the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2025) **121,** e70012 doi: 10.1111/tpj.70012

RESOURCE

Haplotype-resolved genome assembly of *Ficus carica* L. reveals allele-specific expression in the fruit

Gabriele Usai[1,†], Tommaso Giordani[1,†], Alberto Vangelisti[1], Marco Castellacci[1], Samuel Simoni[1], Emanuele Bosi[2], Lucia Natali[1], Flavia Mascagni[1,*] and Andrea Cavallini[1]

[1]*Department of Agriculture, Food and Environment, University of Pisa, Pisa, Italy, and*

[2]*Department of Earth, Environmental and Life Science, University of Genoa, Genoa, Italy*

Received 29 October 2024; revised 9 January 2025; accepted 13 January 2025.

*For correspondence (e-mail flavia.mascagni@unipi.it).

[†]These authors contributed equally to this study.

SUMMARY

In this study, we produced a haplotype-phased genome sequence of the fig tree (*Ficus carica* L.), a non-Rosaceae fruit tree model species, providing a systematic overview of the organization of a heterozygous diploid genome and, for the first time in a fruit tree, evidence of allelic expression direction-shifting among haplotypes. The genome was used for whole genome analysis of heterozygosis, allelic cytosine methylation, and expression profiles in peel and pulp fruit tissues. The two pseudo-haplotypes spanned approximately 355 and 346 Mbp, respectively, and 97% of the sequences were associated with 13 chromosome pairs of the fig tree. Overall, the methylation profile in peel and pulp tissues showed no variations between the homologous chromosomes. However, we detected differential DNA methylation within defined heterozygous allelic gene regions, particularly in upstream regions. Among 6768 heterozygous coding sequences identified, 4024 exhibited allele-specific expression, with approximately 18% specific to the peel and 14% to the pulp. Specifically, 2715 genes were consistent, with one allele always more expressed than the other in both peel and pulp. On the contrary, 22 allele-specific expressed genes switched allele expression among the fig fruit peel and pulp tissues, indicating evidence of overdominance and suggesting that the genome can express one of the two alleles higher or lower depending on developmental or environmental triggers. Notably, these genes were associated with key biological processes, including fruit maturation regulation, seed maturation, and stress response, highlighting their potential role in the plant's developmental and adaptive functions in view of gene editing-based breeding.

Keywords: haplotype-resolved genome, allele-specific expression, differentially methylated regions, heterosis, *Ficus carica* L..

INTRODUCTION

The appreciable variability of plant phenotypic traits arises partly from DNA sequence variations. However, genetic variability among individuals can be determined, besides structural sequence heterozygosity, by allele-specific gene expression and/or by DNA modifications that can alter the genome's activity without changing its sequence (*i.e.*, epigenetic variations).

Cytosine methylation is the most well-studied epigenetic mechanism in mammals and plants: besides contributing to gene regulation, it is involved in transposable element (TE) silencing, imprinting, and X-chromosome inactivation (Law & Jacobsen, 2010). Unlike animals, where methylation is usually limited to the CG context, plants exhibit robust methylation in three possible contexts: CG, CHG, and CHH (where H is any base but G) (Law & Jacobsen, 2010). Different biochemical pathways produce methylation marks in these three contexts, can vary in frequency along the genome, and show different patterns of inheritance. For instance, in Arabidopsis, CGs are the most methylated sequence context (~24% of all CGs), followed by CHG (~7%) and CHH (~1.7%) (Cokus et al., 2008). The insights gained from Arabidopsis research have provided a foundation for exploring methylation impact in a broad spectrum of plant species. However, few species have undergone comprehensive DNA methylation studies

(Ji et al., 2015). Moreover, the majority of investigations into plant DNA methylomes have relied on haplotype-collapsed genomes (Li et al., 2023; López et al., 2022; Wang et al., 2015; Wang et al., 2021), thus overlooking many methylome features specific to a single haplotype, such as the occurrence of epigenetic alleles with identical DNA sequences but distinct heritable differences in DNA methylation and gene expression.

Sequencing plant genomes has dramatically advanced our understanding of plant development, evolution, and the formation of biological traits (Hamilton & Robin Buell, 2012). However, the construction of haplotype-phased plant genomes has always been challenging due to heterozygosity, polyploidy, and high repeat content (Michael & VanBuren, 2020).

The new possibility to perform genome-wide identification of the genomic variations between the two pseudo-haplotypes of the same plant provides additional insights into structural variations (SVs) and imbalanced allelic expression, indicated as allele-specific expression (ASE), between haplotypes (Guk et al., 2022).

ASE may play a crucial role in genetics by contributing to heterosis, the phenomenon implying that heterozygous hybrids show superior performance over homozygous parents. The genetic basis of heterosis can be explained by dominant and overdominant effects, significantly improving adaptability (Liu et al., 2020; Shao et al., 2019; Zhang et al., 2021). Specifically, the first effect occurs when dominant alleles in heterozygous hybrids enhance growth and development by masking the influence of less effective or recessive alleles, while the overdominance effect emerges when a heterozygous allele combination outperforms homozygous alleles.

Heterosis is exploited in F1 hybrids or in clonally propagated crops that are typically heterozygous. Consistent ASE genes that favor one haplotype allele across various tissues and stages are considered to drive the dominant effect. Meanwhile, inconsistent ASEs, which may shift expression direction between tissues or developmental stages, are associated with the overdominant effect (Zhang et al., 2021).

The fig tree (*Ficus carica* L.) is an important commercial species of the genus *Ficus* of the Moraceae family. It has been one of the earliest domesticated species, cultivated for over 12 000 years, and is believed to have originated in Southwest Asia and the Middle East (Kislev et al., 2006; Simsek et al., 2020). *F. carica* is known for its adaptability to a wide range of soils, especially for its tolerance to harsh environmental conditions and poor soils (Vangelisti et al., 2019). In recent years, there has been a growing interest in enhancing the quality and flavor of fig fruits and extending their storability, given that they are highly perishable (Allegra et al., 2017, 2018).

The fig tree is a low/medium heterozygous species with a relatively small genome (~356 Mbp; Loureiro et al., 2007) organized in 13 chromosomes (2n = 26). The first reported reference genome for *F. carica* is that of the Japanese cultivar "Horaishi" (Mori et al., 2017). In 2020, a full haplotype-phased assembly of the fig tree (cultivar "Dottato") genome based on long-read sequencing was released by using the Pacific Biosciences (PacBio) Single-Molecule, Real-Time (SMRT) sequencing technology (Usai et al., 2020). In 2023, another high-quality genome assembly of *F. carica* cultivar "Orphan" was released with transcriptomic and metabolomic resources, further enriching the fig tree genomics (Bao et al., 2023). Lastly, in 2024, an updated version of the cultivar "Horaishi" redefined the fig sex-linked loci through genome-wide association studies (Ikegami et al., 2024). These studies and other transcriptomic and metabolomic analyses (Mascellani et al., 2021; Vangelisti et al., 2019; Vangelisti et al., 2021) indicate the fig tree as a promising model species for non-Rosaceae fruit trees.

Here, we present an updated version of the fully haplotype-phased *F. carica* cultivar "Dottato" assembly. We integrated our previous high-coverage PacBio data with Hi-C, methylome, and transcriptome sequencing. The combination of Hi-C chromosome conformation capture methodology with long-read data allows for improved diploid genome assemblies without parental data, facilitating the study of genomic variation between pseudo-haplotypes (*i.e.*, chromosome-scale phased blocks interspersed with homologous collapsed regions depending on heterozygosity) within an individual's genome (Kronenberg et al., 2021). These methodologies enabled us to decipher genome-wide variations by comparing SVs between pseudo-haplotypes and investigating the pattern of allele-specific events, particularly highlighting the role of differential allelic methylation and expression in the fig tree genome.

RESULTS

Genome sequencing, assembly, and curation

The genomic DNA of the sequenced *F. carica* female plant (cultivar "Dottato"; Usai et al., 2020) was used for the construction and sequencing of the Hi-C library. After quality control, 65 935 417 pairs of 150 bp-long Hi-C reads (~55x genome coverage) were retrieved.

The published assembly generated with PacBio data, composed of primary contigs and haplotigs, was used altogether with Hi-C data to extend the phasing between the unzipped haplotig blocks, resulting in the generation of two new and better-phased assemblies named phase 0 and phase 1. The two phases were validated by mapping back the PacBio dataset, obtaining an overlap between the

reads and the two assemblies of about 98% and 97%, respectively. The produced and polished sequences were manually corrected in case of positioning errors, thus generating the two final sets of haplotype-phased sequences, *i.e.*, pseudo-haplotype 0 and pseudo-haplotype 1 (Table S1; Figure S1).

Overall, we produced 355 244 677 bp for pseudo-haplotype 0 and 346 221 631 bp for pseudo-haplotype 1, respectively (Figure 1a). Loureiro et al. (2007) estimated the fig genome size by flow cytometry at about 356 Mbp; thus, the produced pseudo-haplotype 0 and pseudo-haplotype 1 should represent about 99% and 97% of the fig genome, respectively. Both pseudo-haplotypes were composed of 538 sequences. The pseudo-haplotype 0 had a mean sequence size of 660.306 bp with an N50 of 1 989 800 bp, while pseudo-haplotype 1 had a mean sequence size of 643.535 bp with an N50 of 1 927 249 bp (Figure 1a).

Finally, 400 of the 538 homologous sequences of the two pseudo-haplotypes were anchored to the 13 chromosomes of the fig tree, generating 13 pseudo-chromosome sequence pairs (Figure 1a,b; Figure S2). Those sequences accounted for 346 881 609 bp for pseudo-haplotype 0 (~97% of pseudo-haplotype 0) and 338 526 026 bp for pseudo-haplotype 1 (~95% of pseudo-haplotype 1), respectively. The anchoring process was performed according to the physical data of the cultivar "Horaishi" assembly (Mori et al., 2017), allowing us to produce a genetic-to-physical map for both pseudo-haplotypes (Table S2). Due to the Hi-C-mediated separation of the two pseudo-haplotypes, the new assembly represents a significant improvement over the last assemblies of the fig genome generated using third-generation sequencing methods (Bao et al., 2023; Ikegami et al., 2024).

BUSCO core gene evaluation (Simão et al., 2015) was used to assess the core gene completeness of the two pseudo-haplotypes. The program recovered 1510 (~93%) and 1497 (~92%) of the 1614 highly conserved Embryophyta core genes in pseudo-haplotype 0 and pseudo-haplotype 1, respectively (Table S3). Furthermore, the continuity of the two pseudo-haplotypes was assessed using the genome metric defined as the LTR assembly index (LAI). The analysis resulted in a LAI assembly score of 10.94 and 13.85 for pseudo-haplotype 0 and pseudo-haplotype 1, respectively. According to the proposed classification, these scores referred to a "reference" quality level (Ou et al., 2018).

