Phased T2T genome of a tetraploid grapevine reveals segmental allopolyploid origin and allele-specific alternative splicing during fruit development

Yingchun Zhang[1, 4], Xu Wang[1], Yanshuai Xu[1,5], Cheng Chen[5], Jin Yao[6], Qimin Long[1], Tianhao Zhang[3], Ying Su[1], Sifan Yang[1], Zain Ali[3], Ruo Yang[3], Sheng Yan[3], Yang Dong[3], Lianzhu Zhou[3], Zhongxing Jin[3], Xiucai Fan[2], Qingqing Fu[3], Ling Tian[6], Chonghuai Liu[2], Yongfeng Zhou[1, 3, *], Wei Li[1, *], Yanling Peng[1, *] & Lei Sun[2, *]

1. State Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China
2. National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China
3. State Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China
4. College of Life Sciences, South China Agricultural University, Guangzhou, China
5. College of Horticulture, Hunan Agricultural University, Changsha, China
6. School of Management, Shenzhen Polytechnic University, Shenzhen, China

*Corresponding authors: Lei Sun, sunlei01@caas.cn; Yanling Peng, pengyanling@caas.cn; Yongfeng Zhou, zhouyongfeng@caas.cn, and Wei Li, liwei11@caas.cn
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors is (Lei Sun).

Running head: Multi-omics of phased T2T tetraploid grapevine

© The Author(s) 2026. Published by Oxford University Press on behalf of American Society of Plant Biologists. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Abstract

The tetraploid table grape Kyoho, a cultivar of major global economic importance, has long presented a challenge in understanding its genetic architecture and the regulatory landscapes between its different ancestral genomes. Here, we assemble a haplotype-resolved, near-complete reference genome (480.32-489.47 Mb, contig N50 17.5-23.3 Mb, with a total of only 11 gaps) to investigate regulatory dynamics through integrative multiomic analyses during fruit development. Comparative evolutionary genomics classifies Kyoho as a segmental allotetraploid with a complex mosaic genome, composed of ~71% *V. vinifera* (Vv) and ~29% *V. labrusca* (Vl) ancestry. A total of 67.3% of genes are shared among the four haplotypic genomes with 9.8% haplotype specific genes. Transcriptomic analysis revealed that among 17,750 pairs of ancestral alleles exhibiting directional bias, only 4,086 (23.4%) maintained a consistent ancestral dominance direction across all tissues. Similarly, comparisons of DNA methylation profiles showed no significant differences between genome ancestries Vl (0.009~0.676): Vv (0.010~0.670). However, alternative splicing (AS) analyses demonstrated that allele-specific splicing changes were correlated with ancestral origin across tissues (Vl: 1,529 vs. Vv: 2,052). Integrative omics identified 2,307 differentially expressed ancestral alleles that showed no significant genomic variants or methylation bias but exhibited AS bias. These genes include key regulators of fruit ripening, including *ZEP*, *PIN*, *DOG1* and *MYB*. Overall, this work delivers a landmark genomic resource and characterizes its regulatory architecture, facilitating functional genomic studies and guiding precision polyploid breeding in grapes.

Keywords: Polyploid, Structural variation, Allelic genes, Multi-omics, Alternative splicing, Grape breeding

Introduction

Polyploidy is an important force in plant evolution and crop breeding (Cheng *et al.*, 2025). Polyploids are typically classified into two distinct types: autopolyploid and allopolyploid. Autopolyploids arise from whole-genome duplication within a single species, exhibiting multivalent chromosome paring and tetrasomic inheritance (Xue *et al.*, 2022, Sun *et al.*, 2022).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Conversely, allopolyploids originate from hybridization between divergent species. They exhibit strictly bivalent pairing and tetrasomic inheritance (Zhuang *et al.*, 2019, Yoo *et al.*, 2013, Pfeifer *et al.*, 2014). However, rather than a strict dichotomy, polyploidy actually exists along a continuous spectrum. This spectrum is often bridged by intermediate forms such as segmental allopolyploids, which arise from complex hybridization and display mixed meiotic behaviors (Bertioli *et al.*, 2019, Wu *et al.*, 2025, Song *et al.*, 2023). Polyploid crops often exhibit enhanced vigor and stress tolerance but pose significant challenges in biological studies and breeding applications (Wang *et al.*, 2024a, Zhuang et al., 2019, Yoo et al., 2013, Pfeifer et al., 2014, Xue et al., 2022). The coexistence of highly similar homologous sequences of different ancestral origins and frequent structural variations complicates the analysis of regulatory networks, obscuring the link between genotype and phenotype in polyploids (Wang et al., 2024a, Zhuang et al., 2019, Yoo et al., 2013, Pfeifer et al., 2014, Xue et al., 2022). Recent breakthroughs in phased genomics have underscored that key agronomic traits are often governed by specific haplotype interaction (Brinton *et al.*, 2020, Todesco *et al.*, 2020, Walkowiak *et al.*, 2020). Thus, dissecting polyploid genomes at the haplotype level is paramount for deciphering the genetic basis of complex traits.

The advent of phased genomics is revealing that the boundary between auto- and allopolyploidy is far more blurred than previously thought (Zhang *et al.*, 2025, Song *et al.*, 2024b, Sun et al., 2022). Several economically important crops are not simple combinations of distinct subgenomes but possess chimeric genomic architectures (Zhang et al., 2025, Bertioli et al., 2019, Wu et al., 2025, Song et al., 2023). In such segmental allopolyploids, a single homologous chromosome group may act as a mosaic, containing haplotypes of mixed ancestry alongside those from a single progenitor (Zhang et al., 2025, Bertioli et al., 2019, Wu et al., 2025, Song et al., 2023). This structural mosaicism creates a unique regulatory landscape where divergent ancestral regulatory codes must coexist (Xue et al., 2022, Sun et al., 2022, Zhang *et al.*, 2022, Zhang *et al.*, 2019). How do epigenetic mechanisms, such as DNA methylation, and post-transcriptional processes, like alternative splicing, coordinate across these distinct ancestral haplotypes? The hybridization triggers epigenetic reprogramming (Xue et al., 2022, Sun et al., 2022) and that splicing expands transcriptomic diversity (Zhang et al., 2022, Zhang et al., 2019), the rules governing how these regulatory layers are partitioned and synchronized within a single, structurally chimeric genome remain largely unexplored.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Grapevine (*Vitis vinifera* L.) is a crop of immense economic and cultural importance, yet genomic research has predominantly focused on diploid samples (Shi *et al.*, 2023, Wang *et al.*, 2024b, Peng *et al.*, 2025, Liu *et al.*, 2024, Xiao *et al.*, 2025), leaving superior polyploid cultivars under-explored. The Kyoho grape, one of the world's most cultivated variety (OIV, 2017), epitomizes the complex polyploidy described above. As a tetraploid derived from an interspecific cross between *V. vinifera* and *V. labrusca*—species lacking strong reproductive barriers. Strikingly, our genomic analysis reveals that this hybrid origin has shaped Kyoho into a classic segmental allopolyploid. Rather than exhibiting uniform genetic behavior, it displays a complex mixture of both tetrasomic and disomic inheritance. This mixed meiotic behavior drives extensive but localized recombination, creating a deeply blurred genomic boundary and a highly mosaic chromosomal architecture. Consequently, it serves as an ideal system for dissecting the interplay between mosaic ancestry and regulatory evolution. Although its hybrid origin is known, the precise architecture of its genome-specifically, how its constituent haplotypes trace back to their ancestors and how these distinct ancestries navigate the regulatory landscape-remains an open question.

Here, we present a haplotype-resolved, near-complete genome assembly of Kyoho to dissect its genomic origin and regulatory landscape of a complex hybrid tetraploid. By integrating genomics, epigenomics, and transcriptomics (RNA-seq and Iso-Seq), we address three fundamental questions. (1) What is the precise genomic architecture of this complex tetraploid, and does it exhibit the mosaic ancestry characteristic of segmental alloploids? (2) How are distinct regulatory strategies, including gene expression, DNA methylation and alternative splicing, partitioned between haplotypes of *V. vinifera* versus *V. labrusca* ancestry? (3) Is there a mechanistic link between allele-specific DNA methylation and the regulation of alternative splicing across these ancestral alleles, driving their functional specialization? This work provides a new paradigm for understanding the interplay between mosaic genomic ancestry and regulatory evolution in polyploid crops.

Results

A haplotype-resolved near-complete genome assembly for tetraploid Kyoho table grape

To achieve a high-quality, phased genome assembly for the tetraploid table grape Kyoho, we generated 163.9 Gb of PacBio HiFi (~72×), 42.49 Gb of Oxford Nanopore (ONT) (~34×), and 200 Gb of Hi-C (~110×) data. K-mer analysis (21-mers) estimated a haploid genome size of

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

approximately 435.36 Mb with a high heterozygosity rate of 3.6% (Fig. 1C, Supplementary Fig. 1B). Fluorescence in situ hybridization (FISH) analysis confirmed the tetraploid state with 76 chromosomes (Fig. 1E). By integrating these datasets, we constructed a haplotype-resolved, near-complete reference genome spanning 1.81 Gb, anchored into 76 pseudochromosomes (4 sets of 19). The assembly achieved exceptional continuity, with only 11 gaps remaining across the entire tetraploid genome. The four phased assemblies (designated hap1-hap4) exhibited comparable sizes (480.32-489.47 Mb) and high contiguity, with contig N50 values ranging from 17.5 Mb to 23.3 Mb (Table S1). Importantly, the labels "hap1" through "hap4" are utilized strictly as a technical convention to distinguish the four phased copies of each chromosome within our assembly; they are not intended to imply the existence of four biologically distinct, genome-wide co-segregating haplotypes.

We rigorously validated the quality of genome assembly, confirming its near-telomere-to-telomere (T2T) status. First, the size of each haplotype closely matched the *V. vinifera* reference (PNT2T, ~494.87 Mb), and sequencing reads from all platforms (ONT, HiFi, Illumina) mapped back with rates approaching 100% (Supplementary Fig. S1A). Second, the Hi-C contact maps displayed clear diagonal lines with negligible off-diagonal noise, confirming the precise ordering and orientation of contigs within chromosomes (Fig. 1F). Third, the high completeness of the assembly was demonstrated by Benchmarking Universal Single-Copy Orthologs (BUSCO) scores averaging 97.95% for all haplotypes (range: 97.60%-98.30%) (Fig. 1D, Supplementary Fig. S1C) and high LTR Assembly Index (LAI) scores (range: 16.42-19.55). Most notably, the structural integrity of the assembly allowed for the precise delineation of centromeres for all 76 chromosomes and the end-to-end reconstruction of 122 out of 152 expected telomeres (Supplementary Table S1), providing a robust foundation for downstream analyses.

With a high-quality backbone established, we characterized the genomic landscape of each haplotype. Repeat sequences constituted approximately two-thirds of the genome (range: 66.32%-67.14%), with Long Terminal Repeats (LTRs) being the dominant class (range: 31.70%-32.49%) (Supplementary Table S2). We predicted an average of 29,742 high-confidence protein-coding genes per haplotype (range: 29,599-29,960) (Supplementary Table S1 and Table S3). This comprehensive annotation provides the necessary resolution to dissect the interplay between genomic elements and regulatory layers discussed below.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

To unravel the genome-wide diversity of Kyoho, we aligned the four haplotypes to the PNT2T reference, identifying 14,079,340 SNPs, 2,917,719 Indels, and 373,745 structural variants (SVs >50 bp), with SNPs and Indels affecting ~4% of the genome. Private variation loads were unevenly distributed: we detected 2,278,773, 2,430,253, 1,071,101, and 997,718 haplotype-specific SNPs in hap1-hap4, respectively, along with 547559, 557,280, 261,110, and 244,825 specific Indels; and 70,395, 72,548, 40,854, and 40,378 specific SVs (Supplementary Fig. S2, Fig. S3A, Table S4 and Table S5). PCA confirmed significant divergence from the reference, while chromosomal clustering analyses revealed distinct three-group patterns for SNPs (Chr01, 10, 13, 18), SVs (Chr01, 04, 10, 13, 19), and Indels (Chr01, 10, 11, 13, 18, 19). In total, 59,747 genes harbored haplotype-specific variants, with 34% containing multiple variant types (Supplementary Fig. S3B and Table S6).

To further assess genome-wide diversity, pairwise whole-genome alignments revealed a chromosome-stratified affinity map. While Chr04, 14, and 17 showed high global conservation, other chromosomes exhibited mosaic haplotype grouping: hap1 was distinct on Chr13 and 15; hap2 was divergent on Chr01; and hap2/hap3 clustered closely on Chr02,06, 08, 10, 16, and 18, where hap1/hap4 also showed high consistency on Chr18. This heterogeneous pattern, reflecting complex ancestry and selection, precludes simple autopolyploid or allopolyploid classification. Finally, although we detected 67.3% of genes retained in four copies and 9.8% of genes retaining as single copies (Supplementary Fig. S3C and Table S7), 8,765 single-copy genes were overlapped with private variants (Supplementary Fig. S3D), indicating that significant haplotype-specific loads persist even at loci under strong functional constraints.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

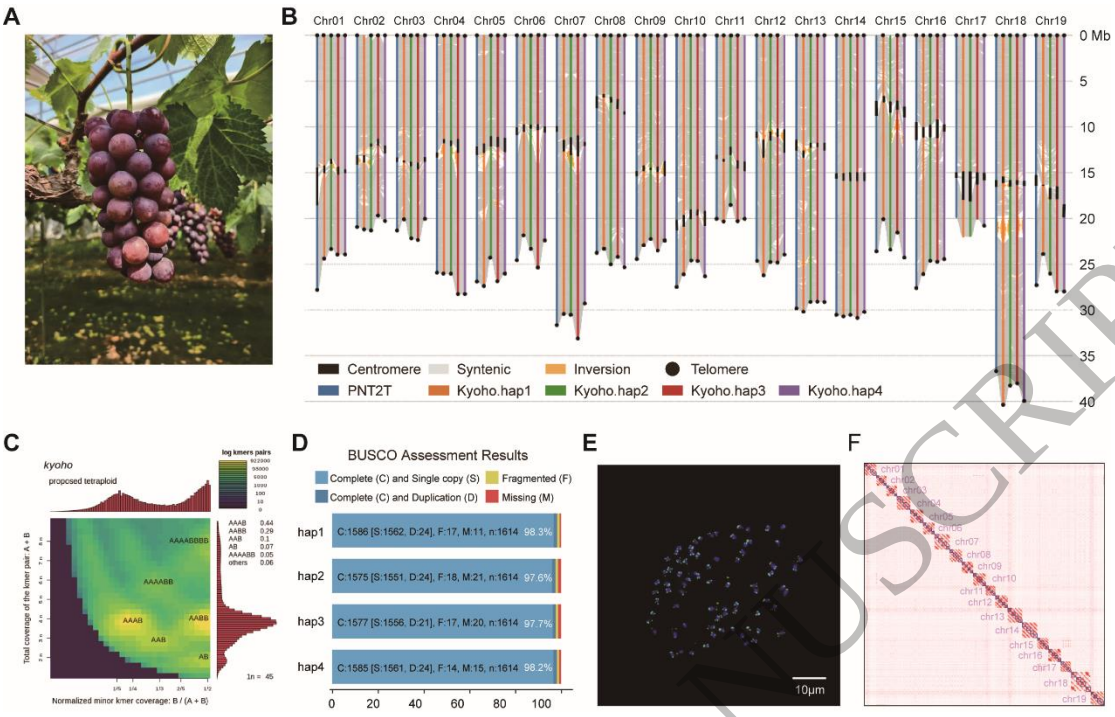


Figure 1. Genome assembly and characterization of the tetraploid Kyoho grape. A. Photograph of a mature Kyoho grape cluster. B. Chromosome-level collinearity analysis between the four Kyoho haplotypes and the reference genome (PNT2T). C. Smudgeplot analysis based on PacBio HiFi reads. D. BUSCO assessment of genome completeness for the four haplotypes. E. Fluorescence in situ hybridization (FISH) image of Kyoho chromosomes. The image shows the Kyoho chromosome set, with chromosomes counterstained in blue with DAPI. F. Hi-C interaction heatmap of the Kyoho genome.

