PaDSPD, a MYB-like transcription factor is linked with the diurnally synchronized protogynous dichogamy flowering type in avocado

Adil Khan[#], Shamseldeen Eltaher[#], Barbie Freeman and Gul Shad Ali[*]

United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Subtropical Horticulture Research Station (SHRS), Miami, FL 33158, USA
[#] Authors contributed equally: Adil Khan, Shamseldeen Eltaher
[*] Correspondence: Gul.Ali@usda.gov (Ali GS)

Abstract

Avocado (*Persea americana*) exhibits a distinctive flowering pattern known as diurnally synchronous protogynous dichogamy (DSPD), in which flowers of Type A cultivars open as female in the morning on day one and as male in the afternoon on day two, while those of Type B cultivars open as female in the afternoon on day one and as male in the morning on day two. The DSPD flowering system promotes cross-pollination and higher yield in some cultivars. However, the genetic and molecular mechanism controlling DSPD is not fully known. Genome-wide association studies of this panel resulted in identifying a precise 20-Kb genomic region on chromosome 10 associated with flowering Type A and B. Remarkably, all significant markers were delineated to a single gene encoding an R2R3 type MYB-like transcription factor, named here as *Persea americana diurnally synchronous protogynous dichogamy* (*PaDSPD*). These results were verified using bulked segregant analyses in a population segregating for flowering types. Analysis of marker data revealed that Type A always occurs as a heterozygous and Type B as a homozygous. Mendelian segregation analysis revealed that flowering type in avocado is inherited in a modified 2:1 (Type A : Type B) ratio, suggesting that one of the homozygous genotypes is likely lethal. For marker-assisted selection, competitive allele-specific PCR markers were developed, which successfully distinguished Type A and Type B accessions with perfect accuracy. These findings contribute to the genetic dissection of avocado flowering type and offer molecular markers for breeding programs aimed at improving avocado pollination efficiency and productivity.

Introduction

Avocado (*Persea americana*), popularly known as a super food, has gained widespread popularity among health-conscious consumers due to its nutrient-rich content and healthy fats[1]. However, production in the US has not kept pace with demand, necessitating heavy reliance on imports, which have reached almost 90% in the last decade. Factors such as adverse weather events, urbanization pressures, and rising production costs, particularly water costs and wages, have contributed to the decline in US avocado production[1,2]. While imports help sustain the US supply chain amidst growing demand, there is an urgent need to develop new avocado varieties with high productivity, expanded cultivation zones, and health benefits to meet future demand and ensure a stable supply.

Avocado yield is contingent upon successful pollination, which is influenced by a multitude of complex factors. Effective pollination significantly impacts fruit set and overall productivity[3]. Limited availability of pollen and the restricted time frame for the pollination process impose a notable challenge in achieving successful pollination in avocados. The unusual flowering habit of avocado trees is a major factor influencing inadequate pollination[4]. Avocado flowers are hermaphroditic, and they exhibit a unique flower opening and closing pattern, in which the flower first opens as female and later reopens as male—a phenomenon first described by Nirody[5] and subsequently confirmed by several investigators[6–11]. According to this pattern, flowering in avocados is categorized into two types, Type A and Type B[11]. Cultivars with Type A (e.g., 'Hass') open as functionally female in the morning of day one and as male in the afternoon of day two. Cultivars with Type B behavior (e.g., 'Fuerte') open as functionally female in the afternoon of day one and as male in the morning of day two. This evolutionary pattern promotes cross-pollination and hinders self-fertilization, as the female and male stages of Type A and B flowers overlap in different cultivars[12–15].

The role of diurnally synchronous protogynous dichogamy (DSPD) in avocado cultivation remains a subject of debate. In the DSPD flowering system, successful pollination requires strategic planting of Type A and Type B varieties. Since the male and female organs in Type A flowers mature at different times, isolated Type A trees may experience inadequate pollination, leading to lower fruit sets. Interplanting Type A and Type B varieties ensures overlapping reproductive phases, improving pollination efficiency and enhancing fruit production. Likewise, research has suggested that DSPD flowering system facilitates cross-pollination between avocado trees, contributing to increased yields in commercial orchards[16–18]. Studies conducted in Israel have demonstrated a significant correlation between outcrossing and fruit yield across various avocado cultivars[19]. In contrast, research in California has reported either a weak or negligible correlation between outcrossing and yield[20,21]. Garner et al.[22] conducted a study on outcrossing using ten six-year-old 'Hass' avocado trees in a commercial grove in California. Employing a single fully informative microsatellite marker, it was determined that outcrossing rates did not correlate with yield or alternate bearing. However, these studies only investigated the cultivar 'Hass' in California. This variety has been shown to efficiently self-pollinate[23], and the lack of correlation between outcrossing and yield may be affected by cultivar and/or environmental conditions.

Efforts have been made to determine the genetic basis of flowering type traits in avocado. For example, Lavi et al.[24] found that the avocado flowering type is controlled by several loci, each having several alleles. In contrast, Ashworth et al.[14] uncovered a single major quantitative trait locus (QTL) for this trait using a small segregating population of 'Gwen' × 'Fuerte' avocados. Using interval mapping (IM) and non-parametric Kruskal-Wallis (KW) test, they found several significant markers in a very large linkage interval between 8.803–59.658 cM (IM only) or 26.808–53.308 cM (both IM and KW) on linkage group 10, which were associated with flowering Type A and B. Recently, Solares et al.[15] conducted *Fst* analysis on 22 avocado accessions, including Type A (*n* = 13) and Type B (*n* = 9) representing Mexican, Guatemalan, and hybrid cultivars. They found *Fst* signals on multiple chromosomes, including chromosomes 6 and 10, which encompassed 466 genes, making it difficult to conclude the causal gene(s) responsible for flowering type. Potential reasons for low resolution for the signal might be that A vs B samples do not represent highly differentiated samples, and the sample number (22) is very small for statistically robust conclusions. Based on these reports, it appears that a single gene on chromosome 10 is likely responsible for controlling flowering type. However, the gene governing flowering type is not known and needs to be explored using a genetically diverse large collection and high-density markers.

Avocado trees raised from seeds typically produce flowers and fruit between the ages of three and five years or longer, so it is necessary to identify the flowering type early in the breeding cycle. Currently, the genetics of dichogamy has not been identified in any system, including avocado[15,25,26]. Identifying the flowering type gene and developing molecular markers would enable breeders to select Type A or Type B at the seedling stage, saving time and resources needed for raising and maintaining populations until fruiting. In a commercial production system, having both Type A and Type B cultivars intermixed would most likely maximize yield. Ideally, breeding new avocado cultivars with a pair of cultivars differing only in flowering types but otherwise similar would greatly maximize pollination efficiency and fruit yield while allowing for uniform orchard management and harvesting.