Genome annotation

To produce an exhaustive annotation of the TEs, the two pseudo-haplotypes were scanned with *ad hoc* predictive tools. The higher contiguity of the new assembly of fig cultivar "Dottato" allowed us to better resolve the repetitive component of the genome in comparison with the previous versions (Bao et al., 2023; Ikegami et al., 2024; Mori et al., 2017; Usai et al., 2020). TEs accounted for about 48% of pseudo-haplotype 0 and pseudo-haplotype 1 (Figure 1a)

(a)

Summary of fig genome assembly and annotation statistics			
Chromosome number (2n)	**2n = 2x = 26**		
Estimated genome size (1n)	**~356 Mbp**		
		Pseudo-haplotype 0	**Pseudo-haplotype 1**
Contig assembly	Genome representation (%)	~99	~97
	Total size of the assembly (bp)	355,244,677	346,221,631
	Number of sequences (No.)	538	538
	Mean sequence size (bp)	660,306	643,53
	N50 sequence length (bp)	1,989,800	1,927,249
Anchored assembly	Genome representation (%)	~97	~95
	Total size of the assembly (bp)	346,881,609	338,526,026
	Number of sequences (No.)	13	13
	Mean sequence size (bp)	26,683,201	26,040,464
	N50 sequence length (bp)	27,941,851	27,454,058
Annotation	BUSCO assessment (%)	93.6	92.7
	Protein-coding genes proportion (%)	30.02	30.01
	Predicted protein-coding genes (No.)	32,985	32,682
	Annotated protein-coding genes (No.)	27,916	27,558
	Rate of annotated protein-coding genes (%)	82.22	82.56
	Average protein-coding genes length (bp)	3,141	3,113
	Average CDS length (bp)	989.2	991.25
	Average exon length (bp)	254.92	254.4
	Average intron length (bp)	554.41	542.66
	Average exon per gene (No.)	4.57	4.59
	Average intron per gene (No.)	3.57	3.59
	LAI assessment (No.)	10.94	13.85
	Transposable elements proportion (%)	48.68	48.59

(b)

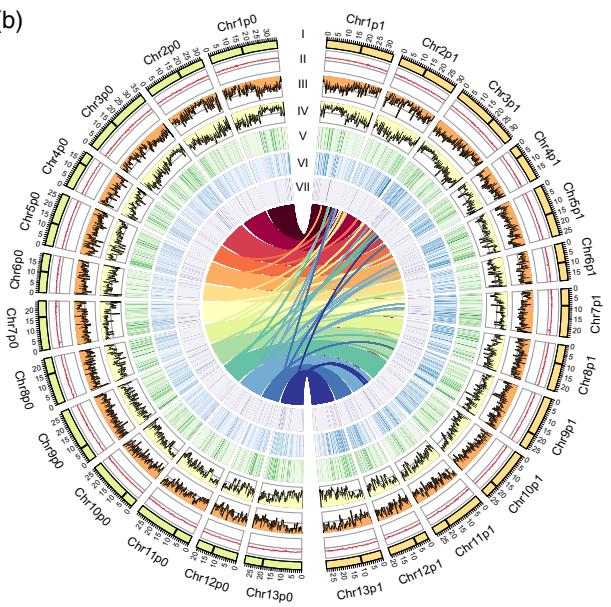

Figure 1. (a) Summary of the haplotype-phased *F. carica* genome assembly and annotation statistics. (b) Distribution of genomic features of the haplotype-phased *F. carica* genome. (I) The 13 pseudo-chromosome pairs. Pseudo-haplotype 0 is reported on the left (light green), and pseudo-haplotype 1 is reported on the right (light orange). The black label on each pseudo-chromosome represents the putative centromeric region. (II) GC content. (III) Gene density. (IV) Transposable elements density. (V) SNPs density. (VI) INDELs density. (VII) Duplications, inversions, and translocations density. The innermost links show the syntenic genes detected between the two pseudo-haplotypes.

and were uniformly identified along the 13 pseudo-chromosomes (Figure 1b; Table S4). Furthermore, the 103 bp-long tandem repeat (Usai et al., 2020) was confirmed as putatively belonging to the centromere (Figure 1b).

For a comprehensive protein-coding gene prediction, we integrated *de novo*, homology-based, and transcriptome-based approaches. In total, 32 985 and 32 682 genes were predicted in fig pseudo-haplotype 0 and pseudo-haplotype 1, representing about 30% of the pseudo-haplotypes, respectively (Figure 1a; Data S1 and S2). Non-coding RNA sequences were also identified, accounting for about 1300 sequences for both pseudo-haplotypes (Data S3).

About 82% of total encoded genes, represented by 27 916 and 27 558 predicted genes for both fig pseudo-haplotypes, matched plant proteins from the National Center for Biotechnology Information (NCBI) non-redundant (nr) database. Overall, 21 725 (~63%) and 21 440 (~64%) predicted genes for pseudo-haplotype 0 and pseudo-haplotype 1, respectively, were functionally characterized through Gene Ontology (GO), Pfam, and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Table S5).

Pseudo-haplotypes comparison and allelic genes characterization

An evaluation of the divergence between the pseudo-haplotypes was performed. The longer pseudo-chromosomes of the homologous pairs were considered the references, and the shorter ones were considered the queries. The higher difference was detected in pseudo-chromosome pair number 11, where the difference between the homologous sequences was 2 040 398 bp (~7%). Conversely, the lower difference was detected in pseudo-chromosome pair number 10, with a size difference of 163 064 bp (less than 1%; Table S6).

The total number of identified genomic variations was 1 828 953 (Table S6), distributed consistently along the assembled pseudo-chromosomes (Figure 1b). In particular, 832 619 single nucleotide polymorphisms (SNPs) and 996 026 insertions/deletions (INDELs) were identified, while 308 SVs were classified as larger rearrangements, including 144 tandem duplications/contractions, 74 inversions, and 90 translocations. The overall intragenomic diversity, *i.e.*, heterozygosity rate, was estimated at 0.52%.

Genomic variations were categorized based on length intervals to elucidate their distribution across the genome (Figure 2a). SNPs and INDELs ranging from 1 to 10 bp were the predominant types of genomic variations, constituting approximately 45% and 53% of the total variations, respectively. Conversely, genomic variations within other length ranges collectively accounted for less than 2%. However, within this small percentage of variations, we find variability in length. Tandem duplications/contractions ranged from 55 to 179 870 bp, inversions ranged from 1614 to 1 273 247 bp, and translocations ranged from 10 009 to 696 148 bp. Forty-seven of these larger rearrangements spanned more than 100 kb.

The p-distances between homologous pseudo-chromosomes, *i.e.*, the proportion of nucleotide sites at which two compared sequences differ, ranged from 0.004 (pseudo-chromosome pair number 1, which showed the fewest differences between pseudo-haplotypes) to 0.0064 (pseudo-chromosome pair number 7; Table S6).

The distribution and potential impact of genomic variations were subsequently analyzed at the gene level, mainly focusing on the allelic gene pairs of the fig tree pseudo-haplotypes. First, 540 syntenic blocks were detected between the two pseudo-haplotypes, covering about 97% of the anchored sequences (about 333 and 331 Mbp per pseudo-haplotype) and about 94% of the total pseudo-haplotype sequences, respectively (Table S7). As mentioned, 32 985 and 32 682 genes were predicted in fig pseudo-haplotypes 0 and 1, respectively. Of them, 19 528 for each pseudo-haplotype were considered reliable allelic gene pairs based on synteny results and manual curation (Table S8). Each allelic gene was analyzed across five distinct regions: the 3 kb upstream region, 5′UTR, CDS, intronic, and 3′UTR to individually evaluate the repartition between homozygous and heterozygous regions (Figure 2b). For each region, except in cases where data were unavailable, metrics including p-distance, Ka, Ks, and Ka/Ks values were computed (Table S8).

Among the paired regions, 10 753 were heterozygous (*i.e.*, with one or more genomic variations on the sequence). The upstream regions (~45%) were most often heterozygous, followed by intronic regions (~41%) and coding sequences (~35%). Conversely, the regions least affected by genomic variations were the 3′UTRs (~26%) and 5′UTRs (~22%).

The distribution of different genomic variations across heterozygous allelic regions was then examined (Figure 2c). Among the five regions analyzed, CDS regions were the only ones exhibiting more SNPs than INDELs (~56%). In contrast, the other regions consistently displayed a higher prevalence of INDELs over SNPs. In particular, the 5′UTRs exhibited the highest occurrence of INDELs over SNPs (~61%). Large rearrangements constituted a minor fraction in this survey, resulting in relatively elevated only in the upstream regions (less than 1%). Within this context, the average p-distance was calculated among the five regions to understand which exhibited a higher ratio of accumulated variations. As expected, the upstream region showed a higher average p-distance, approximately 0.065, while the CDS regions were found to be the least variable, with an average value of 0.002 (Table S8).

Variant annotation analysis identified a total of 349 284 genomic variants across the five analyzed paired regions, comprising 189 677 SNPs (~54%) and 159 607 INDELs (~45%). Notably, most coding variants were non-

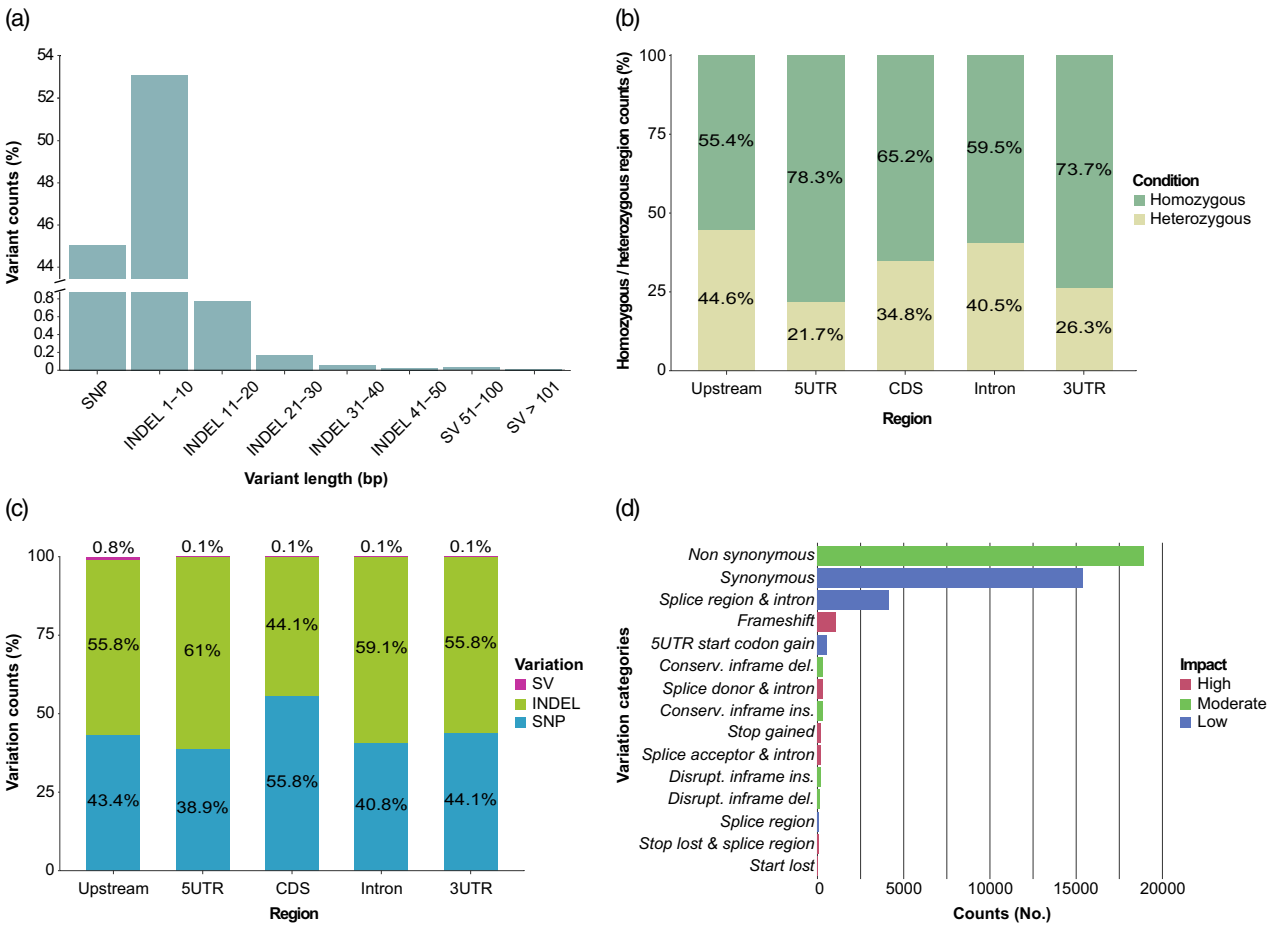

Figure 2. (a) Genome-wide length distribution of SNPs, INDELs, and SVs between pseudo-haplotype 0 and pseudo-haplotype 1 of the *F. carica* genome. SVs include INDELs over 50 bp, inversions, duplications, and translocations. (b) Classification of allelic pair regions (upstream, 5′UTR, CDS, intronic, 3′UTR) as homozygous or heterozygous between pseudo-haplotype 0 and pseudo-haplotype 1 of the *F. carica* genome. (c) Count distribution of genomic variations among the heterozygous allelic pair regions (upstream, 5′UTR, CDS, intronic, 3′UTR) between pseudo-haplotype 0 and pseudo-haplotype 1 of the *F. carica* genome. SVs include INDELs over 50 bp, inversions, duplications, and translocations. (d) Annotation and impact of the 15 most abundant genomic variations on allelic coding regions of the *F. carica* genome, with color-coded bars indicating variation impact (red = high; green = moderate; blue = low).