The evolutionary genomic origin of Kyoho

To resolve the evolutionary origin of the Kyoho tetraploid, we first constructed phylogenomic trees for each chromosome using 31 *Vitis* taxa. While the genome overall reflects its known hybrid parentage from Ishihara Wase (*V. labrusca*) and Centennial (*V. vinifera*) (OIV, 2017) (Fig. 2A, Supplementary Fig. S4), a closer examination revealed striking heterogeneity in ancestral inheritance. While some chromosomes (such as Chr04) showed all four haplotypes grouping with the *V. vinifera* reference (PNT2T), indicating a purely *V. vinifera* origin for this homologous group. In contrast, other chromosomes displayed distinct mixed ancestries. For instance, the phylogenetic tree for Chr01 revealed a clear segregation: while haplotypes hap1, hap3, and hap4 clustered with *V. vinifera* (Group IV), hap2 diverged significantly, grouping with *V. labrusca* (Group I) (Fig. 2A).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

These conflicting phylogenetic signals across chromosomes (Fig. 2A, Supplementary Fig. S4) provided the initial evidence that Kyoho is not a uniform hybrid but possesses a complex, mosaic genomic structure.

By integrating these phylogenetic patterns with local genetic distance calculations, we deciphered the precise genomic architecture of Kyoho. Our analysis confirms that the genome is a complex mosaic lacking the uniform subgenome separation typical of strict allopolyploids. This mosaicism operates at two distinct organizational scales. First, at the chromosomal level, we observed varied combinations of parental haplotypes, ranging from pure *V. vinifera* groups to mixed configurations (Fig. 2A, Supplementary Fig. S4). Second, and most critically, we uncovered extensive intra-chromosomal recombination. At least 9 of the 19 chromosome groups contain chimeric haplotypes, single chromosomes that are themselves patchworks of *V. vinifera* and *V. labrusca* segments (Fig. 2B). The presence of these chimeric structures provides the definitive genomic signature of a segmental allopolyploid, verifying that partially homologous chromosomes from these divergent species paired and exchanged genetic material (homoeologous exchange) during the breeding history. Together, these findings establish that Kyoho was shaped by both the differential segregation of whole chromosomes and extensive interspecific recombination.

To quantify the overall genomic contribution, we aggregated the ancestry assignments across the genome. The Kyoho genome is composed of approximately 71.2% *V. vinifera* and 28.8% *V. labrusca* ancestry, biasing towards the *V. vinifera* parent (Supplementary Fig. S5A and Table S8). This ancestry partitioning independently corroborated by SNPs/Indels/SVs analysis. We found that chromosomal regions assigned to a *V. labrusca* origin harbored a significantly higher density of private SNPs/Indels/SVs compared to *V. vinifera* regions. Given that the variants were called against PNT2T, this density disparity perfectly aligns with the greater evolutionary divergence of the *V. labrusca* segments, validating our ancestry model (Supplementary Fig. S5B).

We further dissected the sex-determining region (SDR) on chromosome 2 to specifically characterize the ancestral origin of this functional locus (Zhou *et al.*, 2017). The genetic architecture within this ~150 kb region across the different haplotypes revealed a clear divergence into two distinct groups, corresponding to the male/hermaphroditic (H) and female (F) forms (Fig. 2C). By leveraging structural topology, we traced the specific parental lineages of these haplotypes. Notably, the SDR on Kyoho.hap1 was structurally homologous to the *V. labrusca* F haplotype (VL.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

hap1), whereas Kyoho.hap4 corresponded to the *V. vinifera* F haplotype (PNT2T). Conversely, both Kyoho.hap2 and Kyoho.hap3 aligned with the *V. labrusca* H haplotype (VL. hap2). These results elucidate the hybrid composition of the SDR locus and confirm that Kyoho is a hermaphrodite (HHFF) carrying both haplotype types.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

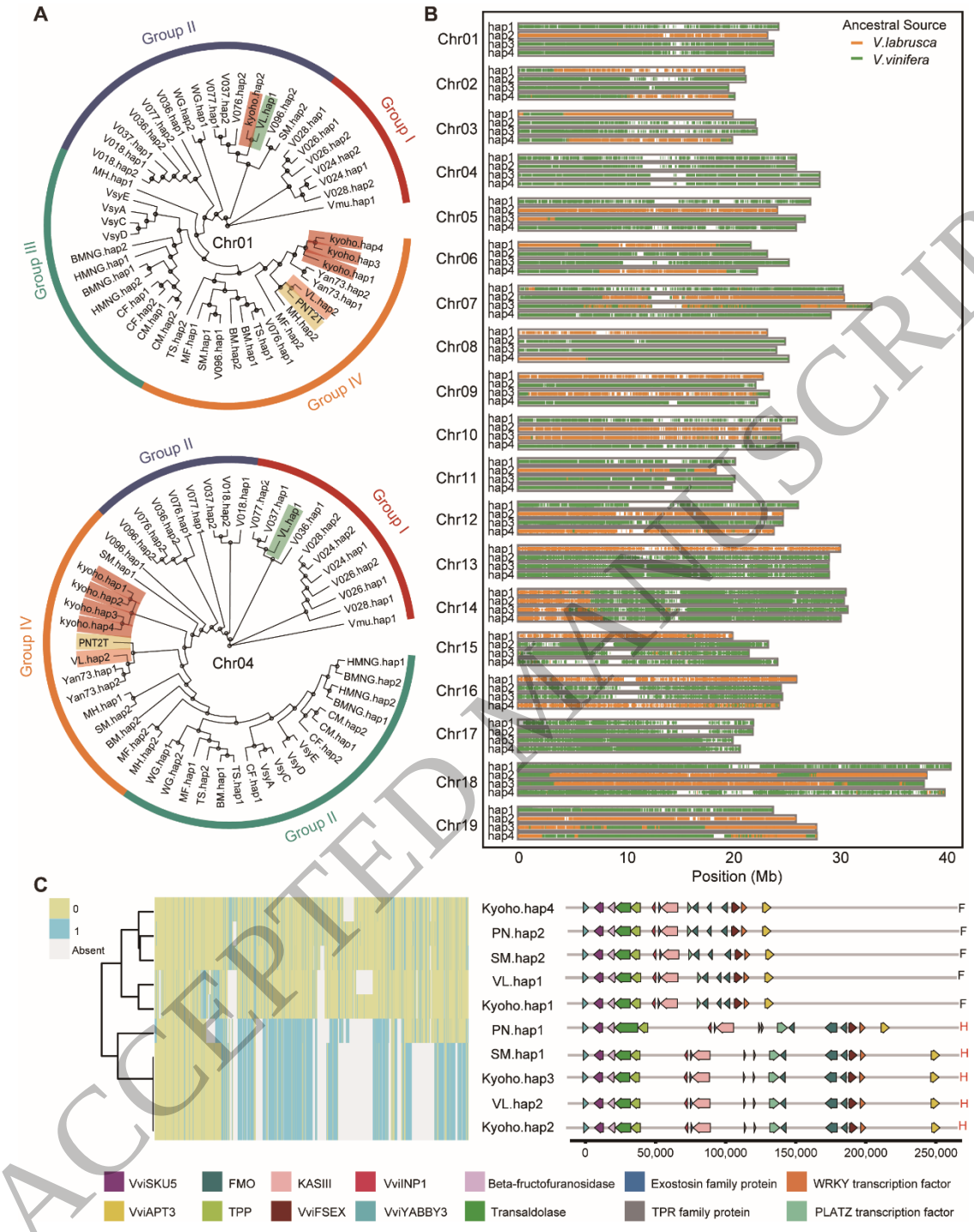


Figure 2. Evolutionary genomic origin and haplotypic composition of Kyoho. A. maximum likelihood phylogenetic tree using single-copy genes from chromosome Chr01 and Chr04 categorizes haplotypes into four groups (see other chromosomes in Figure S4). Group I: Approximate *V. rotundifolia*, Group II: Approximate *V. labrusca*, Group III: Approximate *V. vinifera*, and Group IV: the complete *V. vinifera* (the higher the proportion of single-copy genes

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

from *V. vinifera*, the closer it is to *V. vinifera*). B. Genomic landscape of ancestral segments across the 19 chromosomes of the four Kyoho haplotypes. Fragments inferred to originate from *V. labrusca* are colored orange, and those from *V. vinifera* are colored green, illustrating the mosaic composition of the genome. C. The sex-determination region in Kyoho haplotypes. The heatmap and dendrogram (left) compare this region across different Kyoho and reference haplotypes. The schematic diagram (right) illustrates the organization of candidate functional genes within this locus, linking specific haplotypes to either hermaphrodite (H) or female (F) phenotypes.

Transcriptomic landscape of the Kyoho mosaic genome

Having established the complex mosaic architecture of the Kyoho genome, we next sought to determine how this complex ancestry translates into functional gene expression. We generated comprehensive transcriptome profiles from leaf and three fruit developmental stages, identifying 20,480 homologous quadruplets (Figs. 3A, 3B; Supplementary Table S9). Focusing on 17,186 robustly expressed quadruplets, we categorized their regulatory patterns into 15 distinct motifs (Supplementary Fig. S6). The regulatory landscape was defined by striking local asymmetry; balanced expression was rare (3.0%), whereas asymmetric modes were predominated. These asymmetric patterns were primarily classified into three categories: single-haplotype dominance (42.7%), two-haplotype dominance (43.6%), and single-haplotype suppression (10.7%). Specifically, single-haplotype dominance was relatively evenly distributed among the four haplotypes (ranging from 9.7% to 11.3%), while two-haplotype dominance showed significant variation, with the hap1/hap2 combination being the most prevalent (13.9%). Collectively, these findings reveal a strong expression bias in Kyoho alleles, whereas haplotype-dominant or two-haplotype dominant patterns overwhelmingly surpass balanced expression. Functional enrichment revealed that specific dominant haploids drive distinct biological modules; for instance, hap1*Vitis000012*_hap2*Vitis000206*(*MADS-box*); *hap1Vitis0076_hap3Vitis007617*(*PSY*); *hap2Vitis000205_hap4Vitis000209*(*PL-ATZ*) preferentially regulate floral/embryonic development and light response pathways, respectively (Supplementary Table S10). Further analysis of the four homologous gene copies in the SDR region revealed no antagonistic effects between the genes from *V. labrusca* and *V. vinifera*; instead, both sets of genes were found to contribute positively to the observed traits (Supplementary Table S11).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

To further evaluate whether there is any systematic deviation from a random, symmetrical relationship between *V. labrusca* and *V. vinifera* genomes in the Kyoho genome, we evaluated the competing influences of tissue identity versus genomic origin. We systematically compared 68,164 homologous gene pairs (representing pairwise alignments between *V. labrusca*-derived and *V. vinifera*-derived alleles) by integrating differentially expressed genes (DEGs) (Supplementary Fig. S7 and Table S12). To comprehensively investigate the regulatory divergence between *V. labrusca*- and *V. vinifera*-derived alleles, we classified the differentially expressed homologous pairs into eight distinct patterns based on their expression trajectories (D-D, U-U, D-N, N-D, U-N, N-U, D-U, U-D) (Supplementary Fig. S8). Strikingly, a substantial proportion, ranging from 48.7% to 75.8% across different comparisons, exhibited directionally inconsistent expression patterns (e.g., D-N, N-D, U-N, N-U, D-U, and U-D). This widespread prevalence of inconsistent regulatory responses indicates that allelic pairs do not strictly adhere to a synchronized expression program. Instead, they display significant plasticity, suggesting that the *V. labrusca* and *V. vinifera* alleles independently tune their contributions to meet specific developmental demands (Supplementary Table S13).

Finally, we quantified the stability of these regulatory patterns to understand their evolutionary implications. Among 17,750 homologous gene pairs exhibiting directional bias, only 4,086 (23.4%) showed constant bias (maintaining a fixed dominance direction across all tissues). The vast majority (13,227 pairs, 74.5%) displayed semi-constant bias (biased in some tissues but balanced or silent in others). This distribution confirms that allelic dominance in Kyoho is largely conditional and plastic rather than genetically fixed (Fig. 3C). We further explored this plasticity by analyzing 18,162 tissue-specifically expressed genes (TPGs), which Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction resolved into four distinct co-regulatory archetypes (Supplementary Fig. S9). These genes were categorized into single-origin specific ('Vv' or 'Vl') and shared ('both') patterns (Fig. 3D). As shown in the expression quantification, while a subset of TPGs retains co-specificity ('both'), a significant proportion undergoes subfunctionalization, where expression is exclusively driven by either the *V. vinifera* or *V. labrusca* allele depending on the tissue. The heatmap trajectory (Fig. 3E) vividly illustrates these regulatory archetypes, ranging from strict co-enrichment to mutually exclusive tissue specificity. Collectively, these patterns demonstrate that tissue developmental context, rather than rigid genomic bias, is the primary driver of allelic divergence. By flexibly recruiting distinct ancestral

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

alleles for specific physiological roles, Kyoho maximizes the adaptive potential of its mosaic genome through targeted subfunctionalization.