In this study, using GWAS on a genetically diverse avocado germplasm panel coupled with bulk segregant analysis (BSA) on a segregating mapping population, the aim was to identify gene(s) for flowering type in avocado. Additionally, the goal was to develop molecular markers for application in marker-assisted selection (MAS) in avocado breeding programs. A single gene (*MRB53_031466*) coding for an R2R3-type MYB-like transcription factor, named *Persea americana Diurnally Synchronous Protogynous Dichogamy* (*PaDSPD*), was found to control flowering type in avocado. Moreover, phylogenetic and functional analyses confirmed that *PaDSPD* is a homolog of *EMISSION OF BENZENOIDS I* (*EOBI*) transcription factor, which plays a key role in the development of reproductive organs[27–29]. Competitive allele-specific PCR markers were also developed and validated, successfully distinguishing Type A and Type B accessions in a germplasm collection as well as in breeding populations resulting from the crosses of four genetically diverse parental cultivars: 'Simmonds', 'Tonnage', 'Hass', and 'Bacon'. These findings enhanced understanding of the genetic and molecular basis of flowering types in avocados and will lead to the development of new varieties with the desirable flowering type, enabling better orchard management and higher yields.

Materials and methods

Germplasm collection, mapping populations, and phenotyping

The *Persea americana* germplasm population used in the current study has been previously described[30–32]. Briefly, a total of 153 accessions were used, which are maintained at the United States Department of Agriculture, Agricultural Research Service, Subtropical Horticulture Research Station, Miami, Florida. In addition to the main germplasm collection, the Florida Avocado Mapping Population (FLAMP), which is segregating for flowering type and is maintained at this station[33], was also used for QTL mapping of the flowering type using bulked segregant analysis (BSA). The FLAMP consists of two reciprocal *Persea americana* populations: 'Simmonds' × 'Tonnage' and 'Tonnage' × 'Simmonds'. For each cross, trees were phenotyped for flowering types, and two bulks, each consisting of Type A or Type B, were created by pooling equal numbers of leaf disks from 50 independent trees from each population. In addition to the FLAMP, the trees in the California Avocado Mapping Population (CAMP)[6], which consists of 'Bacon' × 'Hass' cross, were phenotyped for flowering types.

The avocado accessions and segregating populations were scored for flowering type throughout the day, particularly in the morning and evening. At least 50 flowers were scored for flowering type. In addition, two varieties, 'Melendez 2' (Type A) and 'Tonnage' (Type B), were selected for detailed floral microscopy studies. Images of various flowering stages and floral organs were captured using a Leica stereomicroscope (LEICA M205 C, Deerfield, IL, USA). Both varieties were situated in proximity at the research station to ensure consistent environmental conditions.

Whole genome resequencing and SNP calling

The whole genome sequencing was performed by LGC RAPiD Genomics (www.rapid-genomics.com) using the Illumina 150x PE platform. Raw fastq read data were trimmed using FASTP standard trimming parameters with additional options to detect paired end adapters and trim poly G tails: fastp --in1 fastqR1 --in2 fastqR2 --detect_adapter_for_pe --length_required 50 --trim_poly_g[34]. The reads were further processed using Trimmomatic with the following settings: trimmomatic PE ILLUMINACLIP: HPC_TRIMMOMATIC_ADAPTER/TruSeq3-PE-2.fa:2:30:10:8:true SLIDINGWINDOW:10:20 MINLEN:50[35]. The clean reads were then mapped using Burrows-Wheeler Aligner (BWA)[36] to the *Persea americana* Cv. 'Hass' reference genome GCA_029852735.1_ASM2985273v1 v and the duplicates were marked with samtools[37]. Variants, including single-nucleotide polymorphisms (SNPs) and indels, were called using DeepVariant V1.4.0 (https://github.com/google/deepvariant) for each accession separately, and the resulting gVCF files were combined into a single VCF file using GLnexus v1.2.7[38]. The resulting VCF file was filtered with vcftools using the following parameters: MAF > 0.05, missing call rate < 0.1, and quality > 30. To analyze population structure, SNPs were pruned for linkage disequilibrium (LD) using PLINK[39] with the following parameters: --indep-pairwise 100 10 0.5. This pruning process resulted in a refined set of 1,072,934 SNPs. Indels were extracted from the VCF file using vcftools[40].

Population structure, phylogenetic, and principal component analyses

To investigate the population structure of 153 individuals, population structure and phylogenetic analyses were performed using a

filtered and LD-pruned 1,072,934 SNPs. Software fastStructure/1.0 was used for clustering individuals with $K = 1$ to 10 to estimate the optimal number of clusters (K)[41]. Cross-validation error values for each cluster were plotted using the ggplot package in R. The population structure was visualized to subsequently assign accessions to avocado races based on the clustering results. Relationships among races were inferred by constructing a dendrogram using TASSEL 5[42]. Additionally, to explore genetic variation among populations, a Principal Component Analysis (PCA) was performed in PLINK 1.9, and the resulting components were visualized in R (version 4.4.1) using ggplot package.

Genome-wide association studies for mapping the flowering type gene in avocado

Genome-wide association study (GWAS) was conducted using 735,030 biallelic SNPs employing EMMAX (Efficient Mixed Model Association Expedited) to account for population structure and cryptic relatedness. To control for cryptic relatedness, a kinship matrix derived from identity-by-state (IBS) alleles, which considers shared alleles regardless of ancestral origin, was incorporated and calculated using the SNP and Variation Suite (SVS) V8 (Golden Helix, Inc.; www.goldenhelix.com). Results were validated using additional models: Generalized Linear Model (GLM) incorporating principal components (PC), and a Mixed Linear Model (MLM) integrating both PC and the kinship matrix (K) using rMVP (https://github.com/xiaolei-lab/rMVP), an R package designed for memory-efficient, visualization-enhanced, and parallel-accelerated GWAS. A stringent Bonferroni correction was applied with a threshold of $-\log_{10}(0.05/n)$, where n represents the number of independent SNPs. The results were visualized using CMplot (https://github.com/YinLiLin/CMplot).