synonymous mutations, totaling 18 655 variants, constituting approximately 5% of the total variants (Figure 2d; Table S9). Synonymous mutations comprised 15 157 variants, accounting for about 4% of the total variants, and splice region and intron variants, which amounted to 4126 variants, represented approximately 1% of the total variants. The other categories of coding variants were each less than 1%. Six of the 15 most abundant categories of coding variants had a high impact on protein encoding. These categories included "frameshift variant", "splice donor and intron variant", "stop gained", "splice acceptor and intron variant", "stop lost", and "splice region variant and start lost" (Figure 2d).

Whole-genome and allelic methylation analysis in fig fruit

The genome-wide cytosine sites and their corresponding methylation levels were determined for both pseudo-haplotypes and across each tissue. The libraries ranged from about 65 million to about 120 million high-quality paired reads with a fixed length of 150 bp. The coverage achieved spanned from approximately 54x to 101x per library, with an average genome mapping rate of approximately 90% (Table S10). This analysis regarded all three known sequence methylation contexts: CG, CHG, and CHH. The CHH context was the most abundant in the fig genome, with 25.4 million contexts in pseudo-haplotype 0 and 23.6 million contexts in pseudo-haplotype 1 due to the difference in pseudo-haplotype overall length.

Regarding the occurrence of methylation events within individual sequence contexts, quantified as the ratio of methylated contexts to non-methylated contexts, the CG context exhibited more methylation events than the non-methylated state, with a ratio of 1.37 in peel and 1.52 in pulp for both pseudo-haplotypes, respectively, while, in the CHG and CHH contexts, the methylation events were less frequent, with a ratio of 0.73 in peel and 0.82 in pulp

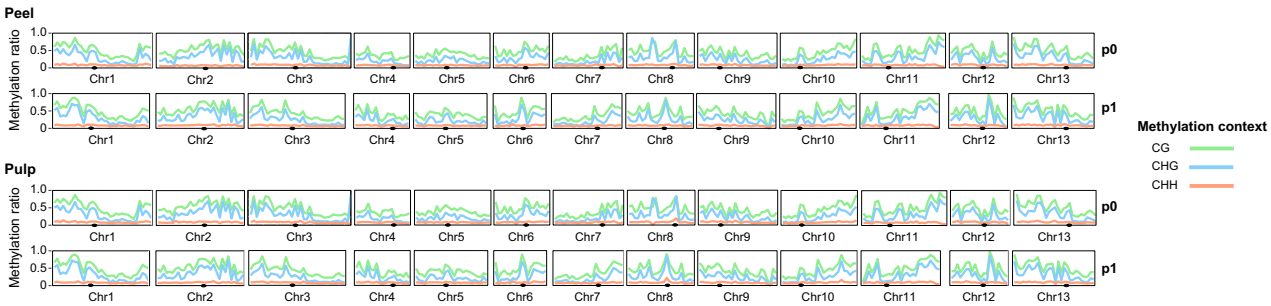

Figure 3. Global methylation profile of the haplotype-phased *F. carica* genome assembly in peel (above) and pulp (below) tissues. CG, CHG, and CHH methylation levels are reported in green, blue, and orange, respectively. Black dots represent the putative centromeric region (p0: pseudo-haplotype 0; p1: pseudo-haplotype 1).

for the CHG context and a ratio of 0.47 in peel and 0.59 in pulp for the CHH context in both pseudo-haplotypes, respectively. These findings indicated that a consistent portion of CG sequence contexts were methylated, while an inferior number of CHH sequence contexts were methylated despite their higher prevalence along the genome sequences.

Figure 3 reports a global overview of the cytosine methylation along the 13 fig pseudo-chromosome pairs, showing that the two pseudo-haplotypes exhibited comparable methylation levels in both peel and pulp tissues, even if slight variations occurred between the two pseudo-haplotypes (*e.g.*, pseudo-chromosome pair number 4 and pseudo-chromosome pair number 7).

To investigate specific methylation levels in fig fruit tissues, we calculated the average methylation ratios across specific regions of the genome sequences, and we performed a comparison between the two pseudo-haplotypes in the two examined tissues (Figure 4). First, we computed the overall methylation level separating coding and non-coding regions (Figure 4a). Coding regions displayed an average methylation level of 0.27, 0.07, and 0.03 for CG, CHG, and CHH contexts in both peel and pulp tissues. Non-coding regions showed an average methylation level of 0.35 and 0.09 for CG and CHH contexts, respectively, in both peel and pulp tissues. The only slight difference was observed in the CHG context, with average methylation levels of 0.24 and 0.23 for peel and pulp, respectively.

After that, the same procedure was applied to the five predefined regions of the identified allelic gene pairs (Figure S3). Even in this analysis, consistent trends were observed across pseudo-haplotypes and in both tissues. Upstream regions exhibited the highest average methylation level across all sequence contexts, with values of 0.22 for CG, 0.16 for CHG, and 0.08 for CHH contexts. The CDS, intronic, and 3′UTR regions displayed relatively high methylation levels only for the CG context, ranging from 0.19 to 0.16. In comparison, very low methylation levels were observed for CHG and CHH contexts, ranging from 0.01 to 0.04 for CHG and from 0.01 to 0.02 for the CHH contexts. The 5′UTRs exhibited poor methylation, even for the CG context (Figure S3).

Subsequently, we scanned the allelic genes for differentially methylated regions (DMRs) between the two pseudo-haplotypes (Figure 4b). In total, we identified 4150 and 4117 allelic gene pairs in peel and pulp, respectively, where at least one of the five regions was differentially methylated for at least one of the three methylation contexts (Tables S11, S12), accounting for approximately 21% of the total allelic genes.

The upstream regions exhibited the highest occurrence of DMRs (Figure 4b). Interestingly, the CHH context was the most frequently found in a differentially methylated state, occurring in approximately 31% of allelic upstream regions in both tissues, followed by the CHG context, found in about 26% of pairs, and by the CG context, approximately 16% of pairs. The other regions exhibited lower occurrences of DMRs and opposite trends in frequency, with DMRs for the CHH context being less frequent than the CG context. In the CDS, intronic, and 3′UTR regions, approximately 6%–7% of allelic pairs exhibited DMRs for the CG context, with occurrences of CHG and CHH contexts below 2%, except for introns, which showed CHG and CHH DMRs at around 3%. The 5′UTRs exhibited a differentially methylated state in less than 2% of the total pairs for each methylation context.

As an additional observation, the three methylation contexts showed different frequencies in DMRs. For instance, among the differentially methylated upstream regions in the peel, approximately 25% were differentially methylated solely for the CHH context, 24% for both CHG and CHH contexts, 22% for all three contexts, 13% for the CHG context, 7% for the CG context, 5% for both CG and CHH contexts, and 5% for both CG and CHG contexts (Tables S11, S12).

From a functional point of view, the allelic genes were classified into “enzyme”, “transcription factor”, or

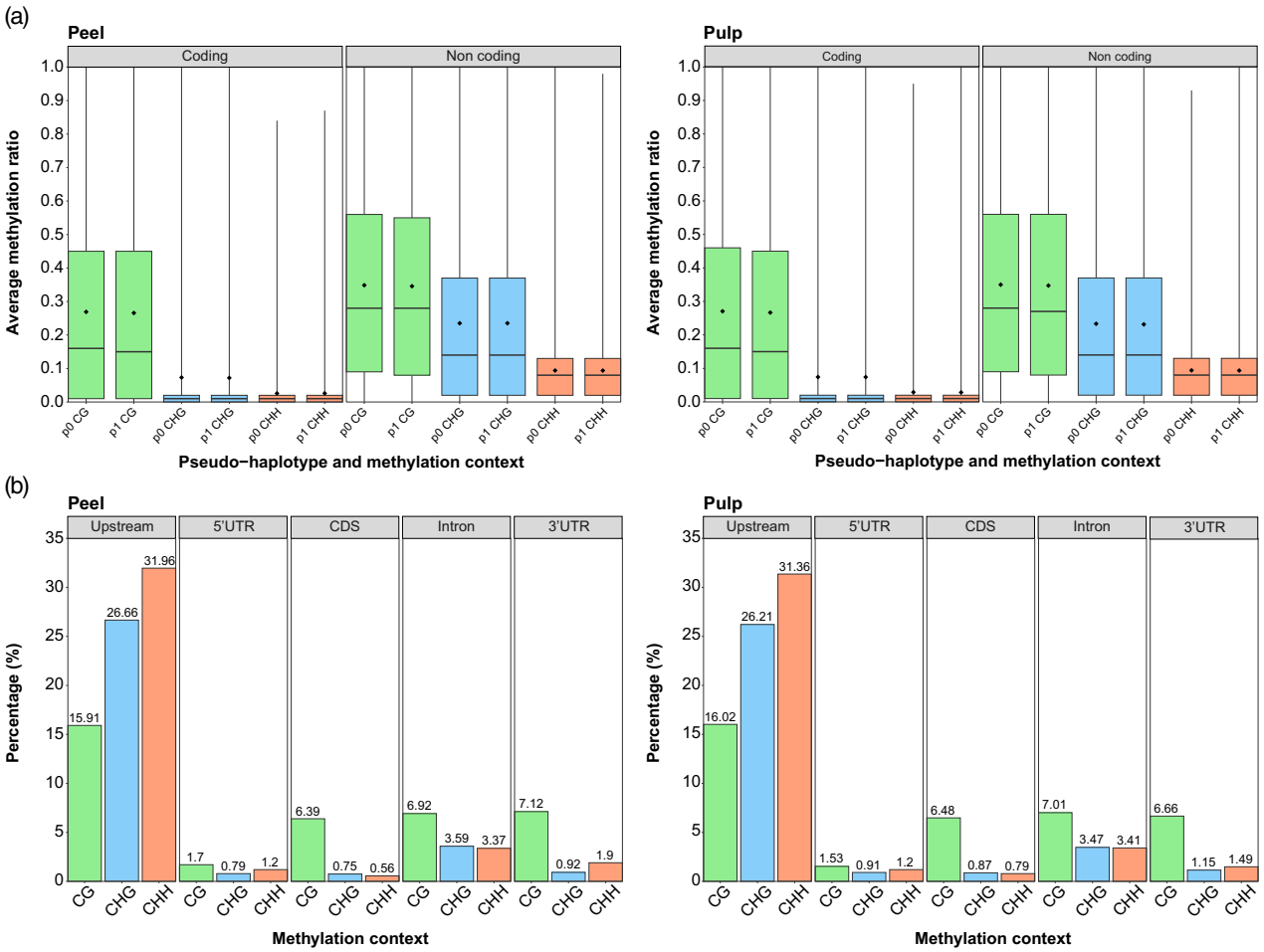

Figure 4. (a) Genome-wide average methylation ratio across coding and non-coding regions of the *F. carica* genome. Results are reported for peel (left) and pulp (right) tissues. Pseudo-haplotypes are indicated as follows: p0 = pseudo-haplotype 0; p1 = pseudo-haplotype 1. Sequence contexts are reported as: CG = green; CHG = blue; CHH = orange. In the bars, the line splitting the box represents the median value, and the dot represents the average value. (b) Percentage of differentially methylated regions across allelic upstream, 5′UTR, CDS, intronic, and 3′UTR regions of the *F. carica* genome. Results are reported for peel (left) and pulp (right) tissues. Sequence contexts are reported as: CG = green; CHG = blue; CHH = orange.