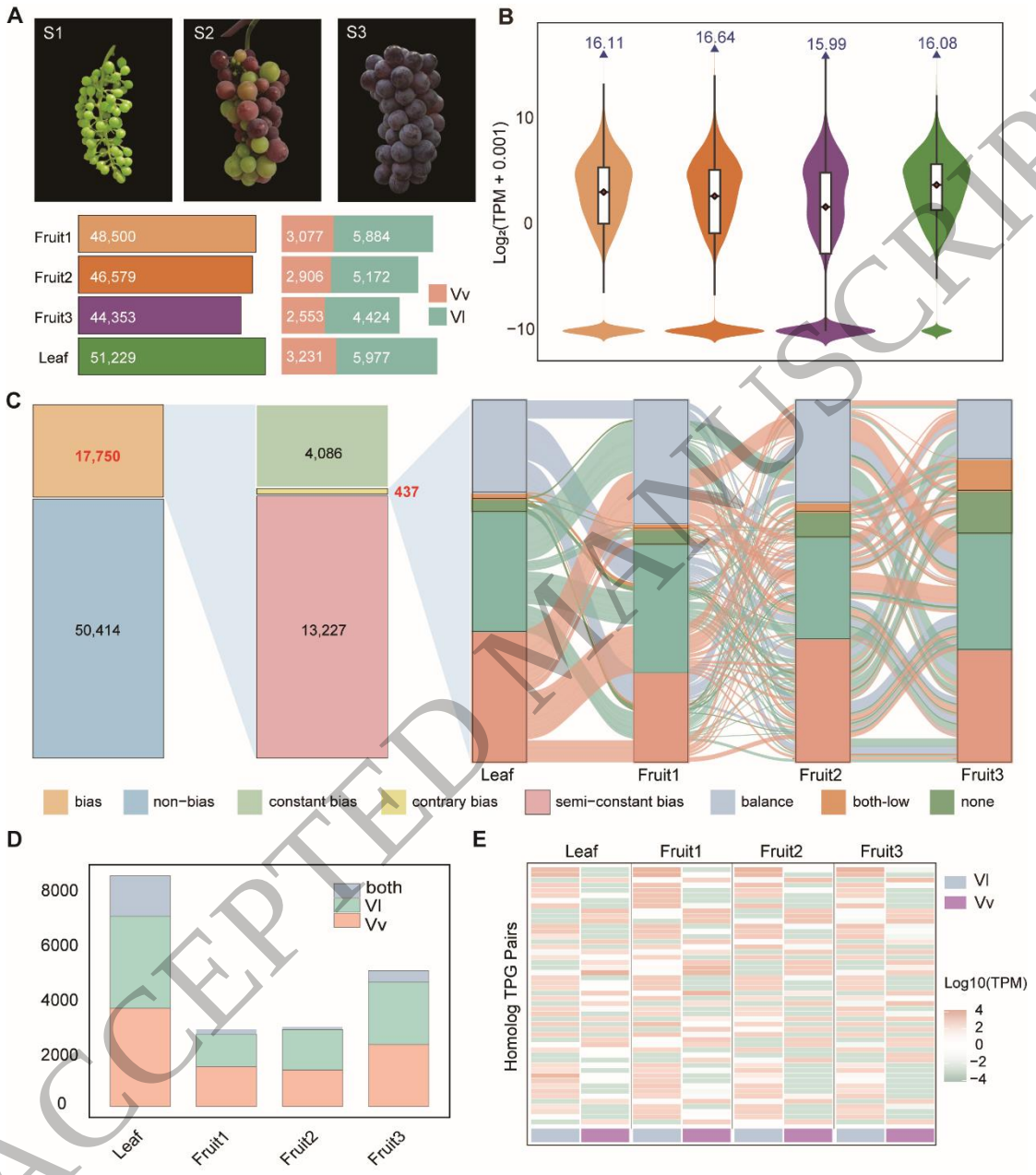


Figure 3. Parentalexpression bias in tetraploid Kyoho grape. A. Fruit at three developmental stages: S1 (young), S2 (veraison), and S3 (mature); The left panel shows the total number of expressed genes, and the right panel shows the number of genes with parental expression bias ('Vl' and 'Vv' represent *V. labrusca-bias* and *V. vinifera-bias*, respectively). B. Violin plots of gene expression levels (log2 TPM) and bar charts of expressed gene counts for each tissue and stage. C. Overall classification (left) and dynamic shifts (right) in parental expression bias across tissues,

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

visualized with an alluvial plot. The alluvial plot on the right systematically illustrates the dynamic transition trajectories of homologous gene pairs in terms of expression preference states during leaf and three fruit development stages, described by flow direction and bifurcation. Each flow line represents a gene pair and depicts its switching path between five possible expression states – no expression, both-low expression (both alleles), VL dominance, Vv dominance, or balanced. D. Tissue-specific homologous gene pair pattern statistics ('Vl' = *V. labrusca* expression; 'Vv' = *V. vinifera* expression; 'both' = bilateral simultaneous expression). E. Heatmap of expression levels for the top 50 homologous gene pairs from the two ancestries across all samples.

DNA methylation landscape of the Kyoho mosaic genome

To elucidate the epigenetic landscape underpinning the Kyoho mosaic genome, we performed whole-genome bisulfite sequencing across leaf and three sequential fruit developmental stages (S1-S3). While CG and CHG methylation levels remained high and stable globally, the dynamic landscape was defined by a substantial increase in the CHH context during fruit maturation (Supplementary Fig. S10). The average CHH methylation levels rose more than five-fold, from 1.45% in the S1 stage to 7.65% in the S3 stage (Fig. 4B). Consistent with previous studies (Peng et al., 2025, Song et al., 2024b), TEs exhibited the highest average methylation (49.3% in CG context), significantly surpassing genic regions (29.97%) (Supplementary Fig. S10 and Table S14). Overall, this specific accumulation of CHH methylation signals a robust activation of the RNA-directed DNA methylation (RdDM) pathway, functioning as a global genomic stabilizer during the metabolically active ripening process (Supplementary Fig. S11).

We next examined whether the epigenetic landscape is symmetrical between the constituent ancestries (*V. vinifera* vs. *V. labrusca*). Weighted methylation analysis revealed a consistent quantitative bias: *V. vinifera*-derived origin displayed higher methylation levels than their *V. labrusca* counterparts across genome. This asymmetry was most pronounced in the CHH context within TEs. For instance, at the S3 stage, the mean CHH methylation level of *V. vinifera* TEs reached approximately 11.5%, distinctly surpassing the 9.8% observed in *V. labrusca* TEs (Supplementary Fig. S12 and Table S15). This divergence likely attributes to the differential TE loads accumulated during the distinct evolutionary histories of the two species. Crucially, however, this ancestral offset was subordinate to the overarching developmental program. As fruit development progressed, methylation levels in both ancestries increased synchronously, following

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

parallel upward trajectories Vl (0.009~0.676): Vv (0.010~0.670) (Supplementary Fig. S13). This finding indicates that the Kyoho genome employs a tightly coordinated regulatory system where developmental drive serves as the primary determinant, while genomic origin plays a subtle modulatory role.

To dissect the mechanism driving these patterns, we analyzed methylation distribution at high resolution. We found that methylation peaks, particularly in the CHH context, colocalized strongly with TE-dense regions regardless of their ancestral origin (Fig. 4A, Supplementary Fig. S14). At the level of individual homologous gene pairs, we uncovered a striking temporal dynamic: significant methylation divergence between *V. vinifera* and *V. labrusca* alleles was not constitutive but transient. Divergence peaked specifically during the fruit expansion stage (S2, $P < 0.05$) before re-converging (Fig. 4C). Comparative analysis across tissues and developmental stages classified gene pairs into two distinct categories: conserved pairs, which maintain consistent methylation states (such as BM-BM, IM-IM, and UM-UM), and remodeled pairs, which exhibit dynamic state transitions (including BM-IM, BM-UM, UM-BM, UM-IM, IM-BM, and IM-UM) (Fig. 4D, Supplementary Table S16). This stage-specific plasticity underscores a dynamic strategy of transient allelic divergence. By temporarily relaxing epigenetic synchrony to accommodate the rapid metabolic flux of expansion, the Kyoho genome effectively facilitates the targeted transcriptional subfunctionalization required for adaptive hybrid development.

Finally, to identify the functional targets of this epigenetic control, we examined genes located in hypermethylated islands. Enrichment analysis revealed a significant overrepresentation of gene families associated with genome defense (e.g., retrotransposon-derived proteins) and large, tandemly duplicated clusters such as receptor-like kinases (RLKs) and proteases (Supplementary Table S16). The hypermethylation of these loci is primarily driven by their proximity to or containment of TEs. This underscores a fundamental trade-off in the hybrid genome: DNA methylation acts as a genomic guardian, strictly silencing repetitive elements to maintain integrity, even if it means imposing a repressive epigenetic environment on adjacent multi-copy gene families.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

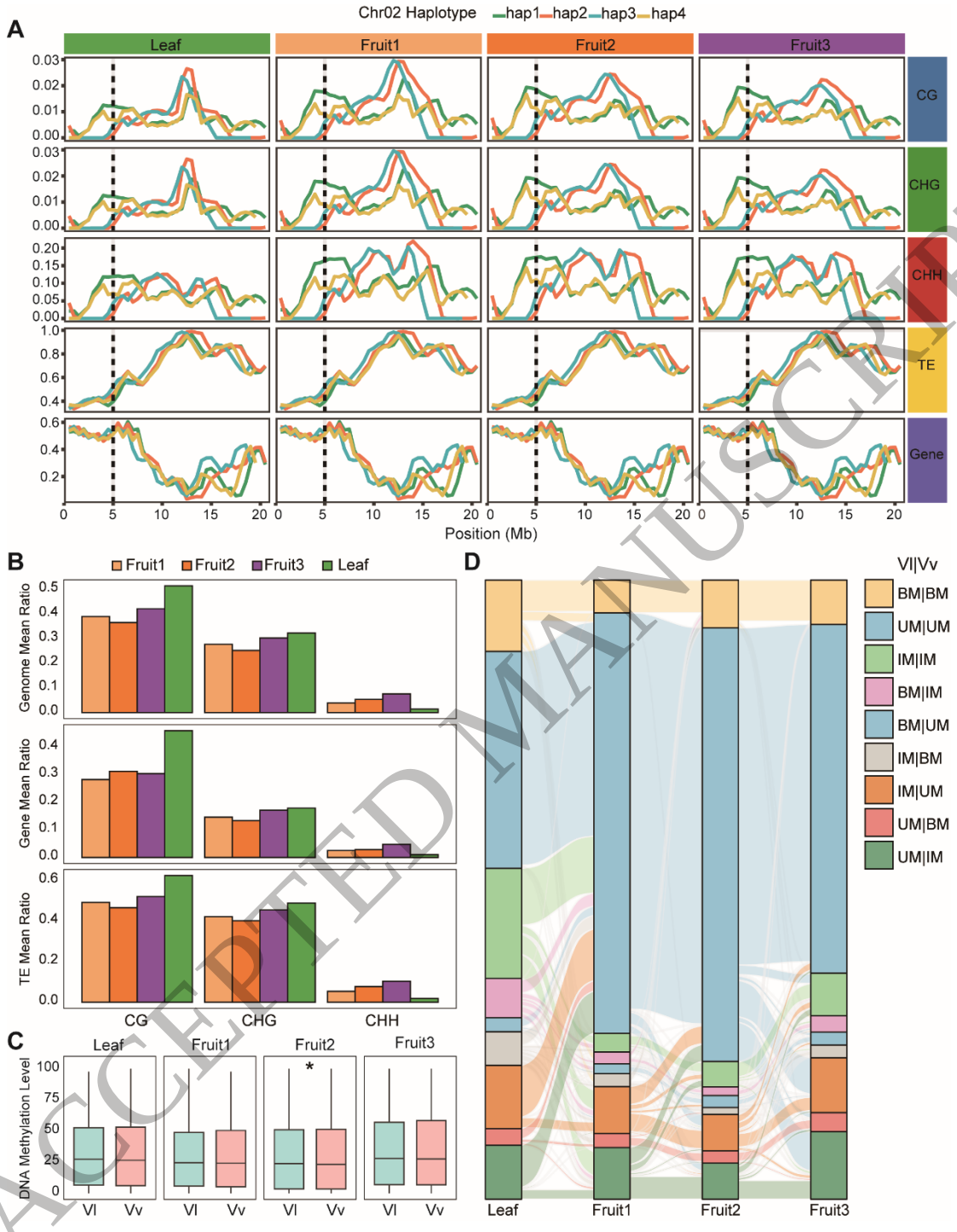


Figure 4. Allelic-specific DNA methylation dynamics and parental bias in Kyoho. A. Distribution of DNA methylation (in CG, CHG, and CHH contexts), transposable element (TE), and gene densities along chromosome 2 for the four phased copies across leaf and three fruit developmental stages. B. Average weighted methylation levels across the genome, gene bodies, and TEs. C. Comparison of gene body methylation between homologous alleles from the *V.*

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

labrusca (Vl) and *V. vinifera* (Vv). D. Alluvial plot showing the transitions of gene body methylation (gbM) states in homologous gene pairs across tissues and developmental stages. Gene body methylation was discretized into three states: Body Methylated (BM; highly methylated gene bodies), Intermediately Methylated (IM; moderate gbM levels), and Unmethylated (UM; low or negligible gbM). For each homologous pair, combined states include symmetric patterns, such as both methylated (BM|BM) and both unmethylated (UM|UM), as well as asymmetric configurations between the *V. labrusca* (Vl) and *V. vinifera* (Vv) ancestries (e.g., BM|UM, BM|IM, IM|UM).

Alternative splicing landscape of the Kyoho during fruit development

To capture the post-transcriptional complexity of the Kyoho genome, we integrated full-length Iso-Seq and short-read RNA-Seq data, establishing a comprehensive catalog of 82,750 alternative splicing (AS) events. Intron retention (RI) emerged as the dominant mode (67.1%), consistent with its prevalence in plants (Fig. 5A). The AS landscape exhibited a distinct temporal dynamic during fruit development. The transition from early fruit expansion (S1 to S2) involved moderate splicing modulation, including 2,510 differentially alternatively spliced (DAS) events. In contrast, the transition to ripening (S2 to S3) was characterized by a marked upregulation in splicing regulation, generating 3,820 DAS events (Fig. 5B). This intensification suggests that AS serves as a critical mechanism for proteomic diversification specifically during the complex metabolic shifts of ripening. Concurrently, isoform usage shifted toward exon inclusion during late development; notably, comparison of splicing inclusion difference (dPSI) revealed a significant increase in positive dPSI values for RI events in the S3 vs. S2 comparison (Fig 5C and 5E), indicating a stage-specific regulation of transcript processing.