Bulked segregant analysis for mapping the flowering type gene in avocado

To identify genomic regions associated with flowering type (Type A vs Type B) in avocado, Bulked Segregant Analysis (BSA) independently was also conducted independently on two reciprocal populations: 'Simmonds' × 'Tonnage' and 'Tonnage' × 'Simmonds', where 'Simmonds' was Type A and 'Tonnage' was Type B. The genotyping data for the parents ('Simmonds' and 'Tonnage') and the Type A and Type B bulks were merged into a single VCF file. Similarly, for the 'Tonnage' × 'Simmonds' reciprocal population, the VCF files for the parents ('Tonnage' and 'Simmonds') and the Type A and Type B bulks were merged into another VCF file. Each merged VCF file was filtered using vcftools to retain SNPs with a minor allele frequency (MAF) > 0.05, maximum missing rate = 0.9, minimum depth = 7, maximum depth = 100, and quality = 50. Subsequently, both filtered datasets were analyzed using the OcBSA pipeline, following the recommended scripts and parameters[43]. The window size for smoothing was set to 200 kb (–w 200,000) to detect genomic regions associated with flowering type differences between the bulks.

Identifying and functional annotation of candidate genes for flowering type

All significant GWAS and BSA signals for flowering type were located within or around a MYB-like transcription (PaDSPD). A BLAST search was performed using the PaDSPD protein sequence, and all matching sequences from various plant species were retrieved using the following criteria: Query Cover > 70%, Percent Identity > 65%, and E-value < 0.001. Given the critical role of MYB transcription factors in floral development and regulation, MYB

gene family members were specifically analyzed using multiple sequence alignments and phylogenetic analysis to classify these MYB genes into subgroups associated with floral initiation, reproductive organ development, and hormone-mediated flowering regulation. The MYB family transcription factor (TF) protein sequences of *Arabidopsis thaliana*, *Solanum lycopersicum*, *Citrus clementina*, *Petunia axillaris*, *Cinnamomum micranthum*, *Musa acuminata*, and *Phoenix dactylifera* were obtained from the Plant Transcription Factor Database v5.0 (PlantTFDB). Phylogenetic analysis was conducted using a model implemented in MEGA11[44]. Sequences were aligned using ClustalW2[45]. The resulting phylogenetic tree was visualized using the web-based tool iTOL (Interactive Tree of Life, https://itol.embl.de). Furthermore, functional analysis of *MRB53_031466* was carried out using gene set enrichment analysis using ShinyGo 0.82 database (http://bioinformatics.sdstate.edu/go/) to infer biological relevance.

Development, and validation of competitive allele-specific markers

To develop PCR Allele Competitive Extension (PACE®) SNP markers, multiple strategically positioned high-confidence SNPs encompassing the *PaDSPD* gene (between 41,389,855–41,410,729 on CM056818.1 chromosome) were selected. The PACE primers were designed for the selected SNPs using the 3CR Bioscience PACE genotyping platform (Supplementary Table S1). Primer sets were optimized for specificity and efficiency via gradient PCR, followed by validation using fluorescence-based endpoint genotyping in 48 Type A and 45 Type B flowering accessions. The developed PACE SNP markers were tested in F₁ mapping populations derived from reciprocal crosses between 'Simmonds' (Type A) and 'Tonnage' (Type B)[33], and 'Hass' (Type A) and 'Bacon' (Type B) (Schnell et al., 2009). DNA was extracted from young leaf tissue using the Omega DNA extraction kit in a 96-well platform and quantified via Nanodrop and Qubit fluorometry. PACE reactions were performed in a 96-well format using the Applied Biosystems QuantStudio Dx Real-Time PCR system. Genotypic data were analyzed to determine Mendelian segregation patterns, validate marker-trait associations, and assess marker predictive accuracy for flowering type. Genotypic and phenotypic concordance was assessed using χ^2 goodness-of-fit tests. Marker efficacy was evaluated based on genotypic call rates and discriminatory power between Type A and Type B phenotypes.

Results

Whole-genome resequencing revealed high-density markers on the avocado genome

A genetically diverse avocado collection consisting of 153 accessions and avocado mapping populations was sequenced using approximately 25× depth (Supplementary Table S2). A total of 11,439,248 variants, comprising 10,863,184 single-nucleotide polymorphisms (SNPs) and 623,953 insertions and deletions (indels), were identified across 153 accessions. High-quality SNPs and indels were randomly distributed across all 12 chromosomes, but their densities were higher towards the chromosome ends and lower in the centromeric regions, likely due to the repetitive and heterochromatic nature of the centromere (Supplementary Fig. S1a, S1b), with an average 13/kb SNP density. Chromosome 2 exhibited the highest number of SNPs (1,473,284), whereas chromosome 12 had the lowest count (626,141) (Supplementary Fig. S1c). The transition-to-transversion ratio in the analyzed SNP dataset was 2.13,

indicating a higher prevalence of transition mutations compared to transversions. Deletion sizes ranged from −1 to −49, while insertion sizes varied from 1 to 150 (Supplementary Fig. S1d). Both deletions and insertions exhibited a higher frequency for shorter variants, whereas longer insertions and deletions were observed in lower numbers (Supplementary Fig. S1d).

Geographic distribution shapes avocado population structure

To delve deeper into the genetic relationships among the diverse avocado accessions, 1,072,934 LD-pruned SNPs were utilized for population structure and principal component analyses (Fig. 1a–d). Admixture analysis was conducted using cluster ranges $K = 2$ to 10. The results identified $K = 4$ and $K = 6$ as the optimal numbers of clusters, effectively capturing the variation within the dataset (Fig. 1a, b). As illustrated in Fig. 1a, at $K = 3$, the analysis revealed three distinct subpopulations, corresponding to the major geographical origins of avocado (i.e., Mexican (Mex.), Guatemalan (Guat.), and West Indian (W.I.) races). Membership probabilities are presented in Supplementary Table S2. Cluster 1 (Q1) comprises 19 accessions with membership probabilities of 1.0, all belonging to the Mexican