"structural protein" categories. Approximately 49% and 47% of the differentially methylated gene pairs were annotated as "structural protein" in peel and pulp, respectively, while about 36% were annotated as "enzyme" in both tissues. Lastly, approximately 5% and 7% were annotated as "transcription factor" in peel and pulp, respectively (Tables S11, S12). Gene annotation was not available for about 10% of the gene pairs identified with DMRs.

Gene and allele-specific expression analysis in fig fruit

Gene expression analyses were performed to assess transcriptional variations at different levels: (I) variations in expression levels of differentially expressed genes (DEGs) between peel and pulp tissues; (II) allelic expression differences, referred to as ASE and defined as allelic gene pairs wherein the expression between two alleles was significantly different; (III) the potential interactions between the two abovementioned categories.

Six paired-end libraries were generated through sequencing, with three libraries prepared from peel tissues and three from pulp tissues. After trimming and filtering out rRNA-like sequences, high-quality libraries ranging from approximately 15 to 21 million reads were obtained (Table S13).

For a general expression overview, RNA-seq libraries from peel and pulp tissues were mapped onto the CDS regions of fig pseudo-haplotype 0, resulting in about 66%–69% mapped reads for tissue. Of 32 985 genes of the fig reference genome, 18 803 (~55%) were identified as expressed in the peel, while 17 632 (~51%) were expressed in the pulp. These results were refined to isolate expression data for the allelic gene pairs (homozygous and heterozygous, Table S14). The majority of these genes, totaling 10 799 (~55%), were found to be expressed in both tissues (Figure 5a). In contrast, the peel exhibited exclusive expression of 1362 genes (~7%), while the pulp showed

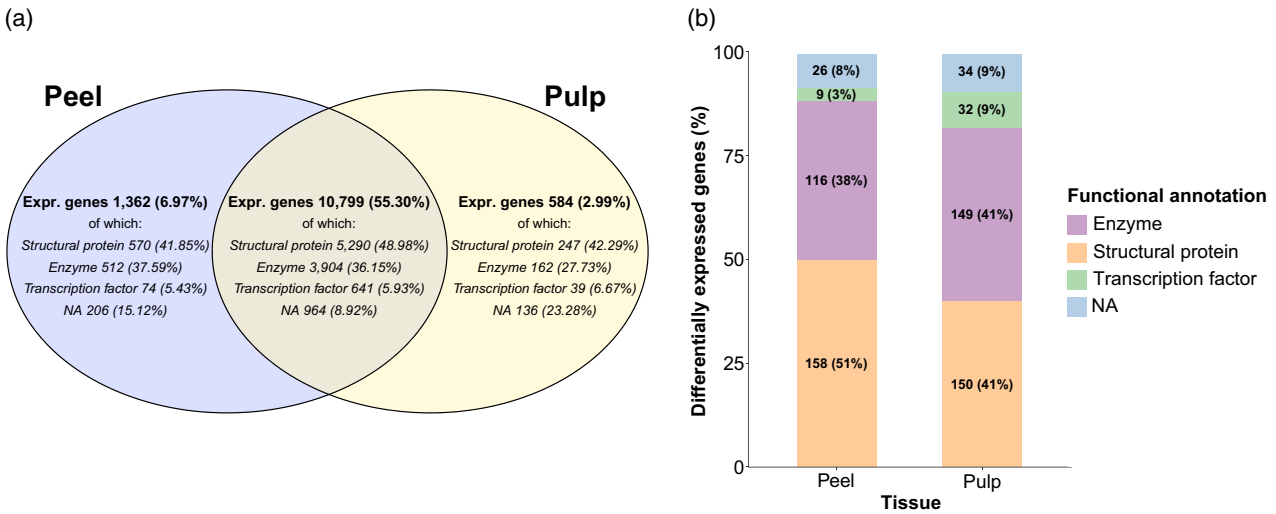

Figure 5. (a) Expression of shared and exclusive allelic gene pairs between peel and pulp tissues of *F. carica*. Functional annotation is indicated in italics. (b) Distribution and functional annotation of differentially expressed genes (DEGs) between peel and pulp tissues of *F. carica*. The count and percentage for each functional category are presented in the bars.

exclusive expression of 584 genes (~3%) (Figure 5a). Finally, 6783 genes (~35%) were not expressed in any tissue analyzed.

Subsequently, gene expression levels were examined to investigate DEGs between peel and pulp tissues (Table S15). This analysis yielded a total of 674 DEGs (Figure 5b). Specifically, 309 DEGs were found to be overexpressed in peel compared to pulp, while 365 DEGs were overexpressed in pulp compared to peel. Functional characterization revealed 158 (~51%) and 150 (~41%) DEGs associated with "structural protein" annotation in peel and pulp, respectively, representing the most abundant functional category. The enzymes were the majority in pulp, with 149 DEGs (~41%), whereas peel exhibited 116 DEGs (~38%). Similarly, the transcription factor-associated genes had more DEGs in pulp, with 32 (~9%), compared to nine in peel (~3%). Twenty-six (~8%) and 34 (~9%) DEGs remained unannotated in peel and pulp, respectively.

After that, we investigated the allelic expression landscape of the two fig tissues. All the allelic gene pairs were found to be expressed in at least one of the two alleles in at least one of the tissues. We then identified and examined the ASE genes, discovering a total of 4024 ASE genes, which accounted for approximately 20% of all allelic genes (Figure 6a; Table S15). 2737 ASE genes (~68% of total ASE genes) were found to be in common in both tissues. Interestingly, the peel tissue exhibited the highest number of exclusive ASE genes, with 728 ones (~18% of total ASE genes), compared to 559 (~14% of total ASE genes) observed only in the pulp tissue. Consistent with the findings from the previous annotation, the functional annotation of ASE genes primarily fell under the category of "structural protein", followed by "enzyme", with a small portion categorized as "transcription factor" (Figure 6a). A percentage ranging from 7% to 13% remained unannotated.

The allelic expression was further evaluated by comparing 2737 ASE genes shared between pulp and peel tissues, aiming to highlight changes in the expression patterns between the two alleles depending on the tissue. The allelic expression patterns remained consistent in both tissues in most ASE genes, accounting for 2715 out of 2737 (Figure 6b). In these cases, the same allele mainly contributes to the overall allelic expression (*i.e.*, one allele contribution ≥90%). This was evident in over 80% of the ASE genes shared in peel and pulp tissues (Figure 6b).

In a smaller number of cases, accounting for 22 ASE genes shared between the two tissues, we detected a switch in the expression pattern between the alleles (Table S16). This group represented the so-called inconsistent ASE genes, wherein the allelic expression contribution underwent a direction-shift among pseudo-haplotypes in relation to the two tissues. In particular, we identified six ASE genes, which, in the peel, were more expressed in the pseudo-haplotype 0 than in the pseudo-haplotype 1, while, in the pulp, the contribution was higher in the pseudo-haplotype 1. On the contrary, the other 16 ASE genes in the peel were more expressed in pseudo-haplotype 1 than in pseudo-haplotype 0, while in the pulp, the contribution was higher than in pseudo-haplotype 0 (Figure 6c). The functional annotation of these 22 switched ASE genes highlighted six ASE genes under the category of "structural protein", two under "enzymes", and four under "transcription factor". These 22 inconsistent ASE genes were associated with the regulation of fruit maturation, seed maturation, and stress response. Specifically, we identified