We investigated the relationship between DAS events and differentially expressed genes (DEGs). The overlap between these two regulatory layers was limited to 11.9%, suggesting that AS operates largely as an independent regulatory mechanism rather than merely mirroring transcriptional changes. Correlation analysis between expression differences (log_2FC) and inclusion level differences (dPSI) revealed no significant linear correlation ($r \approx 0$) across developmental stages (Fig. 5F). However, an inverse relationship was frequently observed where increased isoform inclusion coincided with gene downregulation (58.9%) (Fig. 5F), a pattern consistent with nonsense-mediated decay (NMD) feedback mechanisms. Furthermore, temporal analysis indicated

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

that splicing modulation often precedes steady-state changes in gene expression (Fig. 5G), positing AS as a rapid response mechanism for fine-tuning the functional proteome prior to substantial transcriptional reprogramming.

To determine how the mosaic ancestry influences post-transcriptional regulation, we analyzed homologous alleles derived from *V. labrusca* and *V. vinifera* (Supplementary Fig. S15-17). The distribution of AS event types was consistent across all four haplotypes, with RI remaining the dominant category (Supplementary Fig. S15). Comparison of homologous gene pairs revealed a dual regulatory strategy characterized by both evolutionary conservation and functional divergence. On one hand, a high degree of conservation was evident in the splicing code; inclusion levels (dPSI) between *V. vinifera* and *V. labrusca* homologs were strongly correlated (Pearson r = 0.766, $P < 0.001$; Supplementary Fig. S16D). The majority of events showed identical splicing modalities (e.g., RI in both ancestors), as indicated by the predominance of counts along the diagonal in event-type combination matrices (Supplementary Fig. S16C and S16B).

On the other hand, substantial divergences were identified where homologs adopted distinct splicing fates. Specific loci exhibited differential processing, such as RI in one allele pairing with Exon Skipping (SE) or Alternative 3' Splice Site (A3) usage in its homolog (Supplementary Fig. S17B). This divergence allows the hybrid genome to bypass diploid constraints, enabling one allele to maintain the canonical isoform while the other explores novel splicing variants. Pathway enrichment analysis suggested this divergent splicing is selectively recruited for key developmental processes. For instance, allele-specific intron retention was observed in auxin efflux carriers, and alternative 3' splice site usage was detected in GRF-interacting factors. These findings indicate that differential processing of ancestral alleles contributes to the subfunctionalization of genes regulating traits such as organ size and ripening timing (Supplementary Table S17).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

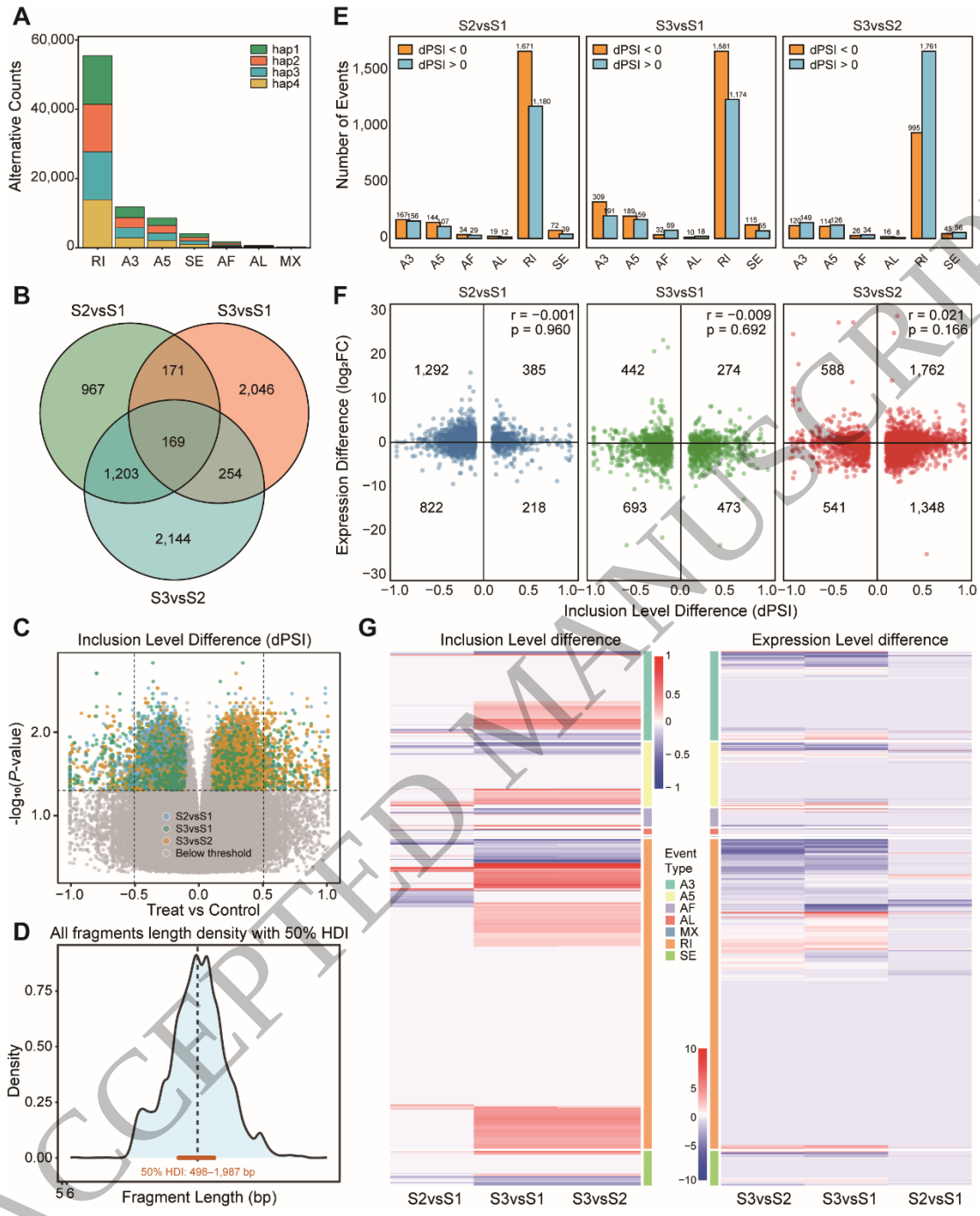


Figure 5. Iso-Seq and RNA-seq of alternative splicing dynamics during grapevine fruit development. A. Counts of alternative splicing (AS) event types identified across the four haplotypes. B. Venn diagram of differentially spliced (DAS) events overlapping between developmental stage comparisons. C. Volcano plot of differential AS events, showing the change in percent spliced-in (dPSI) versus statistical significance. D. Density plot of variable fragment lengths, indicating the 50% highest density interval (HDI). E. Number of significant DAS events

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

per type, showing increased (dPSI > 0) or decreased (dPSI < 0) inclusion for each stage comparison. F. Correlation between splicing inclusion difference (dPSI) and gene expression change (log2FC) for DAS events. G. Heatmaps of dPSI values (left) and corresponding gene expression changes (log_2FC, right) for significant DAS events, clustered by event type.

Integrative multi-omics reveal allele-specific gene regulation

To dissect allele-specific gene regulation, we integrated multi-omics data covering genomic variants (SNPs/Indels/SVs), DNA methylation, transcriptomic changes, and alternative splicing (i.e., HSV, HGBM, DEHG, and HDAS) (Fig.6A). Structural variation emerged as a primary determinant of transcriptional abundance, evidenced by the substantial overlap of 15,032 genes exhibiting both HSV and DEHG. However, the regulatory landscape is not a simple binary output of genomic structure; rather, it exhibits distinct stratification. Transcriptional divergence appeared most plastic, with 21,476 genes showing unique expression changes (Unique DEHG) independent of large structural shifts, epigenetic modifications, or alternative splicing. Conversely, 5,860 genes harbored silent SVs (Unique HSV) that did not trigger downstream regulatory changes, highlighting the genome's capacity to buffer against structural perturbations.

At the intersection of these layers, we identified a core module of 772 genes that exhibited convergence across all four regulatory signals (HSV, DEHG, HGBM, and HDAS) (Fig. 6B and Fig. 6C). This multi-layered coordination suggests that for this specific gene set, structural differentiation anchors a synchronized regulatory recalibration to drive robust sub functionalization. Beyond this core set, we detected a significant group of 2,307 genes showing coordinated HDAS and DEHG changes in the absence of dynamic methylation (HGBM). Visual examination of specific candidates, including *DOG1* (*hap1Vitis019733_hap3Vitis019798*), a sugar transporter (*hap1Vitis023513_hap2Vitis023241*), Abscisic acid receptor *(hap4Vitis024149_hap2Vitis024109*), *ZEP* (*hap1Vitis020191_hap3Vitis020204*), *PIN* (*hap1Vitis011230_hap2Vitis011150)* and a *MYB* transcription factor *hap1Vitis013173_hap2Vitis013123),* confirmed a strong mechanistic link between splicing events and homoloogus gene expression divergence (Fig. 6D; Supplementary Fig. 18 and Table S18). Despite this coordination, the splicing machinery retains the capacity for independent modulation. A distinct subset of genes (including 373 unique DAS events) exhibited differential splicing in the absence of other regulatory changes. This independent mode serves as a rapid response unit,

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

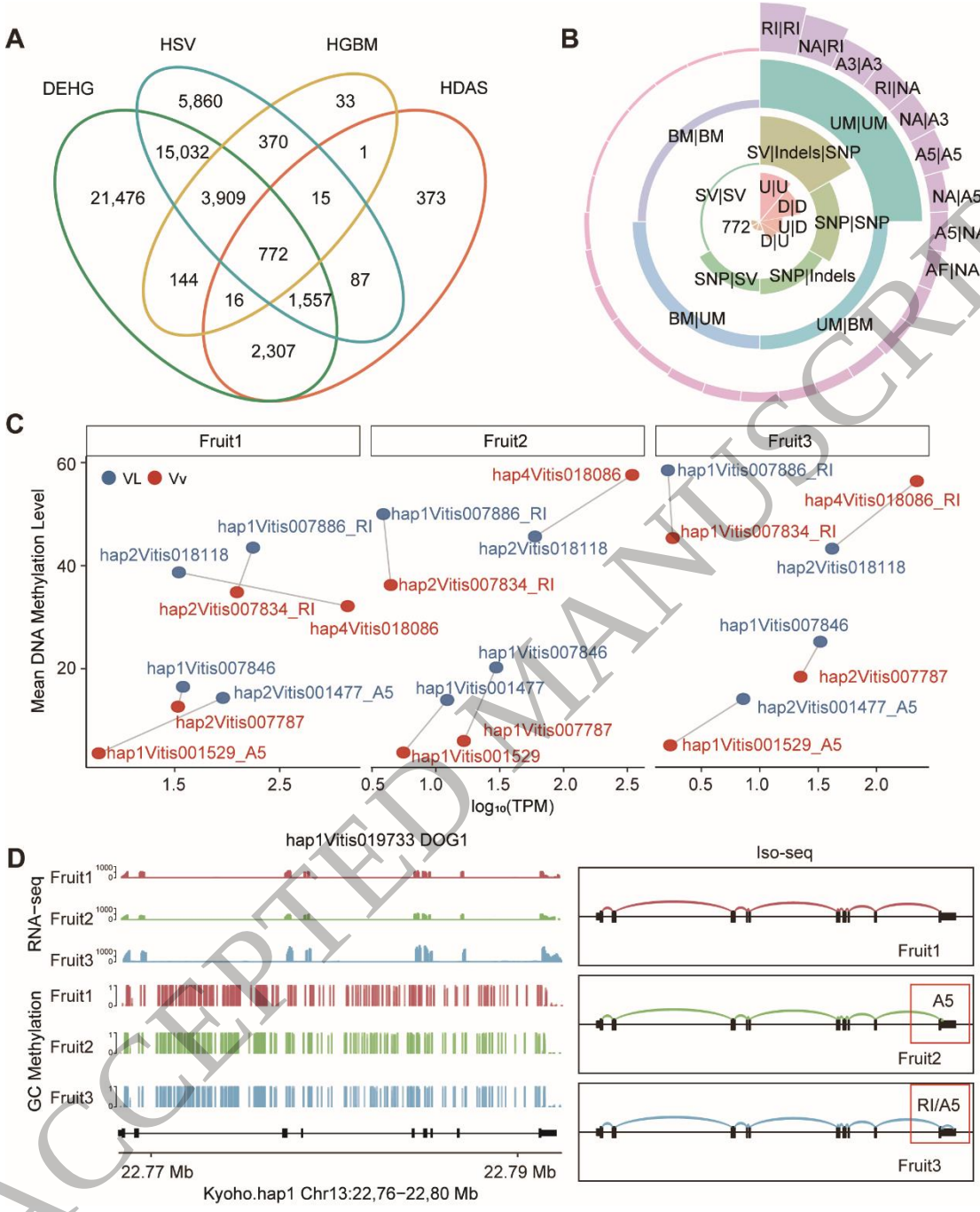
particularly during the transition to ripening, exemplified by isoform switching in *hap2Vitis018118* (Fig. 6C), generating proteomic diversity without requiring the slower, energy-intensive overhaul of the transcriptional baseline. In contrast, DNA methylation functioned primarily as a dependent stabilizer; only 33 genes exhibited unique methylation differences, whereas 3,909 genes tracked with structural and expression divergence.