race. Among these, accession 'La Piscina' with a previously unknown race was assigned the Mexican race. Likewise, accessions 'CRC 195-36', 'Yama-423', and 'Teague' were successfully assigned to the Mex. × Guat. hybrid group, while 'Vero Beach No. 1' was assigned to the Mex. × W.I. group. These accessions were previously categorized as complex hybrids. Cluster 2 (Q2) contained accessions belonging to the Guatemalan race with membership probabilities of 1 for most of the accessions. Additionally, Q2 includes accessions that are hybrids between Guat. × W.I. and Guat. × Mex. Furthermore, it was found that the accession 'San Rafael', which was previously assigned to the Mexican group, belongs to the Guat. × W.I. group, suggesting it might have been mislabeled. Moreover, 'Vero Beach No. 15' was identified as a hybrid among Mex. × Guat. × W.I., which was previously assigned as a complex hybrid. Cluster 3 (Q3) consists of accessions with W.I. genetic background. This cluster also includes accessions that are hybrids, such as Guat. × W.I., as well as individuals that are hybrids among all three major genetic backgrounds: Mex. × Guat. × W.I. Overall, based on the races assigned in this study, the accessions showed a higher proportion of West Indian genetic background (16%) compared to Mexican (8%) and Guatemalan (5%). Furthermore, a significant number of accessions were hybrids, with
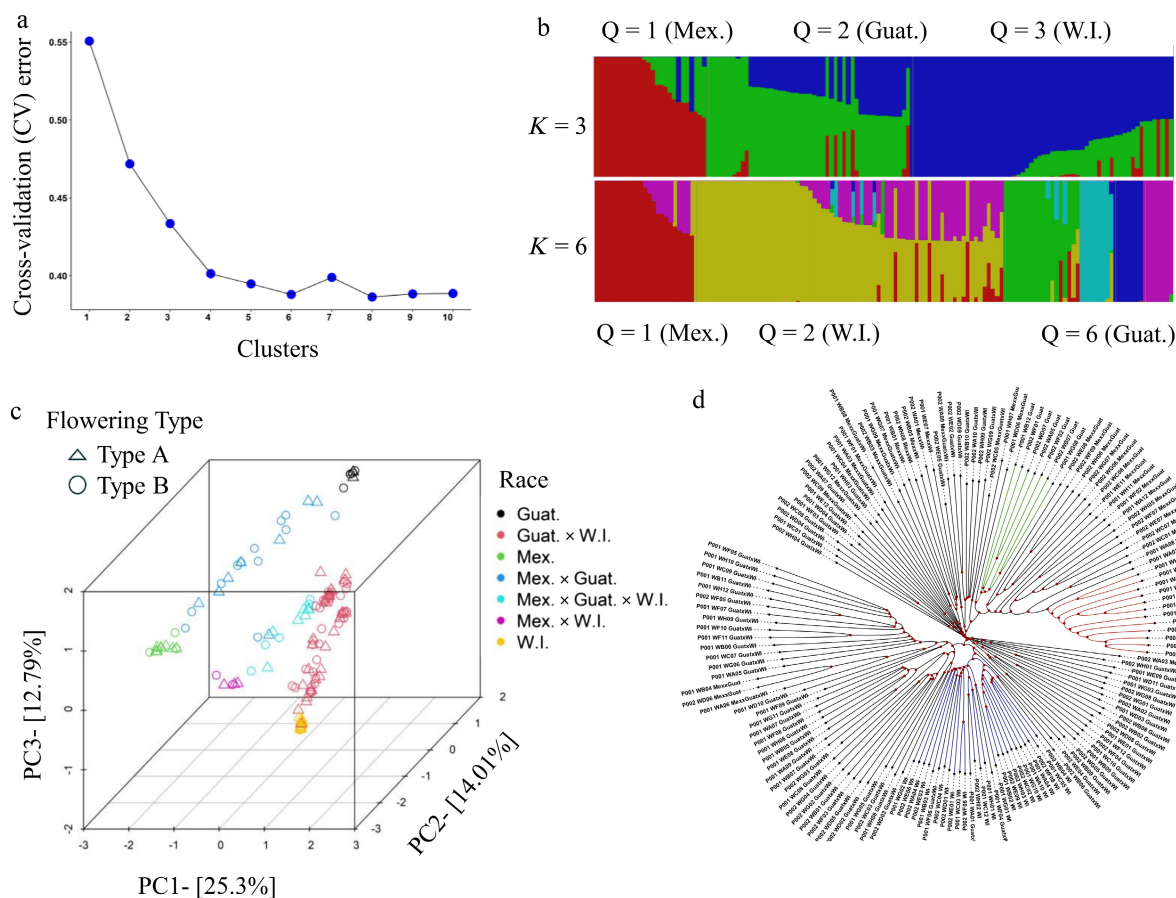

Fig. 1 Population structure analysis of 153 *Persea americana* accessions. (a) Cross-validation error analysis for clustering with hypothetical *K*-values ranging from 1 to 10. The *x*-axis represents *K*-values (1–10), while the *y*-axis represents cross-validation error values. The rate of decrease slows after $K = 4$. (b) Admixture analysis showing population structure for assumed group numbers $K = 3$ and 6. Each accession is represented as a thin vertical segment, with colors indicating population proportions. Each subgroup is designated by a unique color. In $K = 3$, accessions with red, green, and blue represent Mexican, Guatemalan, and West Indian races, respectively, and combinations of colors represent hybrid lines. (c) Scatter plot of the first three principal components (PC1, PC2, and PC3). Each dot represents an accession, with symbol shapes indicating flowering types and colors representing different races. (d) Dendrogram of 153 avocado accessions based on whole-genome SNP data (1,072,934 SNPs). Accessions with red, green, and blue represent Mexican, Guatemalan, and West Indian races, respectively. Each accession label is followed by its corresponding race, and the labels can be matched with the IDs and cultivar names listed in Supplementary Table S2.

the largest group being hybrids between W.I. and Guat. genetic backgrounds (46%), followed by Mex. × Guat. hybrids (14%) (Supplementary Fig. S2a). This study also successfully assigned highly confident genetic races to individuals whose genetic backgrounds were previously unassigned or incorrectly assigned.

Principal Component Analysis (PCA) and hierarchical clustering based on a dendrogram were performed to analyze the population structure of avocado accessions used in this study. The PCA plot (Fig. 1c), like the structure analysis, grouped avocado genotypes into three clusters, reflecting their ecotypes of origin: Guatemalan, Mexican, and West Indian. These results indicate that hybrids between Mexican and Guatemalan races exhibit a genetic composition that is uniformly distributed between these two races. In contrast, hybrids between Guatemalan and West Indian races show a genetic composition skewed more toward the West Indian background, suggesting a stronger influence from the West Indian race (Supplementary Fig. S2a). Similarly, hybrids among all three races demonstrate a genetic composition weighted more heavily toward the West Indian background compared to Mexican and Guatemalan genetic contributions (Fig. 1c). The dendrogram analysis (Fig. 1d) corroborated these findings, revealing a similar distribution pattern. The analysis highlighted considerable overlap among clusters, emphasizing the hybrid nature of many cultivars as inferred through population structure analysis. Furthermore, the distribution of accessions suggests that Mexican and Guatemalan genetic backgrounds are genetically closer to each other than their relationship with West Indian genetic backgrounds.