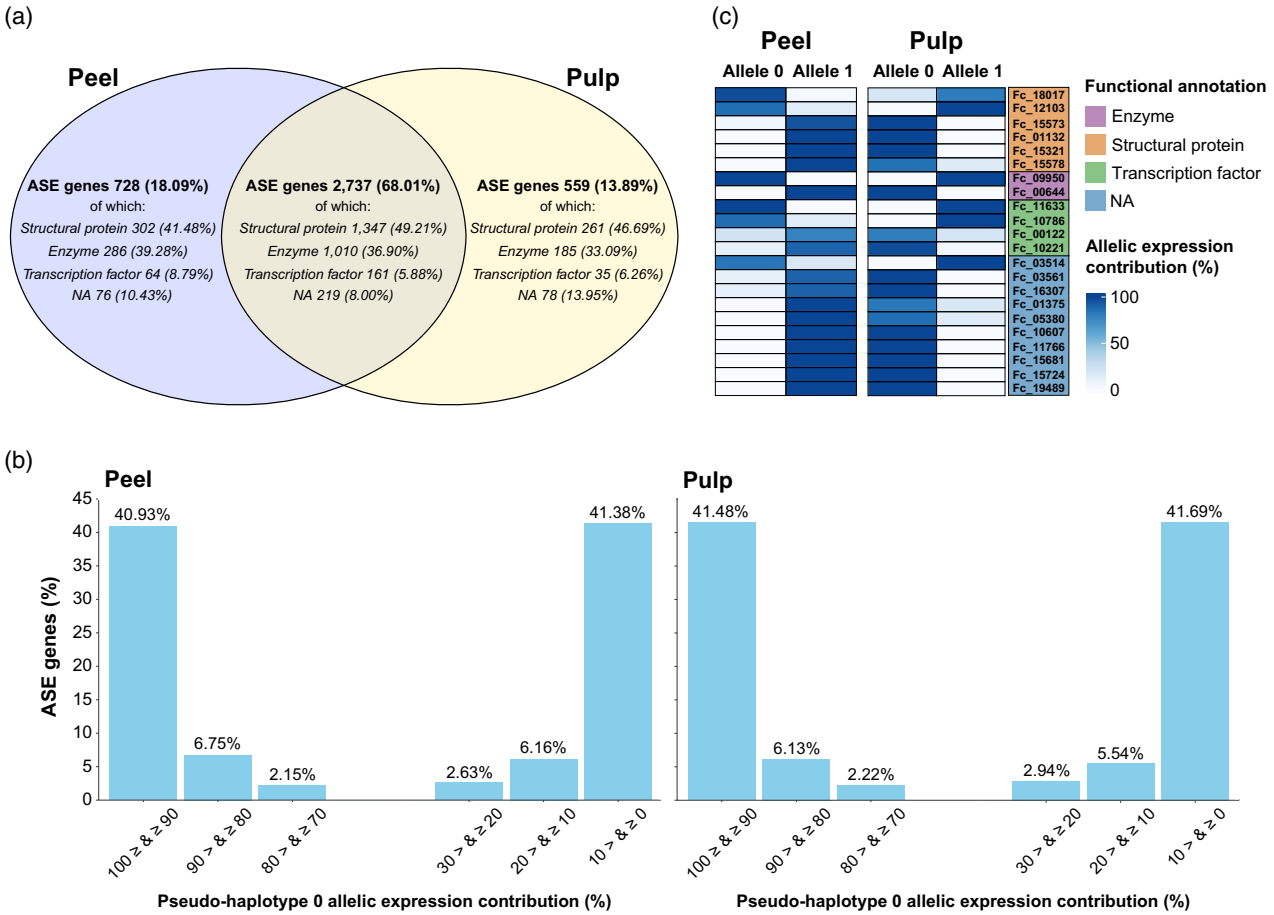

Figure 6. (a) Allele-specific expression (ASE) gene pairs shared and exclusive between peel and pulp tissues of *F. carica*. Functional annotation is indicated in italics. (b) Distribution of consistently expressed ASE genes according to their allelic expression contribution in the pseudo-haplotype 0 of *F. carica*. Results are reported for peel (left) and pulp (right) tissues. The percentage of ASE genes for each range is presented above the bars. (c) Heatmap of ASE genes exhibiting switching behavior in the allelic expression contribution between peel and pulp tissues of *F. carica*. The names of ASE genes are color-coded based on their respective functional annotation.

genes involved in fruit growth and maturation, such as *PSK3*, *LOG1*, *DOF1*, and *CXE12*; genes involved in seed maturation regulation, like *ASIL1*; and genes related to defense mechanisms, like *SRG1* and those with LRR domains (Table S16).

Regarding potential interactions between DEGs and ASE gene expressions, our analysis did not identify any statistically significant interactions (Table S15).

DISCUSSION

The importance of a haplotype-aware for diploid genome assembly

Most current genome sequences are monoploid assemblies, which overlook variations between haplotypes and can lead to errors (Michael & VanBuren, 2020). The use of the Hi-C method and long-read sequencing data allows for improved haplotype genome assemblies without requiring parental data. This approach is crucial for studying genomic and allele-specific structural variations and expression (Pastinen, 2010). Haplotype-aware approaches have been applied in various plant genome assemblies, including diploid potato, vanilla, apple, ginger, and others (Kong et al., 2023).

In this study, we generated a fully phased genome assembly of the fig tree (*Ficus carica* L.) based on long-read sequencing and a Hi-C mapping strategy, representing an updated version of the previously published reference (Usai et al., 2020). Specifically, we performed the accurate haplotype reconstruction of a single genotype of the cultivar "Dottato," which also included gene expression and methylation analyses at both the genome-wide and allele levels.

The fig tree has an estimated haploid genome size of 356 Mbp (Loureiro et al., 2007). Here, we produced two assemblies of haplotype-aware sequences representing the diploid fig tree genome. Both assemblies had a contig N50 of more than 1.9 Mbp, more than twice the N50 value of

the previous version (about 820 kb) that lacked Hi-C integration data (Usai et al., 2020). The sequences were anchored based on markers to generate the 13 pseudo-chromosome pairs of the fig tree, showing an N50 of about 27 Mbp.

Genome assembly improvements were also evident in protein-coding genes and repeat prediction and annotation. The number of genes predicted on the respective pseudo-haplotypes decreased compared to the previous assembly version (from about 37 k to about 33 k genes in the pseudo-haplotypes). This is most likely due to the better resolution of the assembly (*e.g.*, consolidation of allelic variants that were incorrectly considered as regional duplications) and to the comprehensive gene prediction strategy incorporating *de novo*, homology-based, and transcriptome-based predictions that resulted in a more efficient exclusion of false positives. Notably, we observed a relatively high number of duplicated core genes in both pseudo-haplotypes, which may be related to biological phenomena such as duplications and the expansion of gene families. These events are common in plant genomes and can contribute to functional diversification and adaptation. Finally, the improved quality of the assembly also positively affected the identification of the repeats, which in the two pseudo-haplotypes increased by more than 10% compared to the previous version. This could be due to the assembly of previously unresolved (or absent) complex regions that were correctly deciphered through the Hi-C integration.

Uncovering genomic variations and epigenetic landscapes in fig

The new version of the genome assembly, with high completeness and contiguity, allowed us to detect 1 828 953 genomic variations distributed consistently along the assembled pseudo-chromosomes, leading to a more precise estimation of the heterozygosity of the fig tree (about 0.52%). This finding confirms that the species has low/medium heterozygosity when compared to other species in which the phased assembly was obtained similarly, such as *Malus fusca* (about 4.4 million SVs and a heterozygosity of 0.8%; see Mansfeld et al., 2023) and African cassava (heterozygosity of 1.4%; see Mansfeld et al., 2021).

Of the approximately 33 000 genes identified in each pseudo-haplotype, 19 528 gene pairs were identified as reliable alleles. This result was coherent with other published phased plant genomes, in which about 50–60% of genes could be paired reliably between pseudo-haplotypes (Hu et al., 2022; Li et al., 2021; Zhou et al., 2020). In fact, identifying allelic gene pairs can be particularly challenging for genes located in regions with structural variations or tandem duplications. Additionally, specific genes may avoid pairing due to sequence divergence, gene duplications, or their localization in sub-genomic compartments, thus complicating homology determination. The implementation of new strategies, such as the integration of chromatin interaction maps with deep learning approaches (Wang et al., 2024), could assist in recovering unpaired genes.

Among the gene pairs identified, 10 753 were heterozygous, showing modification between the allele sequences. As expected, the upstream region was more variable, while the CDS region was found to be the most conserved, although showing a higher proportion of SNPs than INDELs compared to the other gene regions.

A global examination of cytosine methylation in the fig fruit peel and pulp revealed that the two pseudo-haplotypes exhibited similar methylation levels in both tissues. We also observed that DNA methylation occurred both in the upstream and in transcribed regions of protein-coding genes, suggesting a potential role in regulating gene activity (Greaves et al., 2012; Jin et al., 2008; Zhao et al., 2007). In fact, while promoter methylation is typically associated with gene silencing or downregulation, methylation within the coding sequence can have varying effects on gene expression, either negative or positive (Kumar et al., 2017; Takuno & Gaut, 2013; Williams et al., 2015).

It is generally understood that differences between the two haplotypes of an individual's genome are primarily due to heterozygous SNPs, where the haplotypes contain distinct alleles. In diploid genomes, some alleles exhibit differential DNA methylation patterns, potentially influencing traits such as resistance (Chiba et al., 2018; Do et al., 2016; Kelly et al., 2010; Stern et al., 2017). To explore this further, we scanned DMRs of allelic genes between the two pseudo-haplotypes. Our analysis revealed that approximately 21% of the allelic genes, many of which are annotated as structural proteins, were differentially methylated in at least one of the three methylation contexts within their sequence, especially in the upstream regions. These results must be interpreted considering that the identification of DMRs is biased due to technical factors. One source of underestimation is how sequencing reads are mapped to the reference genome (Akalin et al., 2012), resulting in the aspecific attribution of methylation patterns in the regions where sequence polymorphisms are absent between the pseudo-haplotypes. Consequently, differences between homozygous alleles are overlooked, leading to underestimating DMRs. As a result, this technical limitation can mask actual epigenetic differences, reducing the sensitivity of DMR detection and potentially missing biologically relevant methylation changes.

Hypothesizing dominance and overdominance via allele-specific expression in fig fruit

To investigate gene expression and regulation in fig fruit, we conducted a comprehensive analysis of gene expression patterns in peel and pulp tissues. This analysis

identified a total of 674 DEGs, most of which were associated with the "structural protein" annotation in both tissues. Furthermore, given that genetic variation influences allele-specific expression (Pastinen, 2010), we further explored the allelic expression landscape of the two fig tissues, finding that 4024 allelic gene pairs were differentially expressed in at least one of the tissues, accounting for approximately 20% of the allelic genes. In total, 2715 ASE genes were consistent in our analyses, meaning that one allele was always more highly expressed than the other in both peel and pulp tissues. In particular, 80% of these genes showed a large difference in the expression level between the two alleles; therefore, these genes can be considered genes with a dominant effect. The majority of these genes were annotated as "structural protein", followed by "enzyme" and "transcription factor" categories. In diploid genomes, it is reasonable to assume that when both alleles are functional, they may exhibit different performances depending on the tissue, developmental stage, or environmental conditions. One allele may perform more effectively in certain circumstances, while the other may result superior in other conditions. This sort of regulation favors the condition of heterozygosity over homozygosity, possibly contributing to heterosis and enabling the genome to perform efficiently in different tissues and conditions (Liu et al., 2020; Shao et al., 2019; Zhang et al., 2021).

In support of these considerations, the detection of 22 ASE genes switching allele expression among the fig fruit peel and pulp tissues indicates that direction-shifting may occur, at least for some of the genes. Interestingly, the 22 ASE genes switching allele expression were associated with key biological processes such as fruit maturation, seed maturation, and stress response.

Genes including *PSK3*, *LOG1*, *DOF1*, and *CXE12* are implicated in fruit development, cell differentiation, and proliferation, with evidence supporting their roles in species like strawberry, peach, and kiwifruit (Aghdam et al., 2021; Sutton et al., 2023; Valasiadis et al., 2024). For instance, members of the DOF family of transcription factors are linked to auxin biosynthesis and fruit ripening (Khaksar et al., 2019). Furthermore, the *ASIL1* gene is involved in seed maturation regulation, acting as a negative regulator of seed maturation genes during seed germination and seedling development (Hussain et al., 2020), contributing to the maintenance and control of seed filling and repressing the maturation program during early embryogenesis. Finally, the *SRG1* gene and those containing LRR domains are involved in the plant's defense mechanisms, contributing to the response against stress factors in several fleshy fruits (Cui et al., 2018; Li et al., 2024).

The presence of inconsistent ASE, with one allele more expressed than the other in certain tissues or conditions and less expressed than the other in other tissues or conditions, suggests that the genome may be able to switch the expression of the two alleles in response to the environmental and developmental cues, which may result in an advantage for the heterozygous condition leading to overdominance. Accumulating such differentially advantageous effects of the two alleles may provide a significant cause for heterosis when the genome makes more use of one allele in specific conditions by differential regulation of the two alleles (Zhang et al., 2021). The fact that there were more consistent ASE than inconsistent ASE genes, in line with previous research findings in *Camellia sinensis* and *Phanera championii* (Lu et al., 2024; Zhang et al., 2021), suggests a predominance of the dominant effect in the genome of the fig tree.

Finally, we searched for interactions between DEGs and ASE gene expressions, leading to no statistically significant results. The absence of interactions between DEGs and ASE genes may indicate that their differential expression patterns and allele-specific tendencies operate independently under the investigated conditions rather than being co-regulated or influenced by shared factors. Alternatively, it cannot be excluded that any potential interactions between DEGs and ASE genes were too slight to be detected with the current experimental design or statistical thresholds. Further research will be crucial to uncover potential interactions under different experimental conditions or within specific genomic contexts.