Critically, while these loci lacked dynamic methylation changes, our analysis revealed that splicing stability is intrinsically linked to the static methylation landscape. By analyzing data from different fruit development stages, we focused on key regulatory mechanisms of fruit development, without considering gene integration associated with ancestral structural variations. This structure is based on the analysis of 6,415 differential splicing events, 50,537 DEGs, and 39,095 dynamic methylation sites (Supplementary Fig. S19A and Table S19). We observed that UM were significantly more prone to alternative splicing BM (Supplementary Fig. S19 B and Table S19). The observed splicing events, such as RI and SE, preferentially mapped to hypomethylated regions (Supplementary Fig. S19C). This correlation supports a model in which a stable low-methylation environment ensures sufficient chromatin accessibility. Consequently, this allows the splicing machinery to generate alternative isoforms, facilitating functional specialization in the absence of dynamic epigenetic remodeling. In addition, functional enrichment, *HRGP*, *TBL*, *ZEP*, *PSY*, *SPPase*, *GRAS*, and *Aux/IAA* are involved in a variety of key physiological processes, including cell wall remodeling, pigment synthesis, energy metabolism, transcriptional regulation, and hormone signal transduction, which together maintain the overall regulatory mechanism of fruit development (Supplementary Fig. S20).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Figure 6. Integrative multi-omics analysis of homologous gene pairs. A. Venn diagram showing



the overlap of differential expressed homologous genes (DEHG), genes with differential alternative splicing (HDAS), homologous genes with differential gene-body methylation (HGBM), and homologous genes with structural variants (HSV). B. Sunburst plot illustrating multi-omic divergence between ancestral alleles, including alternative splicing, methylation (Display of homologous differential alternative splicing event types with counts greater than 5), SV, and expression. The concentric rings represent hierarchical expansion only, and categories in

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

different rings are not in a one-to-one correspondence. C. Gene expression, DAS and DNA methylation for homologous gene pairs from the q (blue, *V. labrusca*) and s (red, *V. vinifera*) subgenomes across fruit stages. D. Genome browser view of a candidate gene showing RNA-seq coverage, CG methylation, and Iso-Seq isoforms. An alternative 5' splice site (A5) or intron retention (RI) event is highlighted (red box).

Discussion

In this study, we generated a haplotype-resolved, near-complete genome for the tetraploid Kyoho grape, serving as a landmark resource to dissect regulatory logic of a complex segmental allopolyploid. Our integrative analysis highlights three key insights: (1) we quantitatively resolved its highly asymmetric, mosaic genomic architecture; (2) we overturned the assumption of global ancestral dominance—where alleles from one progenitor consistently out-express the other—revealing instead a flexible, tissue-adaptive regulatory model; (3) we established AS, rather DNA methylation, as the primary driver of functional divergence between ancestral alleles. Collectively, this work provides a new model for understanding how stability and plasticity are maintained in complex polyploid genomes under selective pressures, which will facilitate future polyploid breeding efforts.

Unlike classic allopolyploids with clearly demarcated subgenomes (Wang *et al.*, 2019, Zhang *et al.*, 2015, Hu *et al.*, 2019, Song et al., 2024b, Zhou *et al.*, 2023)), Kyoho is a segmental allopolyploid shaped by extensive inter-ancestral recombination. Its homologous chromosomes are themselves chimeras of *V. vinifera* and *V. labrusca* ancestry. This structural mosaicism challenges the rigid subgenome-dominant models often applied in polyploid studies (OIV, 2017, Chalhoub *et al.*, 2014). Instead, the extensive recombination between ancestral genomes breaking linkage drag (Lynch & Force, 2000) and creating a heterogeneous substrate that necessitates the flexible, chimeric regulatory patterns we subsequently identified.

The functional output of this mosaic genome is governed by dynamic, developmental context-driven trade-offs rather than a static hierarchy. The vast majority of homologous gene pairs exhibiting overwhelming allele-specific expression bias (biased pairs) display variable dominance across tissues. This suggests that the observed expression patterns do not merely reflect an averaging of ancestral differences. Instead, artificial or natural selection has favored the specific ancestral alleles that are most advantageous for given physiological demands (Yoo et al., 2013).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

This independence is particularly critical during fruit maturation, where specific ancestral lineages confer distinct adaptive traits. By retaining conditional dominance, artificial and natural selection have shaped a regulatory landscape that provides both functional robustness and evolutionary plasticity through lineage-specific subfunctionalization (Lynch & Force, 2000).

Beyond transcriptional abundance, DNA methylation acts as a global synchronizer rather than a driver of ancestral divergence. Although *V. vinifera* regions exhibit higher baseline methylation due to differential TE loads (Matzke & Mosher, 2014), both ancestries undergo a synchronized surge in CHH methylation (RdDM) during ripening. This global reinforcement silences the chaotic TE landscape (Tang *et al.*, 2020, Huang *et al.*, 2019, Song *et al.*, 2024a), ensuring that the underlying structural asymmetries do not disrupt the developmental program. Crucially, a transient plasticity window at the S2 stage, where methylation divergence peaks before re-synchronizing, suggests that the epigenome serves as a dynamic filter. It maintains general genomic integrity while permitting brief regulatory relaxation to accommodate the rapid metabolic shifts of hybrid fruit development (Slotkin & Martienssen, 2007, Li *et al.*, 2023).

In contrast to the global synchrony of the methylome, AS is a rapid, orthogonal regulatory layer driving lineage-specific functional divergence (Reddy, 2007, Baralle & Giudice, 2017, Tognacca *et al.*, 2023, Reddy *et al.*, 2013, Alyahya & Taybi, 2024). Operating largely independently of transcriptional abundance, AS resolves a key regulatory puzzle: for thousands of loci lacking explanatory genomic variants or methylation differences, AS acts as the primary determinant for modulating allele-specific expression. Selection has favored this mechanism to fine-tune essential physiological pathways, evidenced by allele-specific splicing patterns in core ripening regulators, including *ZEP* (abscisic acid biosynthesis/ receptor), *Aux/IAA* (auxin signaling), *DOG1* (dormancy regulation), and *Myb* transcription factor (Guo *et al.*, 2021, Alyahya & Taybi, 2024, Hirschberg, 2001). By leveraging these lineage-specific splicing signatures, the hybrid genome bypasses rigid DNA sequence constraints, enabling the precise recruitment of advantageous alleles to coordinate fruit maturation.

Importantly, the AS-driven regulatory strategy is not unique to grape. In other polyploids, subgenome- or allele-biased AS, alongside epigenetic remodeling, repeatedly emerges as a key driver of evolutionary divergence. For example, asymmetric AS profile shaped subgenome adaptation in hexaploid wheat and *Brassica napus* (Gao *et al.*, 2021, Yao *et al.*, 2020, Rousseau-

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Gueutin *et al.*, 2020, de Jong & Adams, 2023, Wang & Wang, 2022). In parallel, polyploid tobacco and *Arabidopsis* illustrate extensive epigenomic and TE reconfiguration during adaptation (Mhiri *et al.*, 2019, Yuan *et al.*, 2020, Jiang *et al.*, 2021, Srikant *et al.*, 2024). Together, these studies indicate that allele-specific AS is a primary driver of functional divergence between ancestral alleles during development.

In conclusion, the haplotype-resolved, multi-omics Kyoho dataset provides a powerful toolkit to overcome long-standing challenges in polyploid crop improvement, facilitating the transition to haplotype-aware breeding. Our findings enable the development of markers tracking specific ancestral segments with desirable traits, such as the *V. labrusca*-derived disease resistance (Cadle-Davidson, 2008) or *V. vinifera*-derived flavor, across a complex recombination landscape. Moreover, the discovery that valuable alleles are mobilized via conditional plasticity implies that latent ancestral traits can be activated through targeted environmental or developmental cues. Ultimately, this work serves as a conceptual and practical blueprint for dissecting complex traits and accelerating genetic gain in segmental allopolyploid crops.

Materials and methods

Plant material and growth conditions

Kyoho grapevines were cultivated in the vineyards of the Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences (Shenzhen, Guangzhou, China). For karyotyping, one-year-old dormant shoots were collected and propagated in 5x5 cm pots in a greenhouse at 26°C to stimulate the growth of active root meristems. For sequencing, samples of leaves, buds, roots, and fruits at three distinct developmental stages (S1: young green, S2: veraison/expansion, S3: ripening) were collected, flash-frozen in liquid nitrogen, and stored at -80°C until use.

Genomic DNA sequencing

High-molecular-weight (HMW) genomic DNA was isolated from a mixture of young leaves, buds, and roots using a modified CTAB/SDS phenol-chloroform extraction protocol combined with a Qiagen column-based purification step. The integrity and purity of the HMW DNA were confirmed before library construction. A PacBio HiFi library was constructed and sequenced on a PacBio Sequel II platform, generating 163.9 Gb of high-fidelity (HiFi) reads. An Oxford Nanopore Technology (ONT) library was prepared and sequenced on GridION X5 and PromethION

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

platforms, yielding 44.9 Gb of ultra-long reads. For chromatin conformation capture, fresh leaves were cross-linked with 2% formaldehyde, and a Hi-C library was prepared by digesting chromatin from isolated nuclei with the *DpnII* restriction endonuclease. The library was sequenced on an Illumina platform (PE150), producing approximately 50 Gb of paired-end reads.

Transcriptome and epigenome sequencing

Total RNA was extracted from leaves and the three fruit developmental stages (S1, S2, S3). For each sample, three biological replicates were utilized for short-read sequencing and two replicates for long-read sequencing.. Two complementary transcriptome sequencing strategies were employed: (1) Illumina short-read RNA-seq libraries were prepared after rRNA depletion (Shanghai Personalbio, China) and sequenced on an Illumina platform, yielding ~12 Gb per library. (2) PacBio full-length cDNA (Iso-Seq) libraries were prepared (Hope Bio, China) and sequenced on a PacBio Sequel II platform, producing a total of ~400 Gb of subreads. For DNA methylation profiling, whole-genome bisulfite sequencing (WGBS) was performed on the same tissues and stages (two biological replicates each). Libraries were constructed and sequenced by Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China) on an Illumina platform, yielding approximately 96 Gb of raw data.

Genome assembly and annotation

To assemble a haplotype-resolved, near-T2T reference genome, we employed a hierarchical strategy. First, the genome size and heterozygosity were estimated from HiFi reads by analyzing the 21-mer frequency distribution with Jellyfish (v2.3.0) (Marcais *et al.*, 2018) and GenomeScope (v2.0) (Ranallo-Benavidez *et al.*, 2020). Then, a haplotype-resolved assembly was generated using hifiasm (v0.19.7-r598) (Cheng *et al.*, 2021) with HiFi and ONT reads, using Hi-C data to phase the four haplotypes (--n-hap 4). The resulting contigs were anchored to 76 pseudo-chromosomes based on synteny to the PNT2T reference genome using RagTag (v2.1.0) (Alonge *et al.*, 2022). Chromosome-scale scaffolding was further refined using the Hi-C data with Juicer (v1.6) (Durand *et al.*, 2016) and 3D-DNA (v190716) (Dudchenko *et al.*, 2017), followed by manual curation of the Hi-C contact maps in Juicebox (v2.17.00) (Durand et al., 2016). Remaining gaps were manually filled by aligning HiFi and ONT reads with Minimap2 (v2.24-r1122) (Li, 2018) and inspecting alignments in the Integrative Genomics Viewer (IGV) (v2.13.1) (Robinson *et al.*, 2023). The final assembly quality was assessed with Seqkit(v2.2.0) (Shen *et al.*, 2016), and gene

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

completeness was evaluated with BUSCO (v5.3.0) (Simao *et al.*, 2015)against the embryophyta_odb10 database. “hap 1–4” are arbitrary technical assemblies used for quality assessment (like BUSCO) and do not represent biologically linked whole-genome haplotypes.

TEs were annotated de novo using a combination of Repeat Modeler/RepeatMasker(Flynn *et al.*, 2020) and Extensive de-novo TE Annotator (EDTA) (v1.9.6)(Ou *et al.*, 2019), with LTR elements further refined using LTR_retriever (v2.9.0)(Ou & Jiang, 2018). Centromeric regions were identified by locating dense arrays of a 107 bp tandem repeat with TRF (v4.09) (Benson, 1999), and telomeric repeats ((TTAGGG)n) were identified with the Telomere Identification Toolkit (tidk) (v0.2.0) (Brown *et al.*, 2025). Protein-coding genes were predicted using a combination of ab initio, homology-based, and transcriptome-assisted approaches.

Predicted protein-coding genes were functionally annotated using InterProScan (v5.48-83.0) (Blum *et al.*, 2025) to assign GO terms, InterPro domains, and Pfam families. Additionally, protein sequences were searched against the UniProt Swiss-Prot database, and functional descriptions were transferred from the best-scoring hits.

Evolutionary genomics and ancestry analysis

To assess genomic diversity and identify structural variants, we performed whole-genome alignments between the four Kyoho haplotypes and the PNT2T reference using Nucmer (v4.0.0) (Marçais *et al.*, 2018) (--maxmatch -c 100 -b 500 -l 50). SVs, including inversions, translocations, and duplications, were identified from the filtered alignments using SyRI (v1.6.4) (Goel *et al.*, 2019, Goel & Schneeberger, 2022) and visualized with plotsr (v1.0.1) (Goel & Schneeberger, 2022). SNPs, Indels, and large-effect SVs (>50 bp) were extracted from the SyRI output for each haplotype, and shared versus private variants were identified using bcftools.

To determine the ancestral origin of the Kyoho genome, a phylogenomic analysis was conducted. Single-copy orthologs were identified across 31 *Vitis* genomes (Supplementary Table S20) using OrthoFinder (v2.5.5) (Emms & Kelly, 2019). Conserved blocks from single-copy gene alignments were extracted with Gblocks (v0.91b) (Castresana, 2000), and the optimal amino acid substitution model (JTT+I+G+F) was determined using ProtTest3 (v3.4.2) (Darriba *et al.*, 2011). Phylogenetic trees were constructed with IQ-TREE (v8.2.12) to infer the ancestral contributions of *V. vinifera* and *V. labrusca* to the Kyoho genome. To further resolve genome-scale relationships, we

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

performed whole-genome alignments of each component pair in VL/PNT2Tusing nucmer and calculated genetic distances with distmat (http://www.bioinformatics.nl/cgi-bin/emboss/distmat). For the Kyohogenome, we then applied Jukes–Cantor correction to non-overlapping 50 kb windows.

Transcriptomic analyses

To quantify allelic-specific expression and analyze homologous gene expression patterns, raw Illumina RNA-seq reads were filtered for quality using fastp (v0.23.2) (Chen *et al.*, 2018). Clean reads were aligned to each of the four haplotype-resolved genome assemblies separately using HISAT2 (v2.2.1)(Kim *et al.*, 2015), and gene-level read counts were quantified with featureCounts (v2.0.8) (Liao *et al.*, 2014).