Genome-wide association studies revealed precise markers associated with a transcription factor controlling flowering type in avocado

To identify markers associated with flowering type, GWAS was conducted using high-density markers on a genetically diverse population segregating for flowering type. Figure 2a illustrates phenotypic differences between Type A ('Melendez 2') and Type B ('Tonnage') cultivars. Their floral opening patterns at different times of the day were recorded. As expected, both types exhibited protogyny opening as female first on day 1, switching to male on day 2. Type A flowers ('Tonnage') opened in their female phase in the morning around 10:30 AM, whereas Type B flowers ('Melendez 2') opened later in the afternoon at approximately 2:00 PM. After this stage, both flower types closed, with Type A closing in the afternoon and Type B in the late evening. On the following day, Type A flowers transitioned to their male phase around 4:00 PM, while Type B flowers did so earlier at approximately 9:00 AM. Both types then closed later in the day. In the subsequent days, the flowers remained closed and gradually developed a brownish coloration. The flowering patterns were recorded for all the avocado accessions in the GWAS panel. The data showed random distribution across the population, as illustrated in Supplementary Fig. S2b.

To account for population structure effect, genome-wide association analysis of flowering type was performed independently using three models: EMAXX, GLM, and MLM. Consistently, analyses across all models revealed a significant GWAS signal for flowering type on chromosome 10 (CM056818.1) (Fig. 2b; Supplementary Fig. S3a, S3b). Since the avocado accessions used in the study exhibited population structuring (Fig. 1), MLM and EMMAX models were used in addition to GLM models to reduce false positive signals. As shown in Fig. 2b (EMMAX model) and Supplementary Fig. S3 (GLM and MLM models), inflation of $-\log_{10}(p)$ was reduced by almost half, from > 50 (GLM model) to < 25 (EMMAX and MLM models). Although confounding effects due to population structure were effectively controlled by applying MLM and EMMAX models, the significance of results remained well above the threshold p-values (Fig. 2b; Supplementary Fig. S3). Notably, most significant markers clustered within a single gene, including its promoter, coding region, and 3' regulatory region, providing strong evidence for the GWAS findings.

Statistical analyses identified an associated chromosomal region encompassing a 20,874 bp haplotype block (CM056818.1: 41,389,855–41,410,729; Supplementary Table S3). Following pruning and filtering of the data, within this block, nine markers were significantly associated with flowering type ($p < 6.59 \times 10^{-8}$; FDR < 5.372936×10^{-3}; Supplementary Table S3). Additional markers were included in an expanded region (CM056818.1: 41,325,077–41,524,944; approximately 200 kb) to enhance marker identification associated with flowering type (Fig. 2b, c; Supplementary Table S4), which resulted in the identification of 436 markers associated with flowering type. Furthermore, haplotype analysis of this QTL revealed two major haploblocks controlling Type A and Type B flowering type respectively, as shown in Fig. 2d. GWAS was also conducted using 553,332 indels, which resulted in the identification of the same genomic region that was determined using SNPs (Fig. 2e). These analyses revealed 12 insertions and 17 deletions significantly associated with flowering type ($p < 4.60 \times 10^{-13}$; FDR < 8.79×10^{-9}) within the haplotype block (CM056818.1:41,385,195–41,410,847; 25,652 bp) (Fig. 2e, f; Supplementary Table S5).

The GWAS peaks were further validated by observing a high fixation index (F_{ST}) of approximately 0.5 within this region (Fig. 2g). An analysis of population differentiation (F_{ST}) of Type A and Type B flower types revealed a strong differential signal (0.3–0.45) around the same region where the GWAS signal was detected (Fig. 2g). This indicates strong genetic differentiation and potential selective sweeps in this region, where divergent selection pressures might have acted to fix alternate alleles for flowering Type A and Type B. The co-localization of high F_{ST} signal and trait-association signal suggests that flowering type might have been the target of positive selection during avocado evolution to perhaps encourage cross-pollination, genetic diversity, and abundant seed production for survival.

Remarkably, only one gene (MRB53_031466) coincided with significant markers (Fig. 2c, f). This gene, which codes for an R2R3-type MYB-like transcription factor, was named *Persea americana Diurnally Synchronous Protogynous Dichogamy* (*PaDSPD*) to be consistent with previous phenological descriptions of flowering type[6]. The GWAS signal was sharply elevated with $-\log_{10}(p)$ values > 8.0 (4.918, 0.05/4141) around this gene from 7,816 bp upstream from the start codon in the promoter region and about 5,153 bp downstream from the stop codon. The rapid decline of GWAS signals beyond the MRB53_031466 gene (Fig. 2c, f) strongly suggests that this gene is controlling flowering type in avocados.