Potential implications and perspectives

Using the PacBio sequencing strategy in combination with Hi-C, we successfully reconstructed a haplotype-aware genome for the diploid fig tree cultivar "Dottato". This advancement enabled us to study the sequence, gene content, gene expression, and genome structure with unprecedented resolution. The phased assembly also provided an ideal reference for the first attempt to assess haplotype-level methylation in two fruit tissues, laying the groundwork for future studies to unravel tissue-specific or stage-specific methylation patterns. Additionally, the identification of direction-shifting allele expression revealed intriguing cases of overdominance, offering new insights into gene regulation.

Together, the haplotype-resolved genome, methylome, and transcriptome represent valuable resources that will support both fig breeding efforts and further research into fig genetics and epigenomics.

EXPERIMENTAL PROCEDURES

Genome assembly and curation

All information regarding the plant material of the Italian *F. carica* cultivar "Dottato," DNA and RNA isolation, and sequencing procedures are reported in the Supporting Information (Methods S1 and Methods S2).

The fig primary contigs and the haplotigs produced by Usai et al. (2020) (accession number: GCA_009761775.1) were curated using Purge Haplotigs v1.1.2 (Roach et al., 2018) to reassign misplaced allelic contigs with a very high degree of heterozygosity (parameters: -l 15 -m 55 -h 100). SAMtools v1.7 (Li et al., 2009) was used for sorting and indexing the PacBio reads, and minimap2 v2.22 (Li, 2018) was used to perform the alignment against the contigs (parameter: -ax map-pb).

The first run of FALCON-Phase v2 (Kronenberg et al., 2021) was performed to correct phase switching errors in the primary contigs and alternative haplotigs and to generate two sets of contigs for each phase, called phase 0 and phase 1 (parameters: min_aln_len 5000 enzyme GATC, GAATC, GATTC, GAGTC, GACTC iter 10 000 000). During the phasing process, BWA-MEM v0.7.17 (Li & Durbin, 2009) was used to align the Hi-C reads (Methods S2) to the contigs (parameter: -5SP), and the alignments were then filtered with SAMtools (parameters: -F 2304) to remove non-primary alignments. The phasing process also required BEDTools v2.27.1 (Quinlan & Hall, 2010) and MUMmer v4.0.0 (Marçais et al., 2018) programs to be installed and run with default parameters.

The Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) was used to map the Hi-C paired-end reads to the contigs necessary for the scaffolding process. The mapping was performed on the phase 0 contigs using the parameters reported in the pipeline. Picard v2.9 (http://broadinstitute.github.io/picard) was used to flag PCR duplicates in the Hi-C dataset for subsequent exclusion using default parameters.

Based on the mapping results, SALSA v2.2 (Ghurye et al., 2017) was used to assemble contigs into scaffolds (parameters: -e GATC, GANTC -o scaffolds -m yes). This method produces scaffolds by maximizing the predicted contact frequency from the aligned Hi-C paired-end reads, normalized by the number of restriction sites (GATC and GANTC) on each contig.

The second run of FALCON-Phase was run to correct eventual phase-switching errors, generating the two sets of haplotype-phased sequences, defined as pseudo-haplotypes. The program parameters were the same as those used in the first run.

The two pseudo-haplotypes were polished with the PacBio reads using Arrow v2.3.3 (https://github.com/PacificBiosciences/GenomicConsensus). The alignment was performed using BLASR v5.3 (Chaisson & Tesler, 2012). Finally, Juicebox v1.11.08 (Durand et al., 2016) was used to visualize the Hi-C interaction heatmap and manually correct scaffolding errors.

FALCON-Phase separates maternal and paternal sequences from the same chromosomes but is limited in establishing which parent the assembled chromosomes come from (Kronenberg et al., 2021). We decided to separate the paired pseudo-chromosomes according to their length. The longer pseudo-chromosomes were attributed to pseudo-haplotype 0, and the shorter ones were attributed to pseudo-haplotype 1.

The quality of both pseudo-haplotypes was assessed using BUSCO v4.0.6 (Simão et al., 2015) to calculate the assembly completeness regarding genes and using LTR_retriever v2.9.0 (Ou & Jiang, 2018) to calculate the LTR assembly Index (LAI). BUSCO was run using the Embryophyta v10 Plantae dataset (https://busco.ezlab.org). Both programs were run with default parameters.

Finally, we used the Illumina sequences of the *F. carica* cultivar "Horaishi" assembly (GCA_002002945.1; Mori et al., 2017) ordered by simple sequence repeat (SSR) and single nucleotide polymorphisms (SNPs) maps to anchor our phased sequences to the 13 pseudo-chromosomes of the fig tree. The anchoring process was performed using BLAST v2.6.0+ (Altschul et al., 1990) through a BLASTN process (parameters: -num_alignments 1 -evalue 1e-10). Additionally, we only considered hits with percent alignment values of more than 85% and query and alignment length values of more than 1000 bp. Sequences were considered to correspond to a specific linkage group when two or more "Horaishi" sequences of the same linkage group were located on the same sequence.

A comprehensive analysis of the repetitive component was performed on the pseudo-haplotypes by using structural and homology-based approaches (Methods S3). Finally, the high-quality protein-coding genes on the two repeat-masked pseudo-haplotypes were predicted using a combination of *de novo*, homology-based, and transcriptome-based approaches (Methods S4).

Assemblathon v2 (Bradnam et al., 2013) was used to calculate all assembly statistics. A global representation of the genomic landscape was produced with Circos (Krzywinski et al., 2009).

Pseudo-haplotypes diversity and allelic genes characterization

The pseudo-haplotype 0 and pseudo-haplotype 1 of the fig tree were aligned to evaluate SNPs, INDELs, and SVs, defined as INDELs greater than 50 bp, duplications/contractions, inversions, and translocations.

MUMmer was used to construct the alignments between the two pseudo-haplotypes (parameters: -c 500 -b 500 -l 100). The output was filtered using the delta-filter function (parameters: -m -i 90 -l 100) and converted into an appropriate format using the show-coords function (parameters: -THrd). Both functions are included in the MUMmer package. After that, SyRI v1.6.5 (Goel et al., 2019) was used to identify all the above-mentioned genomic variations. The process was run with default parameters.

The heterozygosity of the cultivar "Dottato" was estimated by calculating the frequency of the genomic variations between the two pseudo-haplotypes per kilobase (the number of total genomic variations every 1000 bp).

Furthermore, the two pseudo-haplotypes predicted genes were analyzed to identify putative allelic genes. First, we performed a BLASTN process among the CDS predicted on both pseudo-haplotypes (parameter: -evalue 1e-7). Subsequently, MCScanX (Wang et al., 2012) was used to construct the syntenic blocks based on the obtained well-aligned CDS (parameters: MATCH_SCORE: 50 MATCH_SIZE: 4 GAP_PENALTY: -1 OVERLAP_WINDOW: 5 E_VALUE: 1e-05 MAX GAPS: 25). The allelic genes were screened to confirm the presence of paired regions on the homologous sequences. 1000 bp upstream and downstream of the genes were extracted using BEDTools. These sequences were separately subjected to a BLASTN process to select sequence pairs showing a percent alignment value greater than or equal to 85% (parameter: -evalue 1e-10). Genes with matching paired upstream and downstream regions were confirmed as putative allelic genes (representing the gene map).

Each gene pair was analyzed across five distinct regions: 3 kb upstream region, 5′UTR, CDS, intronic, and 3′UTR. The p-distance, Ka and Ks values between the allelic regions were calculated using KaKs_Calculator v3.0 (Zhang, 2022). The previously identified genomic variations were intersected among the allelic regions using BEDTools and annotated using SnpEff v5.1 (Cingolani et al., 2012). The program was run with default parameters applied. Based on the impact level of the genomic variations on the coding regions, they were categorized into four classes: "high", "moderate", "low", and "modifier" (Cingolani et al., 2012).

Whole-genome and allelic methylation analysis in fig fruit

The WGB reads obtained from peel and pulp tissues (Methods S2) were aligned to the fig tree genome assembly using BSMAP v2.90 (Xi & Li, 2009) (parameters: -r 0 -v 0.04). The methylation rates for CG, CHG, and CHH across the phased genome assembly were calculated using the methylated cytosine/total cytosine ratio through the methratio.py script within the BSMAP package. Methylation rates were then calculated genome-wide, between coding and non-coding regions, and in the allelic gene regions (*i.e.*, upstream region, 5′UTR, CDS, intronic, and 3′UTR). MethylKit (Akalin et al., 2012) was used to identify DMRs between the allelic gene regions based on a reads depth of ≥6, differences in methylation rate for CG >0.4, for CHG >0.2, for CHH >0.1, and a *P*-value <0.01 derived from a two-tailed Fisher's exact test (Wang et al., 2018; Zhou et al., 2020). Only DMRs supported by all four replicates were retained for subsequent analysis.

Gene and allele-specific expression analysis in fig fruit

The twelve RNA paired-end libraries (Methods S2) were aligned to the predicted pseudo-haplotype 0 and 1 CDS to assess gene expression levels between peel and pulp tissues using Kallisto v0.46.1 (Bray et al., 2016) with default parameters. The read counts per transcript were converted to reads per kilobase per million mapped reads (RPKM) to quantify gene expression levels (Mortazavi et al., 2008). Only genes exhibiting RPKM values greater than 1 in at least one library were considered expressed and included in further analysis. The raw counts generated from the alignment were subjected to analysis using the R statistical package "edgeR" (Robinson et al., 2010) to identify DEGs between peel and pulp libraries, ASE gene pairs between the two pseudo-haplotypes, and the expression interaction between DEGs and ASE genes. The *P*-values were adjusted using the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995). Fold change (FC) values were normalized via a log2 transformation. Genes were considered differentially expressed in tissues or in alleles if they exhibited an absolute |log2 FC| value of ≥1 and an FDR value of <0.05.

AUTHOR CONTRIBUTIONS

GU, TG, LN, FM, and AC: conceived, conceptualized, and designed the study; GU, AV, MC, SS, EB, and FM: performed the experiments, bioinformatics analysis, and data curation of the study; GU, TG, AV, MC, SS, EB, LN, FM, and AC: investigated and interpreted the results of the study; GU, TG, AV, MC, SS, EB, LN, FM, and AC: discussed the data, wrote the manuscript, and contributed to its final form.

ACKNOWLEDGMENTS

The authors thank Prof. Brandon Gaut for his critical reading of the manuscript. This study was performed in the frame of the FIGGEN/PRIMA19_00197 project, which is part of the PRIMA Programme supported by the European Union through a national MIUR (Italy) grant.