We defined homologous relationships at two levels: four-way tetrads and pairwise ancestral homologs. First, to define homologous tetrads (1:1:1:1), an all-vs-all BLASTP search (*E*-value < $1e^{-5}$) was performed across the four phased copies. Sets of four genes exhibiting reciprocal best-hit relationships across all pairwise comparisons were retained. For expression motif analysis within tetrads, TPM values were normalized to the total expression of the tetrad following established methods (Ramirez-Gonzalez *et al.*, 2018), and each tetrad was assigned to one of 15 ideal expression categories based on minimal Euclidean distance.

Second, to evaluate divergence between ancestral lineages, we identified homologous gene pairs between *V. labrusca* and *V. vinifera*. A BLASTP search was performed using the *V. labrusca*-derived proteins as the query and *V. vinifera*-derived proteins as the subject. High-confidence pairs were filtered using stringent criteria: *E*-value $<1\times10^{-10}$), identity >95%, and alignment coverage ≈ 100%.

To identify differentially and tissue-specifically expressed genes, gene counts were analyzed with DESeq2 (v1.38.3) (Love *et al.*, 2014). Genes were classified as upregulated (U) or downregulated (D) if they met the significance thresholds (adjusted P <0.05 and $|\log_2\text{FoldChange}| > 1$; otherwise, they were classified as non-significant (N).

We systematically compared regulatory modes between ancestral homologs by combining their individual differential states across pairwise tissue comparisons. This resulted in eight regulatory modes: coordinated (U-U, D-D), divergent (U-D, D-U), and unilateral (U-N, N-U, D-N, N-D).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Pairs classifying as stable in both (N-N) were excluded from the divergence analysis. Tissue-specificity was further assessed using the TAU algorithm, with genes exhibiting a tissue-specificity index (TSI) ≥0.85 defined as tissue-specific.

To assess allele-specific expression bias between *V. labrusca* (C_{Lab}) and *V. vinifera* (C_{Vin}) homologs within each tissue, we applied a strict joint-threshold strategy. Homologous pairs with low abundance (average TPM ≤0.5 in both lineages) were labeled "BothLow" and excluded. For expressed pairs, bias was assessed using a binomial test under the null hypothesis $P = 0.5$ (where $n = C_{Lab} + C_{Vin}$), with P-values corrected using the Benjamini-Hochberg (FDR) procedure.

Significant expression bias was declared only if a pair met two criteria: (1) a False Discovery Rate (FDR) ≤0.01; and (2) an absolute fold-change magnitude $| \log2 C_{Lab} / C_{Vin}) | > 1$ for both raw counts and TPM values (calculated with a pseudocount to smooth variance). The direction of bias was determined by the sign of the log fold-change. Pairs not meeting these criteria were classified as "Balanced." Finally, the stability of this bias was categorized across the four tissues into four classes: Constant bias (consistent direction across all significant contexts), Semi-constant bias (consistent direction in some contexts, mixing with Balanced/BothLow), contrary bias (direction shifts between contexts), and No bias.

Epigenomic analyses

To analyze DNA methylation dynamics, raw WGBS reads were quality-trimmed using Trimmomatic (v0.39) (Bolger *et al.*, 2014). Clean reads were aligned to the four haplotype genomes using Bismark (v0.23.1) (Krueger & Andrews, 2011) with the Bowtie2 (v2.1.0) (Langmead & Salzberg, 2012) aligner. PCR duplicates were removed with deduplicate_bismark, and methylation calls were extracted using bismark_methylation_extractor. Only cytosine sites with a minimum coverage of 2x were retained for analysis. Methylation levels were calculated for CG, CHG, and CHH contexts. The distribution of methylation across genomic features (genes, TEs, promoters) and 500 kb chromosomal bins was analyzed, and metaplots were generated using deepTools (v3.5.6) (Ramirez *et al.*, 2014).

To systematically compare gene body methylation (gbM) levels between the *V. labrusca* and *V. vinifera* ancestral lineages across tissues, we harmonized the methylation records for the four haplotypes (hap1-hap4) and mapped them to the identified homologous gene pairs of *V. labrusca*

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

and *V. vinifera* ancestries. The analysis encompassed four conditions (leaf and three fruit developmental stages), each with two biological replicates. We retained homologous pairs that possessed valid methylation data for both lineages in at least one replicate, calculating the mean methylation level across replicates to obtain paired within-tissue observations. Differences in gbM between *V. labrusca* and *V. vinifera* homologs were assessed using a two-sided paired t-test, with significance stratified at $P < 0.05$. Finally, to characterize the spatiotemporal dynamics and stability of the epigenetic landscape, we categorized genes into a three-state model: Body Methylated (BM), Intermediately Methylated (IM), and Unmethylated (UM). This analysis highlighted the regulatory divergence between the *V. labrusca* and *V. vinifera* genomic components.

Alternative splicing analyses

To create a comprehensive reference transcriptome for splicing analysis, raw PacBio Iso-Seq subreads were processed into circular consensus sequences (CCSs) using isoseq3 (v4.0.0) (https://github.com/PacificBiosciences/IsoSeq/blob/master/isoseq-clustering.md). Full-length non-chimeric (FLNC) reads were identified, clustered, and polished to generate high-quality isoforms (accuracy ≥ 99%). These isoforms were further error-corrected using Illumina short-read data with LoRDEC (v0.9) (Salmela & Rivals, 2014). Corrected, non-redundant transcripts were mapped to the Kyoho genome using pbmm2 (v1.17.0) (https://github.com/PacificBiosciences/pbmm2) and collapsed using isoseq collapse to remove 5' end degradation artifacts. A comprehensive transcriptome annotation was generated by merging the Iso-Seq transcripts with the short-read based annotations using gffcompare (v0.11.2) (Pertea & Pertea, 2020).

To identify and quantify DAS, the SUPPA2 pipeline was used. This final annotation served as the reference for AS analysis with SUPPA2 (v2.3) (Alamancos *et al.*, 2015). Transcript abundances were quantified using Salmon (v1.10.3) (Patro *et al.*, 2017), and the percent spliced-in (PSI) value for each AS event was calculated. Differential alternative splicing (DAS) events were identified between conditions using a threshold of $|dPSI| > 0.1$ and an FDR < 0.05. To evaluate the relationship between splicing and transcription, DAS results were integrated with differential gene expression data using normalized gene identifiers. For intersecting genes, global associations were estimated using Pearson and Spearman correlations. Directional consistency was assessed by categorizing

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

genes into quadrants based on the signs of dPSI and log2FoldChange (Q1/Q3: consistent; Q2/Q4: inconsistent).

For orthologous analysis, we adopted a one-to-one *V. vinifera* and *V. labrusca* homology framework. Standardized DAS events were compared across three intra-tissue and three inter-tissue conditions. Pairwise event type combinations flanking homologous gene pairs were enumerated to characterize the total load and compositional spectrum. Splicing conservation was assessed by correlating dPSI values and quantifying effect size differences $dPSI = dPSI_{vin} - dPSI_{lab}$. Using a threshold of $|dPSI| \geq 0.1$, pairwise regulatory patterns were defined as identical, opposite, or unilateral to examine regulatory symmetry.

Finally, structure characterization was analyzed for RI, A5, A3, AF, AL, and SE event. Event lengths were defined as the distance between endpoints (for RI, SE) or alternative splice sites (for A3, A5, AF, AL). Length distributions were modeled using kernel density estimation. We reported median, and interquartile range (IQR), and 50% Highest Density Intervals (HDIs) on both linear and log1p scales to accommodate heavy-tailed distributions.

Statistics and reproducibility

Statistical comparisons of continuous variables between groups were performed using the two-sided Wilcoxon signed-rank test, and differences in proportions were assessed using the two-sided Pearson chi-square test. All statistical analyses were conducted in R (v4.5).

Accession numbers

Iso-Seq, BS-seq and RNA-seq data have been deposited in the NCBI Sequence Read Archive under project number PRJNA1372951 and the National Genomics Data Center (NGDC) Genome Sequence Archive (GSA) (https://ngdc.cncb.ac.cn/gsa/), with BioProject number PRJCA052795. The haplotype-resolved genome assemblies of Kyoho have been deposited in the National Center for Biotechnology Information (NCBI) under the BioProject accession PRJNA1372984.

Acknowledgements

We thank members of the Zhou lab at AGIS and CATAS for discussion and comments on the project.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Author contributions

Y.Z., W.L. and C.L. conceived and designed the study. L.S. collected the plant materials. Yingchun Z. and X.W. assembled and annotated the haplotype-resolved genomes. Yingchun Z., Y.X. and C.C. performed grape photography. Yingchun Z., Z.J., X.F. and Q.F. analyzed the evolutionary genomic origins. Yingchun Z., J.Y., Q.L. and T.Z. conducted transcriptome analysis. Yingchun Z., Y.S., S.Y., Z.A. and R.Y. performed DNA methylation analysis. Y.Z., Sheng Y., Y.D., L.Z., and L.T. conducted Iso-seq transcriptome analysis. Yingchun Z. and Y.P. wrote the manuscript. All authors provided critical feedback and approved the final manuscript.

Funding

This work was supported by the SZPU Research Project-6024330001K, the Science Fund Program for Excellent Young Scholars of the National Natural Science Foundation of China (Overseas) to Yongfeng Zhou, and the National Key Research and Development Program of China (No. 2023YFF1000100; 2023YFD2200700), the Project of National Key Laboratory of Tropical Crop Breeding (No. NKLTCB-RC202501).

Data availability

The haplotype-resolved genome assemblies of Kyoho have been deposited in Zendo (https://doi.org/10.5281/zenodo.18047369). The complete genome assembly and gap-filling workflow, along with sequencing data, are available on our lab's GitHub repository at GitHub@zhouyflab.

Competing interests

The authors declare no competing interests.

References

Alamancos, G. P., Pages, A., Trincado, J. L., Bellora, N. and Eyras, E. (2015) Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA,* **21,** 1521-1531.

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., *et al.* (2022) Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol,* **23,** 258.

Alyahya, N. and Taybi, T. (2024) Transcriptome-wide characterization of alternative splicing regulation in Najran wheat under salt stress. *Curr Plant Biol,* **38**.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Baralle, F. E. and Giudice, J. (2017) Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol,* **18,** 437-451.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res,* **27,** 573-580.

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., *et al.* (2019) The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nat Genet,* **51,** 877-884.

Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., *et al.* (2025) InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res,* **53,** D444-D456.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30,** 2114-2120.

Brinton, J., Ramirez-Gonzalez, R. H., Simmonds, J., Wingen, L., Orford, S., Griffiths, S., *et al.* (2020) A haplotype-led approach to increase the precision of wheat breeding. *Commun Biol,* **3,** 712.

Brown, M. R., Manuel Gonzalez de La Rosa, P. and Blaxter, M. (2025) tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics,* **41,** btaf049.

Cadle-Davidson, L. (2008) Variation Within and Between spp. for Foliar Resistance to the Downy Mildew Pathogen. *Plant Dis,* **92,** 1577-1584.

Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution,* **17,** 540-552.

Chalhoub, B., Denoeud, F., Liu, S. Y., Parkin, I. A. P., Tang, H. B., Wang, X. Y., *et al.* (2014) Early allopolyploid evolution in the post-Neolithic oilseed genome. *Science,* **345,** 950-953.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics,* **34,** i884-i890.

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. and Li, H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods,* **18,** 170-175.

Cheng, L., Bao, Z., Kong, Q., Lassois, L., Stein, N., Huang, S., *et al.* (2025) Genome analyses and breeding of polyploid crops. *Nat Plants,* **11,** 1714-1728.

Darriba, D., Taboada, G. L., Doallo, R. and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics,* **27,** 1164-1165.

de Jong, G. W. and Adams, K. L. (2023) Subgenome-dominant expression and alternative splicing in response to Sclerotinia infection in polyploid Brassica napus and progenitors. *Plant J,* **114,** 142-158.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., *et al.* (2017) De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science,* **356,** 92-95.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., *et al.* (2016) Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst,* **3,** 95-98.

Emms, D. M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol,* **20**.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., *et al.* (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A,* **117,** 9451-9457.

Gao, P., Quilichini, T. D., Zhai, C., Qin, L., Nilsen, K. T., Li, Q., *et al.* (2021) Alternative splicing dynamics and evolutionary divergence during embryogenesis in wheat species. *Plant Biotechnol J,* **19,** 1624-1643.

Goel, M. and Schneeberger, K. (2022) plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics,* **38,** 2922-2926.

Goel, M., Sun, H., Jiao, W. B. and Schneeberger, K. (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol,* **20,** 277.

Guo, S. H., Yang, B. H., Wang, X. W., Li, J. N., Li, S., Yang, X., *et al.* (2021) ABA signaling plays a key role in regulated deficit irrigation-driven anthocyanins accumulation in 'Cabernet Sauvignon' grape berries. *Environ Exp Bot,* **181**.

Hirschberg, J. (2001) Carotenoid biosynthesis in flowering plants. *Curr Opin Plant Biol,* **4,** 210-218.

Hu, Y., Chen, J. D., Fang, L., Zhang, Z. Y., Ma, W., Niu, Y. C., *et al.* (2019) Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature Genetics,* **51,** 739-+.

Huang, H., Liu, R., Niu, Q., Tang, K., Zhang, B., Zhang, H., *et al.* (2019) Global increase in DNA methylation during orange fruit development and ripening. *Proc Natl Acad Sci U S A,* **116,** 1430-1436.

Jiang, X., Song, Q., Ye, W. and Chen, Z. J. (2021) Concerted genomic and epigenomic changes accompany stabilization of Arabidopsis allopolyploids. *Nat Ecol Evol,* **5,** 1382-1393.

Kim, D., Langmead, B. and Salzberg, S. L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods,* **12,** 357-360.

Krueger, F. and Andrews, S. R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics,* **27,** 1571-1572.

Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods,* **9,** 357-359.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics,* **34,** 3094-3100.

Li, M., Ou, M., He, X., Ye, H., Ma, J., Liu, H., *et al.* (2023) DNA methylation role in subgenome expression dominance of Juglans regia and its wild relative J. mandshurica. *Plant Physiol,* **193,** 1313-1329.

Liao, Y., Smyth, G. K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* **30,** 923-930.

Liu, Z. J., Wang, N., Su, Y., Long, Q. M., Peng, Y. L., Shangguan, L. F., *et al.* (2024) Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics*.

Love, M. I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* **15,** 550.

Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics,* **154,** 459-473.

Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L. and Zimin, A. (2018) MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol,* **14,** e1005944.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L. and Zimin, A. (2018) MUMmer4: A fast and versatile genome alignment system. *Plos Computational Biology,* **14**.