Markers, including both SNPs and SVs associated with flowering type, were found to be consistently heterozygous in Type A varieties and homozygous in Type B varieties (Supplementary Fig. S4a). The SNPs of the minor allele, named as F^M for 'female in the morning', were significantly less common (frequency of ~0.1–0.2) compared to the ones in the major allele, F^E for 'female in the evening' (frequency ~0.8–0.9) (Supplementary Fig. S4; Supplementary Table S5). Similarly, the genotype ratio showed $F^M F^E$ genotypes occurring twice as frequently as $F^E F^E$ genotypes, while $F^M F^M$ genotypes were entirely absent, suggesting that it is lethal. These results are consistent with the Mendelian genetic analysis of mapping population segregating for flowering type as described below in the next section.
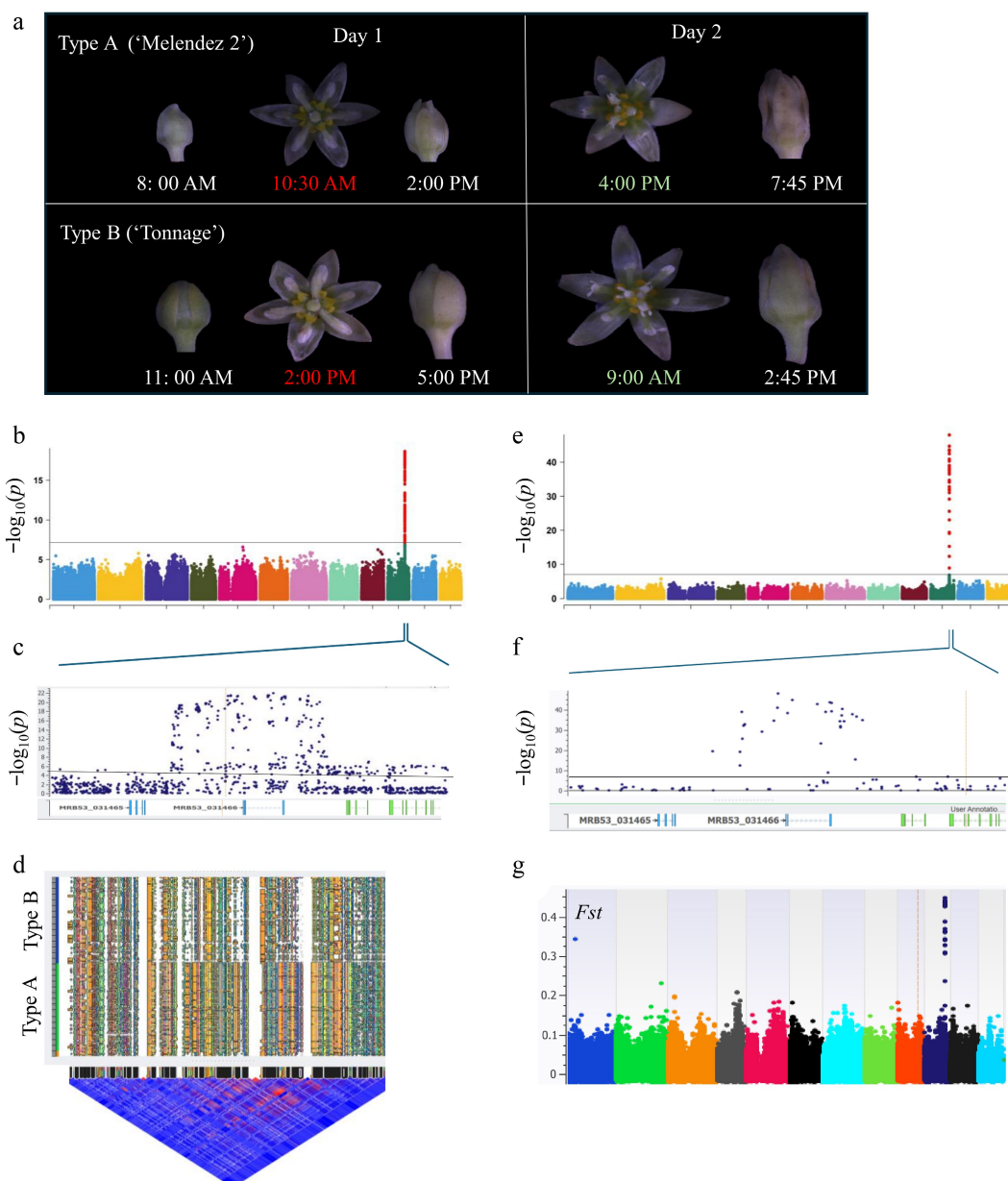
Fig. 2 Genome-wide association study (GWAS) of the flowering type in avocado germplasm using SNP and SV markers. (a) Floral development of Type A ('Melendez 2') and Type B ('Tonnage') from bud to full bloom till flower closing, observed at different times of the day. The photographs illustrate the distinct blooming patterns of both varieties. Both Type A and Type B open as functionally female on the first day with different timing (highlighted in red), followed by functionally male stages on the second day (highlighted in light green). After pollination, the sepals get brown, indicating the end of the pollination period. (b), (e) Manhattan plots illustrating GWAS results for flowering type using (b) 735,030 SNPs and (e) 553,332 structural variants (SV) using the EMAXX model; associations were verified using additional GWAS models (Supplementary Fig. S3). The y-axis represents $-\log_{10}(p)$ values. The horizontal line represents 5% Bonferroni correction significance threshold. (c), (f) Zoomed-in genomic regions representing the peak in panels B and E with gene models displayed below the graph. Blue dots represent markers. All significant markers are primarily localized around the *MRB53_031466* gene encoding R2R3-type Myb-like transcription factor. (d) Haplotype blocks and Linkage disequilibrium (LD) plots of the flowering type locus. The upper part of the figure illustrates different haplotype blocks in Type A and Type B. The lower part shows an LD heat map of SNPs with red regions representing high LD between adjacent SNPs, while blue regions represent low LD. (g) Fixation Index (F_{ST}) plot showcasing genetic differentiation between Type A and Type B avocado accessions. Elevated Fst values indicate regions under potential positive selection, aligning with GWAS signals.

Bulked segregant analysis and genetic analysis validate a single gene on chromosome 10 as controlling flowering type in avocado

Consistent with GWAS analyses, bulk segregant analyses (BSA) conducted on both reciprocal avocado populations – 'Tonnage' × 'Simmonds' and 'Simmonds' × 'Tonnage' – revealed a prominent peak on chromosome 10 (Fig. 3a, b). Each cross involved phenotypically distinct Type A and Type B flowering bulks. Independent analyses for each population showed statistically significant signals on chromosome 10, the genomic interval, which strongly supports the presence of a robust quantitative trait locus (QTL) associated with flowering type. Because the genetic background of these populations is essentially the same, but paternal and maternal parents are reversed, this consistent signal rules out cytoplasmic
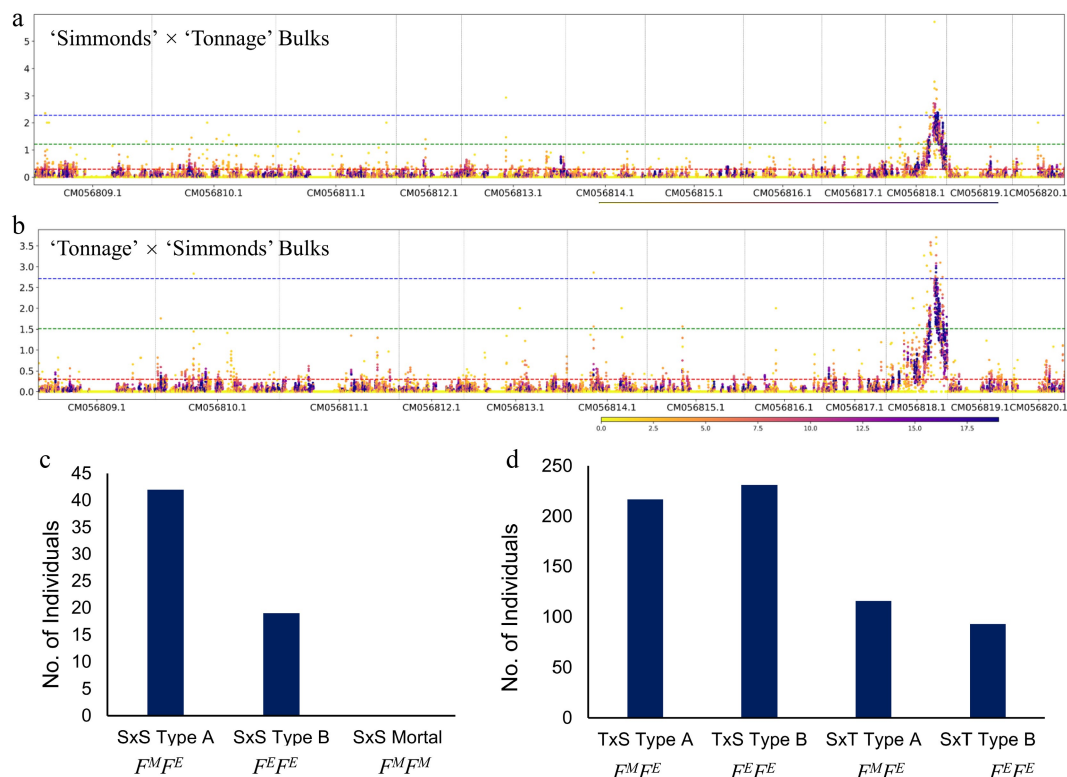
Fig. 3 Bulk segregant analysis maps flowering type to the same loci as GWAS signals on chromosome 10. (a), (b) Bulk segregant analyses (BSA) conducted on both reciprocal avocado populations. (a) 'Simmonds' × 'Tonnage' ($n = 50$), and (b) 'Tonnage' × 'Simmonds' ($n = 50$). Segregation ratio of the Type A ($F^M F^E$) and Type B ($F^E F^E$) genotype in segregating populations: (c) Cv 'Simmonds' selfed, and (d) 'Tonnage' × 'Simmonds' and reciprocal ('Simmonds' × 'Tonnage').