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data has been deposited in the National Center for Biotechnology Information (NCBI) repository under BioProject PRJNA1111048 and PRJNA1111063. The haplotype-phased fig genome has been deposited at DDBJ/ENA/GenBank under accession JBDYCZ000000000 and JBDYDA000000000. The versions described in this paper are JBDYCZ010000000 and JBDYDA010000000. The Hi-C sequences have been deposited in the Sequence Read Archive (SRA) with accession numbers SRR29142444 and SRR29142443. The whole-genome bisulfite (WGB) sequences have been deposited in the SRA with accession numbers SRR29156273, SRR29156274, SRR29156275, SRR29156276, SRR29156277, and SRR29156278. The RNA-Seq sequences have been deposited in the SRA with accession numbers SRR29160843, SRR29160844, SRR29160845, SRR29160846, SRR29160847, and SRR29160848.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Data S1. Phase 0 protein-coding gene prediction.
Data S2. Phase 1 protein-coding gene prediction.
Data S3. Phase 0 and Phase 1 non-coding RNA prediction.
Figure S1. Genome assembly workflow.
Figure S2. Hi-C interaction heatmap.
Figure S3. Methylation ratios of allelic regions.
Table S1. Intermediate assembly statistics.
Table S2. Genetic-to-physical map of the fig genome.
Table S3. BUSCO assessment of the fig genome assembly.
Table S4. Repeat annotation of the fig genome assembly.
Table S5. Gene annotation of the fig genome assembly.
Table S6. Genomic variations of the fig genome assembly.
Table S7. Syntenic blocks of the fig genome assembly.
Table S8. Allelic gene pairs of the fig genome assembly.
Table S9. Variant annotation of the allelic gene pairs.
Table S10. Whole-genome bisulfite library statistics.
Table S11. Peel differentially methylated regions.
Table S12. Pulp differentially methylated regions.
Table S13. RNA-Seq library statistics.
Table S14. RPKM values of allelic genes.
Table S15. DEGs, ASE genes, and interaction statistics.
Table S16. Inconsistent ASE genes.
Method S1. Plant material, DNA, and RNA isolation.
Method S2. Hi-C, transcriptome, and methylome sequencing.
Method S3. Repeat prediction and annotation.
Method S4. Gene prediction and annotation.

REFERENCES

Aghdam, M.S., **Sayyari, M.** & **Luo, Z.** (2021) Exogenous phytosulfokine α application delays senescence and promotes antioxidant nutrient accumulation in strawberry fruit during cold storage by triggering endogenous phytosulfokine α signaling. *Postharvest Biology and Technology*, **175**(111), 473. Available from: https://doi.org/10.1016/j.postharvbio.2021.111473

Akalin, A., **Kormaksson, M.**, **Li, S.**, **Garrett-Bakelman, F.E.**, **Figueroa, M.E.**, **Melnick, A.** ***et al.*** (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, **13**, R87. Available from: https://doi.org/10.1186/gb-2012-13-10-r87

Allegra, A., **Gallotta, A.**, **Carimi, F.**, **Mercati, F.**, **Inglese, P.** & **Martinelli, F.** (2018) Metabolic profiling and post-harvest behavior of "Dottato" fig (Ficus carica L.) fruit covered with an edible coating from O. Ficus-indica. *Frontiers in Plant Science*, **9**, 856. Available from: https://doi.org/10.3389/fpls.2018.01321

Allegra, A., **Sortino, G.**, **Inglese, P.**, **Settanni, L.**, **Todaro, A.** & **Gallotta, A.** (2017) The effectiveness of Opuntia ficus-indica mucilage edible coating on post-harvest maintenance of 'Dottato' fig (Ficus carica L.) fruit. *Food Packaging and Shelf Life*, **12**, 135–141. Available from: https://doi.org/10.1016/j.fpsl.2017.04.010

Altschul, S.F., **Gish, W.**, **Miller, W.**, **Myers, E.W.** & **Lipman, D.J.** (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410. Available from: https://doi.org/10.1016/S0022-2836(05)80360-2

Bao, Y., **He, M.**, **Zhang, C.**, **Jiang, S.**, **Zhao, L.**, **Ye, Z.** ***et al.*** (2023) Advancing understanding of Ficus carica: a comprehensive genomic analysis reveals evolutionary patterns and metabolic pathway insights. *Frontiers in Plant Science*, **14**, 1214. Available from: https://doi.org/10.3389/fpls.2023.1298417

Benjamini, Y. & **Hochberg, Y.** (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, **57**, 289–300. Available from: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bradnam, K.R., **Fass, J.N.**, **Alexandrov, A.**, **Baranay, P.**, **Bechner, M.**, **Birol, I.** ***et al.*** (2013) Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, **2**, 10. Available from: https://doi.org/10.1186/2047-217X-2-10

Bray, N.L., **Pimentel, H.**, **Melsted, P.** & **Pachter, L.** (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, **34**, 525–527. Available from: https://doi.org/10.1038/nbt.3519

Chaisson, M.J. & **Tesler, G.** (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*, **13**, 238. Available from: https://doi.org/10.1186/1471-2105-13-238

Chiba, H., **Kakuta, Y.**, **Kinouchi, Y.**, **Kawai, Y.**, **Watanabe, K.**, **Nagao, M.** ***et al.*** (2018) Allele-specific DNA methylation of disease susceptibility genes in Japanese patients with inflammatory bowel disease. *PLoS One*, **13**, e0194036. Available from: https://doi.org/10.1371/journal.pone.0194036

Cingolani, P., **Platts, A.**, **Wang, L.L.**, **Coon, M.**, **Nguyen, T.**, **Wang, L.** ***et al.*** (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*, **6**, 80–92. Available from: https://doi.org/10.4161/fly.19695

Cokus, S.J., **Feng, S.**, **Zhang, X.**, **Chen, Z.**, **Merriman, B.**, **Haudenschild, C.D.** ***et al.*** (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **452**, 215–219. Available from: https://doi.org/10.1038/nature06745

Cui, Y., **Jiang, J.**, **Yang, H.**, **Zhao, T.**, **Xu, X.** & **Li, J.** (2018) Virus-induced gene silencing (VIGS) of the NBS-LRR gene SLNLC1 compromises Sm-mediated disease resistance to Stemphylium lycopersici in tomato. *Biochemical and Biophysical Research Communications*, **503**, 1524–1529. Available from: https://doi.org/10.1016/j.bbrc.2018.07.074

Do, C., **Lang, C.F.**, **Lin, J.**, **Darbary, H.**, **Krupska, I.**, **Gaba, A.** ***et al.*** (2016) Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *American Journal of Human Genetics*, **98**, 934–955. Available from: https://doi.org/10.1016/j.ajhg.2016.03.027

Durand, N.C., **Robinson, J.T.**, **Shamim, M.S.**, **Machol, I.**, **Mesirov, J.P.**, **Lander, E.S.** ***et al.*** (2016) Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Systems*, **3**, 99–101. Available from: https://doi.org/10.1016/j.cels.2015.07.012

Ghurye, J., **Pop, M.**, **Koren, S.**, **Bickhart, D.** & **Chin, C.-S.** (2017) Scaffolding of long read assemblies using long range contact information. *BMC Genomics*, **18**, 527. Available from: https://doi.org/10.1186/s12864-017-3879-z

Goel, M., **Sun, H.**, **Jiao, W.-B.** & **Schneeberger, K.** (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, **20**, 277. Available from: https://doi.org/10.1186/s13059-019-1911-0

Greaves, I.K., **Groszmann, M.**, **Ying, H.**, **Taylor, J.M.**, **Peacock, W.J.** & **Dennis, E.S.** (2012) Trans chromosomal methylation in Arabidopsis hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3570–3575. Available from: https://doi.org/10.1073/pnas.1201043109

Guk, J.Y., **Jang, M.J.**, **Choi, J.W.**, **Lee, Y.M.** & **Kim, S.** (2022) De novo phasing resolves haplotype sequences in complex plant genomes. *Plant Biotechnology Journal*, **20**, 1031–1041. Available from: https://doi.org/10.1111/pbi.13815

Hamilton, J.P. & **Robin Buell, C.** (2012) Advances in plant genome sequencing. *The Plant Journal*, **70**, 177–190. Available from: https://doi.org/10.1111/j.1365-313X.2012.04894.x

Hu, G., **Feng, J.**, **Xiang, X.**, **Wang, J.**, **Salojärvi, J.**, **Liu, C.** ***et al.*** (2022) Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nature Genetics*, **54**, 73–83. Available from: https://doi.org/10.1038/s41588-021-00971-3

Hussain, Q., **Shi, J.**, **Scheben, A.**, **Zhan, J.**, **Wang, X.**, **Liu, G.** ***et al.*** (2020) Genetic and signalling pathways of dry fruit size: targets for genome editing-based crop improvement. *Plant Biotechnology Journal*, **18**, 1124–1140. Available from: https://doi.org/10.1111/pbi.13318

Ikegami, H., **Shirasawa, K.**, **Yakushiji, H.**, **Toyoda, A.**, **Hayashi, T.**, **Yabe, S.** ***et al.*** (2024) Genome-wide association studies using chromosome-scale genomes of male and female lines redefines two sex-linked loci in linkage disequilibrium in Ficus carica L. *Scientia Horticulturae*, **336**(113), 424. Available from: https://doi.org/10.1016/j.scienta.2024.113424

Ji, L., **Neumann, D.A.** & **Schmitz, R.J.** (2015) Crop epigenomics: identifying, unlocking, and harnessing cryptic variation in crop genomes. *Molecular Plant*, **8**, 860–870. Available from: https://doi.org/10.1016/j.molp.2015.01.021

Jin, H., **Hu, W.**, **Wei, Z.**, **Wan, L.**, **Li, G.**, **Tan, G.** ***et al.*** (2008) Alterations in cytosine methylation and species-specific transcription induced by interspecific hybridization between Oryza sativa and O. Officinalis. *Theoretical and Applied Genetics*, **117**, 1271–1279. Available from: https://doi.org/10.1007/s00122-008-0861-9

Kelly, T.K., **De Carvalho, D.D.** & **Jones, P.A.** (2010) Epigenetic modifications as therapeutic targets. *Nature Biotechnology*, **28**, 1069–1078. Available from: https://doi.org/10.1038/nbt.1678

Khaksar, G., **Sangchay, W.**, **Pinsorn, P.**, **Sangpong, L.** & **Sirikantaramas, S.** (2019) Genome-wide analysis of the Dof gene family in durian reveals fruit ripening-associated and cultivar-dependent Dof transcription factors. *Scientific Reports*, **9**(12), 109. Available from: https://doi.org/10.1038/s41598-019-48601-7

Kislev, M.E., **Hartmann, A.** & **Bar-Yosef, O.** (2006) Early domesticated fig in the Jordan Valley. *Science*, **312**, 1372–1374. Available from: https://doi.org/10.1126/science.1125910

Kong, W., **Wang, Y.**, **Zhang, S.**, **Yu, J.** & **Zhang, X.** (2023) Recent advances in assembly of complex plant genomes. *Genomics, Proteomics & Bioinformatics*, **21**, 427–439. Available from: https://doi.org/10.1016/j.gpb.2023.04.004

Kronenberg, Z.N., **Rhie, A.**, **Koren, S.**, **Concepcion, G.T.**, **Peluso, P.**, **Munson, K.M.** ***et al.*** (2021) Extended haplotype-phasing of long-read de novo genome assemblies using hi-C. *Nature Communications*, **12**, 1935. Available from: https://doi.org/10.1038/s41467-020-20536-y

Krzywinski, M., **Schein, J.**, **Birol, İ.**, **Connors, J.**, **Gascoyne, R.**, **Horsman, D.** ***et al.*** (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645. Available from: https://doi.org/10.1101/gr.092759.109

Kumar, S., **Beena, A.S.**, **Awana, M.** & **Singh, A.** (2017) Physiological, biochemical, epigenetic and molecular analyses of wheat (Triticum aestivum) genotypes with contrasting salt tolerance. *Frontiers in Plant Science*, **8**, 1151. Available from: https://doi.org/10.3389/fpls.2017.01151

Law, J.A. & **Jacobsen, S.E.** (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews, Genetics*, **11**, 204–220. Available from: https://doi.org/10.1038/nrg2719

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100. Available from: https://doi.org/10.1093/bioinformatics/bty191