Matzke, M. A. and Mosher, R. A. (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics,* **15,** 394-408.

Mhiri, C., Parisod, C., Daniel, J., Petit, M., Lim, K. Y., Dorlhac de Borne, F., *et al.* (2019) Parental transposable element loads influence their dynamics in young Nicotiana hybrids and allotetraploids. *New Phytol,* **221,** 1619-1633.

OIV (2017) Distribution of the World's Grapevine Varieties (International Organisation of Vine and Wine).

Ou, S. and Jiang, N. (2018) LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol,* **176,** 1410-1422.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., *et al.* (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol,* **20,** 275.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods,* **14,** 417-419.

Peng, Y., Wang, Y., Liu, Y., Fang, X., Cheng, L., Long, Q., *et al.* (2025) The genomic and epigenomic landscapes of hemizygous genes across crops with contrasting reproductive systems. *Proceedings of the National Academy of Sciences,* **122,** e2422487122.

Pertea, G. and Pertea, M. (2020) GFF Utilities: GffRead and GffCompare. *F1000Res,* **9**.

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., *et al.* (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science,* **345,** 1250091.

Ramirez-Gonzalez, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., *et al.* (2018) The transcriptional landscape of polyploid wheat. *Science,* **361**.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res,* **42,** W187-191.

Ranallo-Benavidez, T. R., Jaron, K. S. and Schatz, M. C. (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun,* **11,** 1432.

Reddy, A. S. (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol,* **58,** 267-294.

Reddy, A. S. N., Marquez, Y., Kalyna, M. and Barta, A. (2013) Complexity of the Alternative Splicing Landscape in Plants. *Plant Cell,* **25,** 3657-3683.

Robinson, J. T., Thorvaldsdottir, H., Turner, D. and Mesirov, J. P. (2023) igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics,* **39**.

Rousseau-Gueutin, M., Belser, C., Da Silva, C., Richard, G., Istace, B., Cruaud, C., *et al.* (2020) Long-read assembly of the Brassica napus reference genome Darmor-bzh. *Gigascience,* **9**.

Salmela, L. and Rivals, E. (2014) LoRDEC: accurate and efficient long read error correction. *Bioinformatics,* **30,** 3506-3514.

Shen, W., Le, S., Li, Y. and Hu, F. (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One,* **11,** e0163962.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Shi, X., Cao, S., Wang, X., Huang, S., Wang, Y., Liu, Z., *et al.* (2023) The complete reference genome for grapevine (Vitis vinifera L.) genetics and breeding. *Hortic Res,* **10,** uhad061.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics,* **31,** 3210-3212.

Slotkin, R. K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics,* **8,** 272-285.

Song, A., Su, J., Wang, H., Zhang, Z., Zhang, X., Van de Peer, Y., *et al.* (2023) Analyses of a chromosome-scale genome assembly reveal the origin and evolution of cultivated chrysanthemum. *Nat Commun,* **14,** 2021.

Song, B., Yu, J., Li, X., Li, J., Fan, J., Liu, H., *et al.* (2024a) Increased DNA methylation contributes to the early ripening of pear fruits during domestication and improvement. *Genome Biol,* **25,** 87.

Song, Y., Peng, Y., Liu, L., Li, G., Zhao, X., Wang, X., *et al.* (2024b) Phased gap-free genome assembly of octoploid cultivated strawberry illustrates the genetic and epigenetic divergence among subgenomes. *Hortic Res,* **11,** uhad252.

Srikant, T., Gonzalo, A. and Bomblies, K. (2024) Chromatin Accessibility and Gene Expression Vary Between a New and Evolved Autopolyploid of Arabidopsis arenosa. *Mol Biol Evol,* **41**.

Sun, H., Jiao, W. B., Krause, K., Campoy, J. A., Goel, M., Folz-Donahue, K., *et al.* (2022) Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat Genet,* **54,** 342-348.

Tang, D., Gallusci, P. and Lang, Z. (2020) Fruit development and epigenetic modifications. *New Phytol,* **228,** 839-844.

Todesco, M., Owens, G. L., Bercovich, N., Legare, J. S., Soudi, S., Burge, D. O., *et al.* (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature,* **584,** 602-607.

Tognacca, R. S., Rodríguez, F. S., Aballay, F. E., Cartagena, C. M., Servi, L. and Petrillo, E. (2023) Alternative splicing in plants: current knowledge and future directions for assessing the biological relevance of splice variants. *J Exp Bot,* **74,** 2251-2272.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., *et al.* (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature,* **588,** 277-283.

Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., *et al.* (2019) Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nat Genet,* **51,** 224-229.

Wang, M. and Wang, J. (2022) Transcriptome and DNA methylome analyses provide insight into the heterosis in flag leaf of inter-subspecific hybrid rice. *Plant Mol Biol,* **108,** 105-125.

Wang, T., van Dijk, A. D. J., Zhao, R., Bonnema, G. and Wang, X. (2024a) Contribution of homoeologous exchange to domestication of polyploid Brassica. *Genome Biol,* **25,** 231.

Wang, X., Liu, Z., Zhang, F., Xiao, H., Cao, S., Xue, H., *et al.* (2024b) Integrative genomics reveals the polygenic basis of seedlessness in grapevine. *Curr Biol,* **34,** 3763-3777 e3765.

Wu, S., Sun, H., Zhao, X., Hamilton, J. P., Mollinari, M., Gesteira, G. S., *et al.* (2025) Phased chromosome-level assembly provides insight into the genome architecture of hexaploid sweetpotato. *Nat Plants,* **11,** 1951-1959.

Xiao, H., Wang, Y., Liu, W., Shi, X., Huang, S., Cao, S., *et al.* (2025) Impacts of reproductive systems on grapevine genome and breeding. *Nat Commun,* **16,** 2031.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

Xue, Y. J., Shi, Y., Qi, Y. C., Yu, H., Zou, C. W., Cao, X. Y., *et al.* (2022) Epigenetic and Genetic Contribution for Expression Bias of Homologous Alleles in Polyploid Sugarcane. *Agronomy-Basel,* **12**.

Yao, S., Liang, F., Gill, R. A., Huang, J., Cheng, X., Liu, Y., *et al.* (2020) A global survey of the transcriptome of allopolyploid Brassica napus based on single-molecule long-read isoform sequencing and Illumina-based RNA sequencing data. *Plant J,* **103,** 843-857.

Yoo, M. J., Szadkowski, E. and Wendel, J. F. (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb),* **110,** 171-180.

Yuan, J., Jiao, W., Liu, Y., Ye, W., Wang, X., Liu, B., *et al.* (2020) Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. *BMC Biol,* **18,** 171.

Zhang, J., Qi, Y., Hua, X., Wang, Y., Wang, B., Qi, Y., *et al.* (2025) The highly allo-autopolyploid modern sugarcane genome and very recent allopolyploidization in Saccharum. *Nat Genet,* **57,** 242-253.

Zhang, T. Z., Hu, Y., Jiang, W. K., Fang, L., Guan, X. Y., Chen, J. D., *et al.* (2015) Sequencing of allotetraploid cotton (L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology,* **33,** 531-U252.

Zhang, Z., Xun, H., Lv, R., Gou, X., Ma, X., Li, J., *et al.* (2022) Effects of homoeologous exchange on gene expression and alternative splicing in a newly formed allotetraploid wheat. *Plant J,* **111,** 1267-1282.

Zhang, Z. B., Fu, T. S., Liu, Z. J., Wang, X. T., Xun, H. W., Li, G., *et al.* (2019) Extensive changes in gene expression and alternative splicing due to homoeologous exchange in rice segmental allopolyploids. *Theor Appl Genet,* **132,** 2295-2308.

Zhou, Y., Duvaux, L., Ren, G., Zhang, L., Savolainen, O. and Liu, J. (2017) Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity (Edinb),* **118,** 211-220.

Zhou, Y., Xiong, J., Shu, Z., Dong, C., Gu, T., Sun, P., *et al.* (2023) The telomere-to-telomere genome of Fragaria vesca reveals the genomic evolution of Fragaria and the origin of cultivated octoploid strawberry. *Hortic Res,* **10,** uhad027.

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., *et al.* (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet,* **51,** 865-876.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