inheritance. These analyses further suggest that this locus is under Mendelian nuclear control and reaffirm evidence for a major-effect locus on chromosome 10 that differentiates flowering types. Importantly, the peak alignment between the 'Tonnage' × 'Simmonds' and 'Simmonds' × 'Tonnage' crosses reinforces the biological relevance of this region, suggesting that it harbors the *PaDSPD* gene as shown in the GWAS results.

In the FLAMP, 61 trees of selfed 'Simmonds' were available, which allowed us to test single gene Mendelian inheritance of the *PaDSPD* gene. Based on the GWAS analyses, where all Type A accessions were found to be heterozygous for all significantly associated SNPs, and all Type B accessions were homozygous, a hypothesis was proposed that Type A alleles are heterozygous, designated as $F^M F^E$, representing Female first in the Morning and Female first in the Evening, and Type B are either homozygous $F^E F^E$ or $F^M F^M$. Since 'Simmonds' is Type A and is heterozygous for all significantly associated alleles ($F^M F^E$), a selfed 'Simmonds' population, according to a one gene hypothesis, should segregate in a 1:2:1 ($F^M F^M : F^M F^E : F^E F^E$) ratio. As is shown in Fig. 3c, selfed 'Simmonds' progeny segregated in a modified 2:1 (Type A : Type B) ratio ($n = 61$; $\chi^2 = 0.717$, Table 1), suggesting that flowering type in avocado is controlled by a single gene with $F^M F^E$ as Type A and $F^E F^E$ as Type B and $F^M F^M$ homozygous genotype as lethal; the expected 3:1 phenotypic segregation ratio was not met ($\chi^2 = 2.78^{E-11}$, Table 1), indicating a deviation from classical mendelian inheritance. Likewise, in both reciprocal avocado populations – 'Tonnage' × 'Simmonds' and 'Simmonds' × 'Tonnage', where 'Simmonds' is heterozygous ($F^M F^E$) and 'Tonnage' is homozygous ($F^E F^E$), a 1:1 (Type A : Type B; $F^M F^E : F^E F^E$) segregation ratio is expected. Indeed, a 1:1 (Type A : Type B) ratio was observed for both the 'Tonnage' × 'Simmonds' population ($n = 448$; $\chi^2 = 0.508$) and the 'Simmonds' × 'Tonnage' ($n =$

209; $\chi^2 = 0.111$) (Fig. 3d; Table 2). In conclusion, the flowering type in avocado is controlled by one gene, with Type A always occurring as a heterozygous and Type B as a homozygous, with one of the homozygous genotypes being lethal.

Phylogenetic and functional annotation of the *PaDSPD* MYB-like transcription factor

Homology searches and phylogenetic analyses showed that *PaDSPD* is a close homolog of (*Cinnamomum micranthum*, *Magnolia sinica*) EMISSION OF BENZENOIDS I (*EOBI*) and *EOBII* of *Aristolochia*. Phylogenetic results based on BLAST analysis from NCBI further confirm its close relationship with *EOBI*, *EOBII*, and MYB-related protein 305-like (Fig. 4a; Supplementary Table S6). To gain deeper insight into the genetic relationship of *MRB53_031466*, a phylogenetic tree was developed incorporating MYB-related genes from *Arabidopsis thaliana*, *Cinnamomum micranthum*, *Citrus clementina*, *Musa acuminata*, *Petunia axillaris*, *Phoenix dactylifera*, and *Solanum lycopersicum* (Fig. 4b). The results suggest that *PaDSPD* (*MRB53_*

Table 1. Segregation of *Persea americana* cv. 'Simmonds' selfed population for flowering type.

	Observed	Expected under a 2:1 expected ratio assuming one homozygotes is lethal	Expected under a 3:1 expected ratio
SxS Type A ($F^M F^E$)[1]	42	40.67	45.75
SxS Type B ($F^E F^E$)	19	20.33	15.25
Mortal ($F^M F^M$)	0		
Total	61		
χ^2 test		0.717	0.268

Table 2. Segregation of flowering type in the Florida Avocado Mapping Populations (FLAMP) resulting from a cross between 'Tonnage' × 'Simmonds' and its reciprocal cross 'Simmonds' × 'Tonnage' cultivars.

	Observed	Expected under a 1:1 expected ratio	Expected under a 3:1 expected ratio
'Tonnage' × 'Simmonds' Type A ($F^M F^E$)	217	224	336
'Tonnage' × 'Simmonds' Type B ($F^E F^E$)	231	224	112
Total	448		
χ^2 test		0.508	1.50865^{E-38}
Reciprocal cross			
'Simmonds' × 'Tonnage' Type A ($F^M F^E$)	116	104.5	156.75
'Simmonds' × 'Tonnage' Type B ($F^E F^E$)	93	104.5	52.25
Total	209		
χ^2 test		0.112	7.53547^{E-11}

1, Genotypes of Type A and Type B flowering types; F^M = 'Female in the morning' allele, and F^E = 'Female in the evening' allele of the *Persea americana Diurnally Synchronous Protogynous Dichogamy* (*PaDSPD*) gene.

