, Press, 2022,
- Al-Khayri. *Advances in Plant Breeding Strategies: Fruits*. Cham: Springer International Publishing; 2018:
- Cimen B, Yesiloglu T, Yilmaz B. *et al.* Genetic improvement in citrus. In: *Citrus Production*. Boca Raton: CRC Taylor and Francis Group, Press, 2022,
- Alves MN, Raiol-Junior LL, Girardi EA. *et al.* Insight into resistance to ‘*Candidatus Liberibacter asiaticus*,’ associated with Huanglongbing, in Oceanian citrus genotypes. *Front Plant Sci*. 2022;13:1009350
- Mahmoud LM, et 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;16:1614845
- Peng Z, Bredeson JV, Wu GA. *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;104:1215–32
- Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010;11:204–20
- Mason AS, Batley J. Creating new interspecific hybrid and polyploid crops. *Trends Biotechnol*. 2015;33:436–41
- Birchler JA, Yao H, Chudalayandi S. *et al.* Heterosis. *Plant Cell*. 2010;22:2105–12
- Hochholding F, Peng Y. Molecular concepts to explain heterosis in crops. *Trends Plant Sci*. 2025;30:95–104
- Greaves IK, Gonzalez-Bayon R, Wang L. *et al.* Epigenetic changes in hybrids. *Plant Physiol*. 2015;168:1197–205
- Koren S, Rhie A, Walenz BP. *et al.* De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol*. 2018;36:1174–82
- Ollitrault P, Hufnagel B, Curk F. *et al.* Comparative genetic mapping and a consensus interspecific genetic map reveal strong synteny and collinearity within the citrus genus. *Front Plant Sci*. 2024;15:1475965
- Martin G, Baurens F-C, Droc G. *et al.* Improvement of the banana “*Musa acuminata*” reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics*. 2016;17:243
- Droc G, Giraud D, Belser C. *et al.* A super-pangenome for cultivated citrus reveals evolutive features during the allopatric phase of their reticulate evolution. *Plant Biotechnol J*. 24:3345–67
- Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinf (Oxf)*. 2021;37:1639–43
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100
- Giraud D, Choise N, Summo M. *et al.* Construction of a comprehensive library of repeated sequences for the annotation of citrus genomes. *BMC Genomic Data*. 2025;26:30
- Kakoulidou I, et Johannes F. DNA methylation remodeling in F1 hybrids. *Plant J*. 2024;118:671–81
- Lauss K, Wardenaar R, Oka R. *et al.* Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiol*. 2018;176:1627–45
- Kakoulidou I, Piecyk RS, Meyer RC. *et al.* Mapping parental DMRs predictive of local and distal methylome remodeling in epigenetic F1 hybrids. *Life Sci Alliance*. 2024;7:e202402599
- 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;66:243–67
- Zhan J, Meyers BC. Plant small RNAs: their biogenesis, regulatory roles, and functions. *Annu Rev Plant Biol*. 2023;74:21–51
- Xia Q-M, Miao L-K, Xie K-D. *et al.* Localization and characterization of citrus centromeres by combining half-tetrad analysis and CenH3-associated sequence profiling. *Plant Cell Rep*. 2020;39:1609–22

33. He L, Zhao C, Zhang Q. *et al.* Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proc Natl Acad Sci.* 2021;118:e2107320118
34. Stroud H, Do T, Jiamu D. *et al.* Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol.* 2014;21:64–72
35. Yu H, Zhang C, Lu C. *et al.* The lemon genome and DNA methylome unveil epigenetic regulation of citric acid biosynthesis during fruit development. *Hortic Res.* 2024;11:uhae005
36. Huang Y, Xu Y, Jiang X. *et al.* Genome of a citrus rootstock and global DNA demethylation caused by heterografting. *Hortic Res.* 2021;8:69
37. Greaves IK, Groszmann M, Ying H. *et al.* Trans chromosomal methylation in Arabidopsis hybrids. *Proc Natl Acad Sci.* 2012;109:3570–5
38. Rymen B, Ferrafiat L, Blevins T. Non-coding RNA polymerases that silence transposable elements and reprogram gene expression in plants. *Transcription.* 2020;11:172–91
39. 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
40. Wang Y, Zuo L, Wei T. *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;64:2327–43
41. Pontvianne F, Blevins T, Pikaard CS. Arabidopsis histone lysine methyltransferases. *Adv Bot Res.* 2010;53:
42. Harris CJ, Scheibe M, Wongpalee SP. *et al.* A DNA methylation reader complex that enhances gene transcription. *Science.* 2018;362:1182–6
43. Earley KW, Frederic P, Andrzej W. *et al.* Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev.* 2010;24:1119–32
44. Gallego-Bartolomé J. DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytol.* 2020;227:38–44
45. Wang M, Qin Y-Y, Wei N-N. *et al.* Highly efficient *Agrobacterium rhizogenes*-mediated hairy root transformation in citrus seeds and its application in gene functional analysis. *Front Plant Sci.* 2023;14:1293374
46. Picart-Piccolo A, Grob S, Picault N. *et al.* Large tandem duplications affect gene expression, 3D organization, and plant-pathogen response. *Genome Res.* 2020;30:1583–92
47. Emmerson R, et Catoni M. The role of mobile DNA elements in the dynamic of plants genome plasticity. *J Exp Bot.* 2025;76:erae523
48. Thieme M, Brêchet A, Bourgeois Y. *et al.* Experimentally heat-induced transposition increases drought tolerance in *Arabidopsis thaliana*. *New Phytol.* 2022;236:182–94
49. Zhang P, Mbodj A, Soundiramourty A. *et al.* Extrachromosomal circular DNA and structural variants highlight genome instability in Arabidopsis epigenetic mutants. *Nat Commun.* 2023;14:5236
50. Baduel P, Quadra L, Hunter B. *et al.* Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat Commun.* 2019;10:5818
51. Li Y, Varala K, Moose SP. *et al.* The inheritance pattern of 24 nt siRNA clusters in Arabidopsis hybrids is influenced by proximity to transposable elements. *PLoS One.* 2012;7:e47043
52. Graham JH, Bassanezi RB, Dawson WO. *et al.* Management of Huanglongbing of citrus: lessons from São Paulo and Florida. *Annu Rev Phytopathol.* 2024;62:243–62
53. Bonnin M, Soriano A, Favreau B. *et al.* Comparative transcriptomic analyses of diploid and tetraploid citrus reveal how ploidy level influences salt stress tolerance. *Front Plant Sci.* 2024;15:1469115
54. Koren S, Walenz BP, Berlin K. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27:722–36
55. Wick R. (2016) 2025. *rrwick/Unicycler*. C++. mars 14, released juin 14. <https://github.com/rrwick/Unicycler>.
56. Walker BJ, Abeel T, Shea T. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014;9:e112963