Li, H. & **Durbin, R.** (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760. Available from: https://doi.org/10.1093/bioinformatics/btp324

Li, H., **Handsaker, B.**, **Wysoker, A.**, **Fennell, T.**, **Ruan, J.**, **Homer, N.** ***et al.*** (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079. Available from: https://doi.org/10.1093/bioinformatics/btp352

Li, H.L., **Wu, L.**, **Dong, Z.**, **Jiang, Y.**, **Jiang, S.**, **Xing, H.** ***et al.*** (2021) Haplotype-resolved genome of diploid ginger (Zingiber officinale) and its unique gingerol biosynthetic pathway. *Horticulture Research*, **8**, 189. Available from: https://doi.org/10.1038/s41438-021-00627-7

Li, J., **Han, F.**, **Yuan, T.**, **Li, W.**, **Li, Y.**, **Wu, H.X.** ***et al.*** (2023) The methylation landscape of giga-genome and the epigenetic timer of age in Chinese pine. *Nature Communications*, **14**, 1947. Available from: https://doi.org/10.1038/s41467-023-37684-6

Li, S., **Zhao, Y.**, **Wu, P.**, **Grierson, D.** & **Gao, L.** (2024) Ripening and rot: how ripening processes influence disease susceptibility in fleshy fruits. *Journal of Integrative Plant Biology*, **66**, 1831–1863. Available from: https://doi.org/10.1111/jipb.13739

Liu, J., **Li, M.**, **Zhang, Q.**, **Wei, X.** & **Huang, X.** (2020) Exploring the molecular basis of heterosis for plant breeding. *Journal of Integrative Plant Biology*, **62**, 287–298. Available from: https://doi.org/10.1111/jipb.12804

López, M.-E., **Roquis, D.**, **Becker, C.**, **Denoyes, B.** & **Bucher, E.** (2022) DNA methylation dynamics during stress response in woodland strawberry (Fragaria vesca). *Horticulture Research*, **9**, uhac174. Available from: https://doi.org/10.1093/hr/uhac174

Loureiro, J., **Rodriguez, E.**, **Dolezel, J.** & **Santos, C.** (2007) Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany*, **100**, 875–888. Available from: https://doi.org/10.1093/aob/mcm152

Lu, Y., **Chen, X.**, **Yu, H.**, **Zhang, C.**, **Xue, Y.**, **Zhang, Q.** ***et al.*** (2024) Haplotype-resolved genome assembly of Phanera championii reveals molecular mechanisms of flavonoid synthesis and adaptive evolution. *The Plant Journal*, **118**, 488–505. Available from: https://doi.org/10.1111/tpj.16620

Mansfeld, B.N., **Boyher, A.**, **Berry, J.C.**, **Wilson, M.**, **Ou, S.**, **Polydore, S.** ***et al.*** (2021) Large structural variations in the haplotype-resolved African cassava genome. *The Plant Journal*, **108**, 1830–1848. Available from: https://doi.org/10.1111/tpj.15543

Mansfeld, B.N., **Yocca, A.**, **Ou, S.**, **Harkess, A.**, **Burchard, E.**, **Gutierrez, B.** ***et al.*** (2023) A haplotype-resolved chromosome-scale assembly of north American wild apple Malus fusca and comparative genomics of the fire blight Mfu10 locus. *The Plant Journal*, **116**, 989–1002. Available from: https://doi.org/10.1111/tpj.16433

Marçais, G., **Delcher, A.L.**, **Phillippy, A.M.**, **Coston, R.**, **Salzberg, S.L.** & **Zimin, A.** (2018) MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, **14**, e1005944. Available from: https://doi.org/10.1371/journal.pcbi.1005944

Mascellani, A., **Natali, L.**, **Cavallini, A.**, **Mascagni, F.**, **Caruso, G.**, **Gucci, R.** ***et al.*** (2021) Moderate salinity stress affects expression of main sugar metabolism and transport genes and soluble carbohydrate content in ripe fig fruits (Ficus carica L. cv. Dottato). *Plants*, **10**, 1861. Available from: https://doi.org/10.3390/plants10091861

Michael, T.P. & **VanBuren, R.** (2020) Building near-complete plant genomes. *Current Opinion in Plant Biology*, **54**, 26–33. Available from: https://doi.org/10.1016/j.pbi.2019.12.009

Mori, K., **Shirasawa, K.**, **Nogata, H.**, **Hirata, C.**, **Tashiro, K.**, **Habu, T.** ***et al.*** (2017) Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (Ficus carica L.). *Scientific Reports*, **7**, 41124. Available from: https://doi.org/10.1038/srep41124

Mortazavi, A., **Williams, B.A.**, **McCue, K.**, **Schaeffer, L.** & **Wold, B.** (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**, 621–628. Available from: https://doi.org/10.1038/nmeth.1226

Ou, S., **Chen, J.** & **Jiang, N.** (2018) Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Research*, **46**, e126. Available from: https://doi.org/10.1093/nar/gky730

Ou, S. & **Jiang, N.** (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**, 1410–1422. Available from: https://doi.org/10.1104/pp.17.01310

Pastinen, T. (2010) Genome-wide allele-specific analysis: insights into regulatory variation. *Nature Reviews, Genetics*, **11**, 533–538. Available from: https://doi.org/10.1038/nrg2815

Quinlan, A.R. & **Hall, I.M.** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842. Available from: https://doi.org/10.1093/bioinformatics/btq033

Roach, M.J., **Schmidt, S.A.** & **Borneman, A.R.** (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, **19**, 460. Available from: https://doi.org/10.1186/s12859-018-2485-7

Robinson, M.D., **McCarthy, D.J.** & **Smyth, G.K.** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140. Available from: https://doi.org/10.1093/bioinformatics/btp616

Shao, L., **Xing, F.**, **Xu, C.**, **Zhang, Q.**, **Che, J.**, **Wang, X.** ***et al.*** (2019) Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 5653–5658. Available from: https://doi.org/10.1073/pnas.1820513116

Simão, F.A., **Waterhouse, R.M.**, **Ioannidis, P.**, **Kriventseva, E.V.** & **Zdobnov, E.M.** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212. Available from: https://doi.org/10.1093/bioinformatics/btv351

Simsek, E., **Kilic, D.** & **Caliskan, O.** (2020) Phenotypic variation of fig genotypes (Ficus carica L.) in the eastern Mediterranean of Turkey. *Genetika*, **52**, 957–972. Available from: https://doi.org/10.2298/GENSR2003957S

Stern, J.L., **Paucek, R.D.**, **Huang, F.W.**, **Ghandi, M.**, **Nwumeh, R.**, **Costello, J.C.** ***et al.*** (2017) Allele-specific DNA methylation and its interplay with repressive histone marks at promoter-mutant TERT genes. *Cell Reports*, **21**, 3700–3707. Available from: https://doi.org/10.1016/j.celrep.2017.12.001

Sutton, M., **Roussel, B.**, **Chavez, D.J.** & **Malladi, A.** (2023) Synthesis of active cytokinins mediated by LONELY GUY is associated with cell production during early fruit growth in peach (Prunus persica (L.) Batsch). *Frontiers in Plant Science*, **14**, 1155755. Available from: https://doi.org/10.3389/fpls.2023.1155755

Takuno, S. & **Gaut, B.S.** (2013) Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1797–1802. Available from: https://doi.org/10.1073/pnas.1215380110

Usai, G., **Mascagni, F.**, **Giordani, T.**, **Vangelisti, A.**, **Bosi, E.**, **Zuccolo, A.** ***et al.*** (2020) Epigenetic patterns within the haplotype-phased fig (Ficus carica L.) genome. *The Plant Journal*, **102**, 600–614. Available from: https://doi.org/10.1111/tpj.14635

Valasiadis, D., **Kollaros, M.G.**, **Michailidis, M.**, **Polychroniadou, C.**, **Tanou, G.**, **Bazakos, C.** ***et al.*** (2024) Wide-characterization of high and low dry matter kiwifruit through spatiotemporal multi-omic approach. *Postharvest Biology and Technology*, **209**(112), 727. Available from: https://doi.org/10.1016/j.postharvbio.2023.112727

Vangelisti, A., **Simoni, S.**, **Usai, G.**, **Ventimiglia, M.**, **Natali, L.**, **Cavallini, A.** ***et al.*** (2021) LTR-retrotransposon dynamics in common fig (Ficus carica L.) genome. *BMC Plant Biology*, **21**, 221. Available from: https://doi.org/10.1186/s12870-021-02991-x

Vangelisti, A., **Zambrano, L.S.**, **Caruso, G.**, **Macheda, D.**, **Bernardi, R.**, **Usai, G.** ***et al.*** (2019) How an ancient, salt-tolerant fruit crop, Ficus carica L., copes with salinity: a transcriptome analysis. *Scientific Reports*, **9**, 2561. Available from: https://doi.org/10.1038/s41598-019-39114-4

Wang, H., **Beyene, G.**, **Zhai, J.**, **Feng, S.**, **Fahlgren, N.**, **Taylor, N.J.** ***et al.*** (2015) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 13729–13734. Available from: https://doi.org/10.1073/pnas.1519067112

Wang, L., **Jia, G.**, **Jiang, X.**, **Cao, S.**, **Chen, Z.J.** & **Song, Q.** (2021) Altered chromatin architecture and gene expression during polyploidization and domestication of soybean. *Plant Cell*, **33**, 1430–1446. Available from: https://doi.org/10.1093/plcell/koab081

Wang, L., **Xie, J.**, **Hu, J.**, **Lan, B.**, **You, C.**, **Li, F.** ***et al.*** (2018) Comparative epigenomics reveals evolution of duplicated genes in potato and tomato. *The Plant Journal*, **93**, 460–471. Available from: https://doi.org/10.1111/tpj.13790

Wang, Y., **Tang, H.**, **DeBarry, J.D.**, **Tan, X.**, **Li, J.**, **Wang, X.** ***et al.*** (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49. Available from: https://doi.org/10.1093/nar/gkr1293

Wang, Z., **Yuan, H.**, **Yan, J.** & **Liu, J.** (2024) Identification, characterization, and design of plant genome sequences using deep learning. *The Plant Journal*, **121**, e17190. Available from: https://doi.org/10.1111/tpj.17190

Williams, B.P., **Pignatta, D.**, **Henikoff, S.** & **Gehring, M.** (2015) Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genetics*, **11**, e1005142. Available from: https://doi.org/10.1371/journal.pgen.1005142

Xi, Y. & **Li, W.** (2009) BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics*, **10**, 232. Available from: https://doi.org/10.1186/1471-2105-10-232

Zhang, X., **Chen, S.**, **Shi, L.**, **Gong, D.**, **Zhang, S.**, **Zhao, Q.** ***et al.*** (2021) Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant Camellia sinensis. *Nature Genetics*, **53**, 1250–1259. Available from: https://doi.org/10.1038/s41588-021-00895-y

Zhang, Z. (2022) KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. *Genomics, Proteomics & Bioinformatics*, **20**, 536–540. Available from: https://doi.org/10.1016/j.gpb.2021.12.002

Zhao, X., **Chai, Y.** & **Liu, B.** (2007) Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Science*, **172**, 930–938. Available from: https://doi.org/10.1016/j.plantsci.2007.01.002

Zhou, Q., **Tang, D.**, **Huang, W.**, **Yang, Z.**, **Zhang, Y.**, **Hamilton, J.P.** ***et al.*** (2020) Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nature Genetics*, **52**, 1018–1023. Available from: https://doi.org/10.1038/s41588-020-0699-x