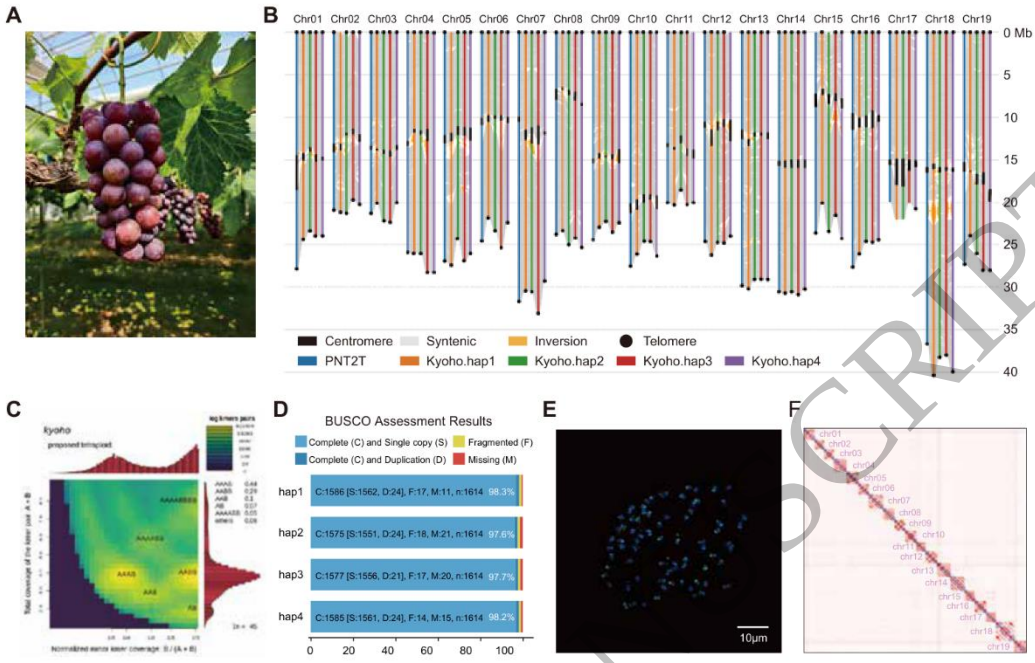


Figure 1
210x297 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

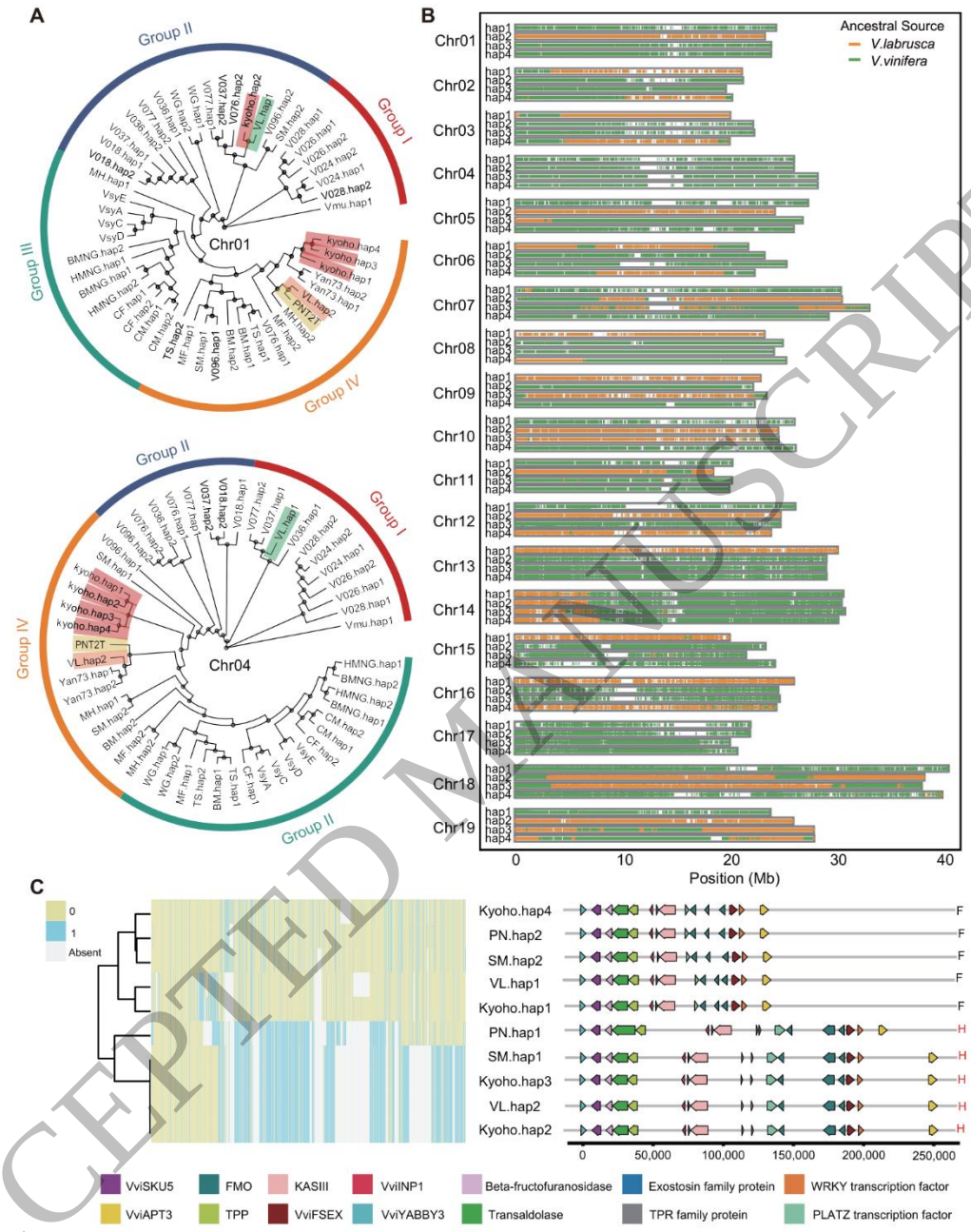


Figure 2
210x297 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

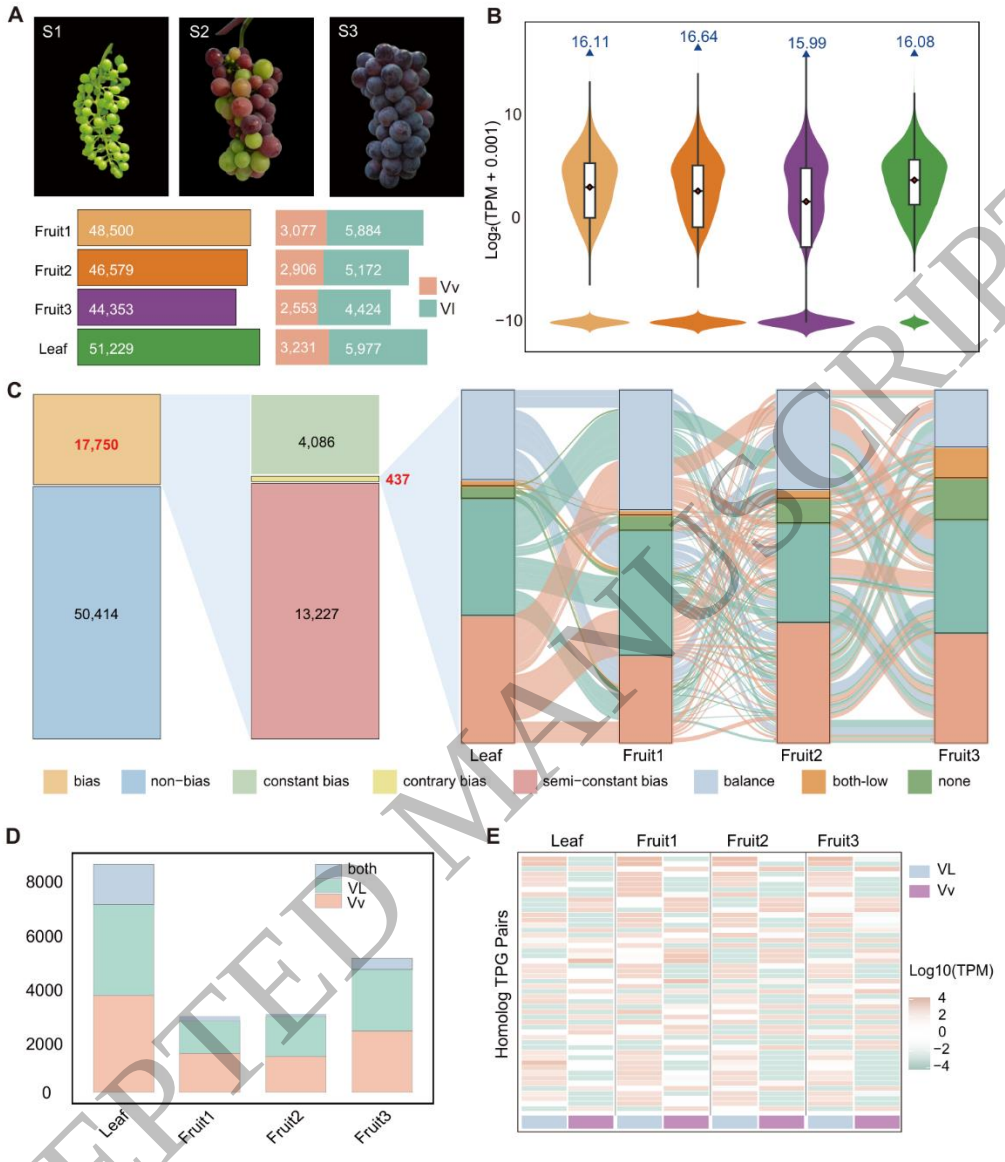


Figure 3
210x297 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

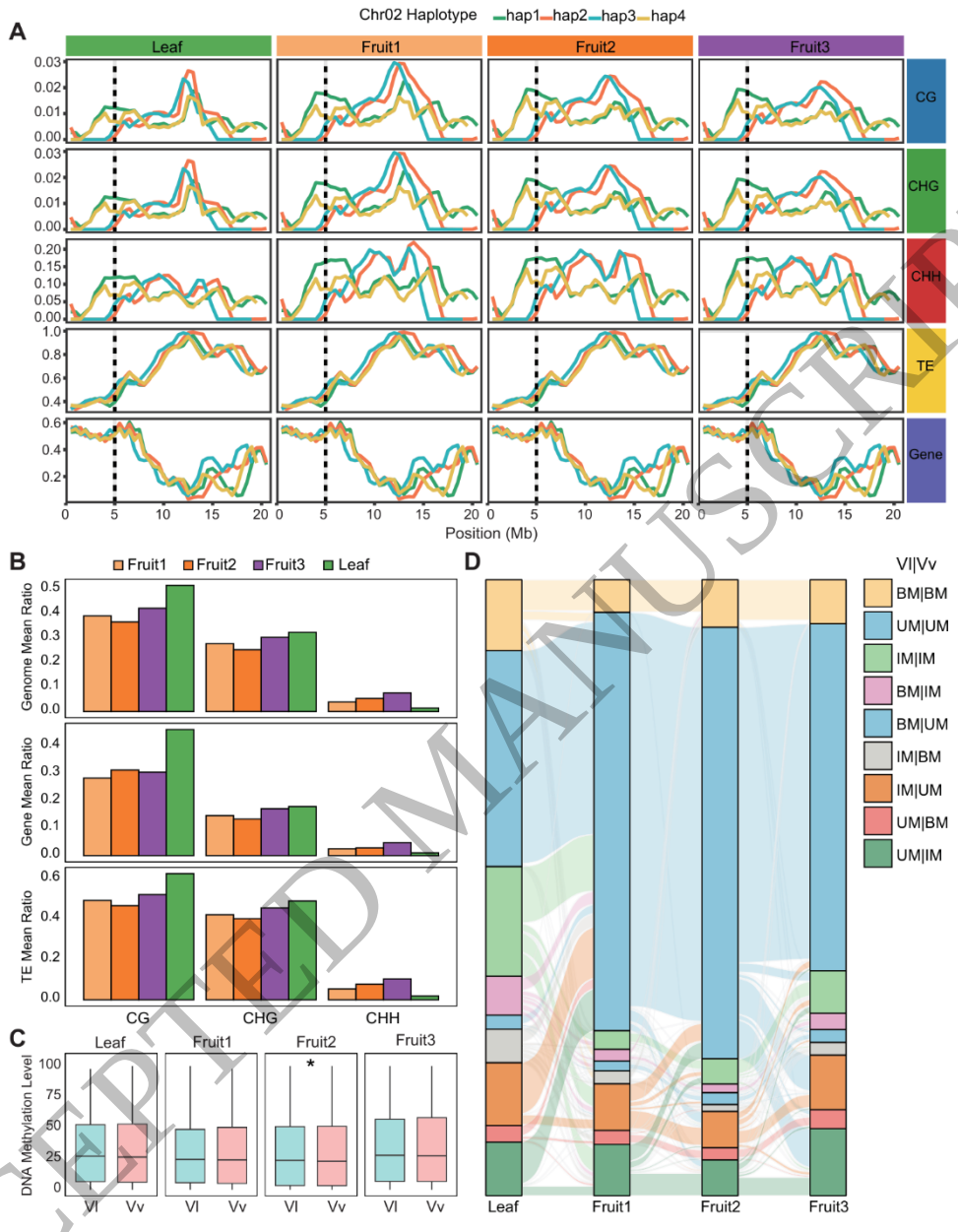


Figure 4
210x297 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

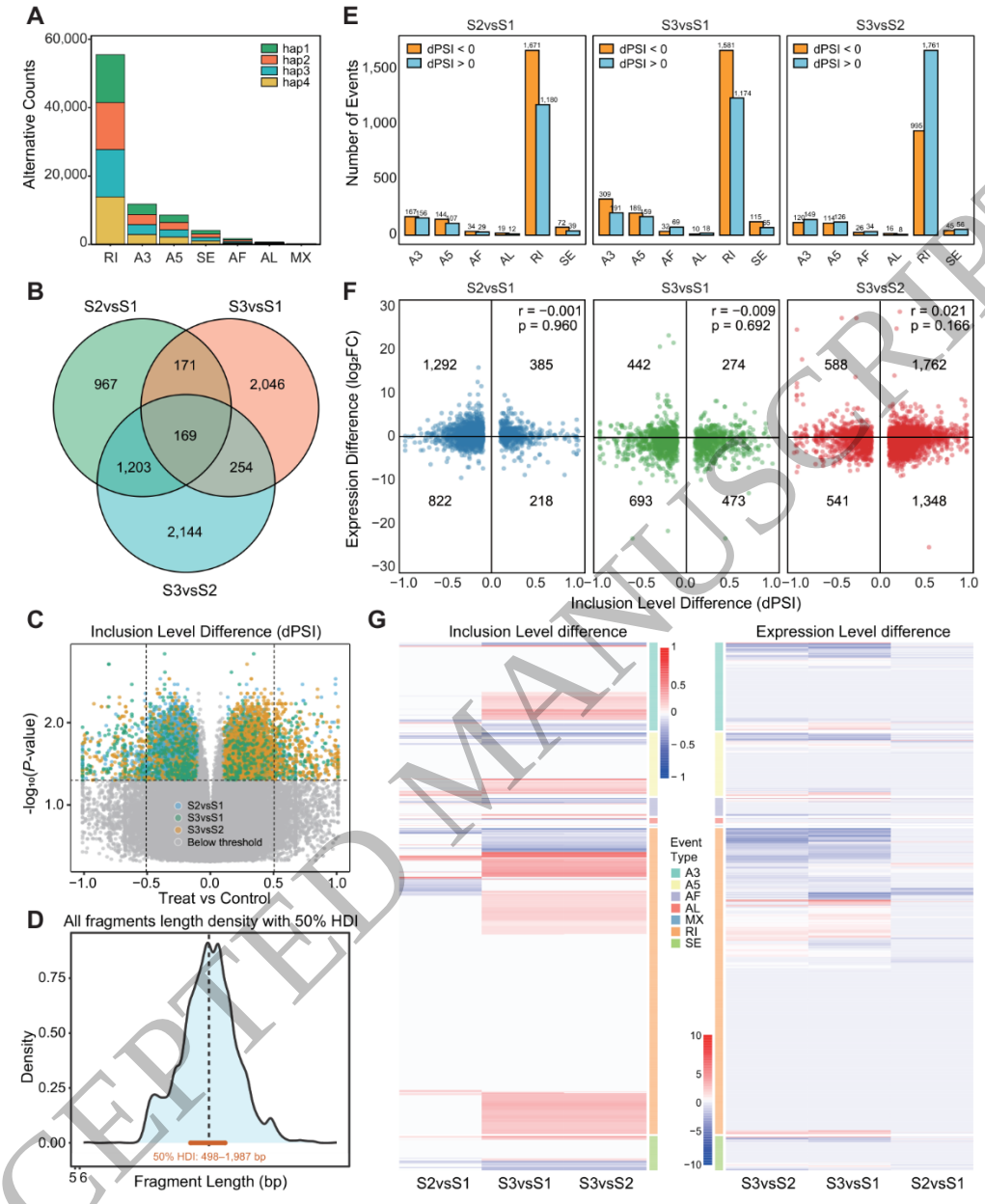


Figure 5
210x297 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026

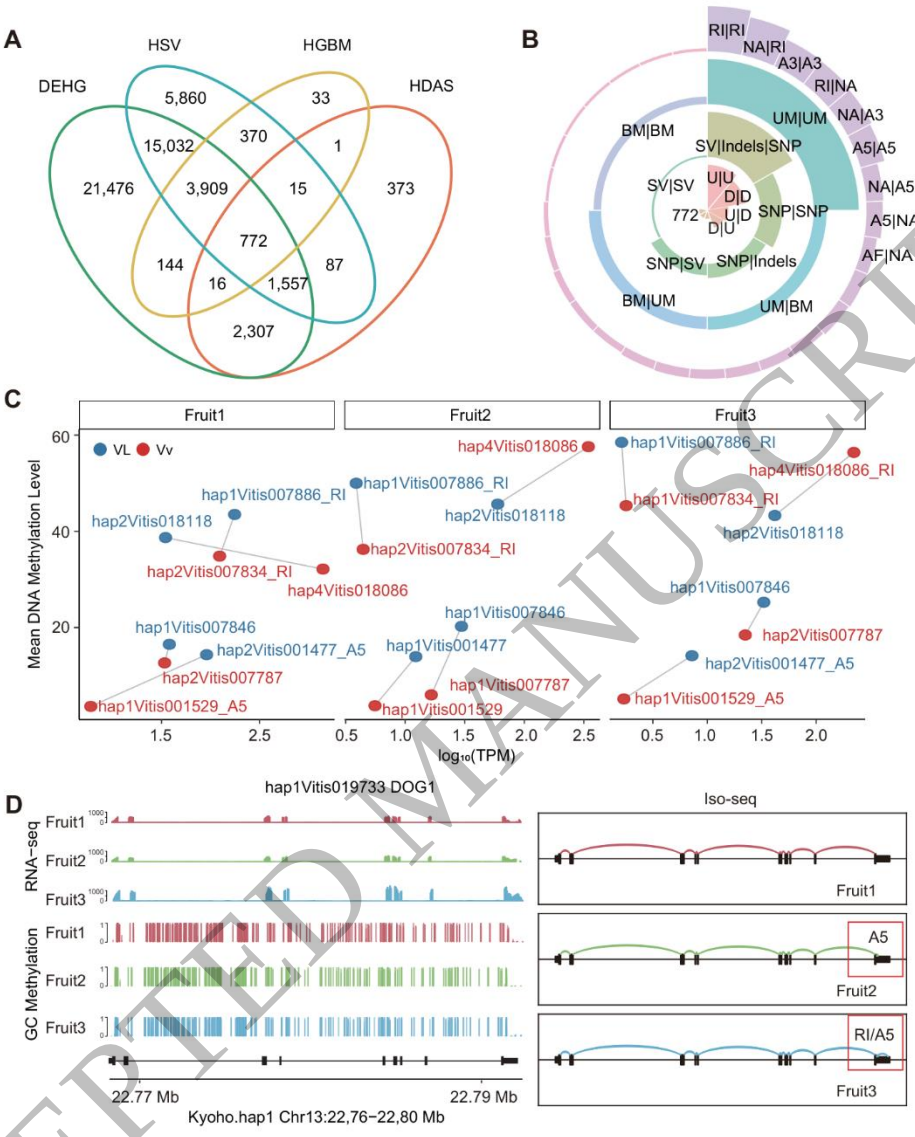


Figure 6
210x297 mm (x DPI)

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/plphys/advance-article/doi/10.1093/plphys/kiag330/8705307 by guest on 23 June 2026