031466) forms a distinct clade, including MYB genes such as *Myb24*, *Myb21*, and *myb-related 305-like*, all of which belong to subgroup 19 (SG19)[27,46] within the R2R3-MYB transcription factor family (Fig. 4c). The alignment of protein sequences of *Persea americana*, *Arabidopsis thaliana*, *Cinnamomum micranthum*, *Citrus clementina*, *Musa acuminata*, *Petunia axillaris*, *Phoenix dactylifera*, and *Solanum lycopersicum* can be seen in Supplementary Fig. S5. As expected, the R2R3-MYB domain is more conserved compared to the TAD domain, suggesting a more crucial functional role of the R2R3-MYB domain and species-specific function for the TAD across these species. Comparing the PaDSPD type A allele to the type B allele revealed a total of fifteen non-synonymous mutations, with a D65G mutation located in the loop of the R2R3 domain, and the rest were mostly localized to the C-terminal region of the protein, with S183F and S184H mutations localized to the TAD domain (Supplementary Fig. S5).

Since the *PaDSPD* gene shows close homology to the *Arabidopsis* At3G27810 and At5G40350 genes, these genes were used as a reference to infer the potential function of *PaDSPD* (Fig. 4d). Gene enrichment analysis revealed their dominant role in reproductive organ development in plants. Additionally, the analyses indicated involvement in hormone-mediated signaling pathways, including the jasmonic acid-mediated signaling pathway. Functional analyses using *Arabidopsis* homologous genes demonstrated that *PaDSPD* is primarily involved in anther filament development, suggesting an evolutionarily conserved mechanism for the function of *PaDSPD*. Furthermore, the *Arabidopsis* ePlant database (https://bar.utoronto.ca/) provides extensive gene expression data related to various plant tissues and organ development, including flowers. Using this resource, the expression of the *Arabidopsis* homologs, *AtMYB21/AtMYB3* of the *PaDSPD* gene was mentioned. As shown in Supplementary Fig. S6a, *Arabidopsis* homologs of this gene are expressed in stamens and petals, and Supplementary Fig. S6b shows its interaction with other genes. The co-expressed genes are involved in responses to hormonal signaling, particularly within the jasmonic acid-mediated signaling pathway.

Development of competitive allele-specific markers

To facilitate marker-assisted selection for flowering type in avocado, competitive allele-specific PCR-based genotyping assays were developed based on the PCR Allele Competitive Extension (PACE) chemistry. Multiple highly specific markers encompassing the entire *PaDSPD* locus (41,389,707–41,411,647) were designed targeting the promoter, CDS (41,397,672–41,405,576) and 3' regulatory region of *PaDSPD* gene (Supplementary Table S1). Since all significant SNPs were localized in the promoter, terminator and coding regions of a single gene, *PaDSPD*, these markers were designed to comprehensively cover the gene body and its promoter region. In total, seven markers were screened, of which four markers worked successfully. To assess marker accuracy, PACE PCR was conducted on individual accessions (*n* = 47) representing different races from the diversity germplasm with known flowering types (Type A = 26; Type B = 21). Of the seven markers, four markers (M1–41392642_T_C; –5,030 bp promoter, M2–41398033_A_G; +361 bp from start codon causing a D65G amino acid change; M3–41398041_C_A; +369 bp from start codon; M4–41405555_CAT_; +7,883 bp from start codon causing three nucleotide deletion) reliably identified flowering type perfectly (Fig. 5a–c; Supplementary Fig. S7; Supplementary Table S7). Since these markers successfully identified flowering types with high accuracy in genetically diverse accessions representing Mex., W.I., Guat. and hybrid avocado races, they would be applicable in diverse breeding programs using complex crosses involving diverse parents. To validate and demonstrate the practical application of these markers in breeding, three avocado hybrid populations segregating for flowering type were screened (Simmonds' × 'Tonnage'; *n* = 18, 'Tonnage' × 'Simmonds'; *n* = 18, and 'Hass' × 'Bacon'; *n* = 10)[33]. In all three populations, all markers identified flowering type with 100% accuracy (Supplementary Table S7). Notably, the marker analyses confirmed the absence of an $F^M F^M$ homozygous genotype in the diversity panel as well as the breeding populations, suggesting that this allelic combination is lethal. Furthermore, these analyses also validated the Mendelian inheritance of the *PaDSPD* gene, with Type A consistently exhibiting heterozygosity and Type B homozygosity as described above.

Discussion

Population structure analyses reveal ecogeographic origins of avocado accessions

Avocados are native to Mesoamerica and are cultivated in tropical, subtropical, and Mediterranean climates[1,30,47]. Population structure analyses in this study, based on high-density markers, supported previous studies on the ecological adaptations and geographic origins of avocado races based on smaller sets of markers.[31,48,49]. The genetic asymmetry exhibited by Guat. × W. I. hybrids in relation to their West Indian ancestry align with historical patterns of introgression and may reflect conferring adaptive benefits or selective breeding biases[50]. The genetic predominance of W.I. ancestry in tri-racial hybrids significantly influences this extreme introgression pattern, likely reflecting adaptation to the environmental conditions associated with tropical agricultural practices[51]. The genetic similarity between Mex. and Guat. subpopulations, when compared to their relationship with the W.I. subpopulation, supports evolutionary theories suggesting that the West Indian ancestry separated earlier[48,52]. This phylogenetic arrangement for the avocado ecological races may indicate the biogeographical history associated with the domestication and variation of avocados throughout the middle and tropical regions of the Americas.

Future breeding efforts should prioritize crosses between the Guatemalan and Mexican races, as hybridization between these groups appears relatively uncommon. The low occurrence of hybrids between these races may stem from pollination incompatibility,

which could be overcome through targeted crossbreeding strategies. Successful crosses between pure Mexican and Guatemalan genotypes could generate new varieties with enhanced genetic diversity. Similarly, breeding programs should explore hybridization among all three races, leveraging different genomic proportions to develop cultivars with improved adaptability, nutritional, and agronomic traits.

Genetic structure of flowering type locus

The exact genetic makeup of the flowering type in avocado has been debated in previous studies, with an earlier report suggesting multiple loci[24], whereas later studies indicated one locus[14,15]. In the current study, several lines of evidence, including a multi-model GWAS approach and BSA, consistently revealed a strong association signal delineated to a precise haplotype block on chromosome 10. The GWAS analysis using SNPs as well as structural variants identified significant markers at the same locus, further supporting the evidence for the role of this region in regulating flowering type. Previous research has highlighted the significance of SVs in influencing plant phenotypes, especially concerning traits such as flowering time, disease resistance, and fruit development[53,54]. Secondly, bulked segregation analysis (BSA) of two breeding populations
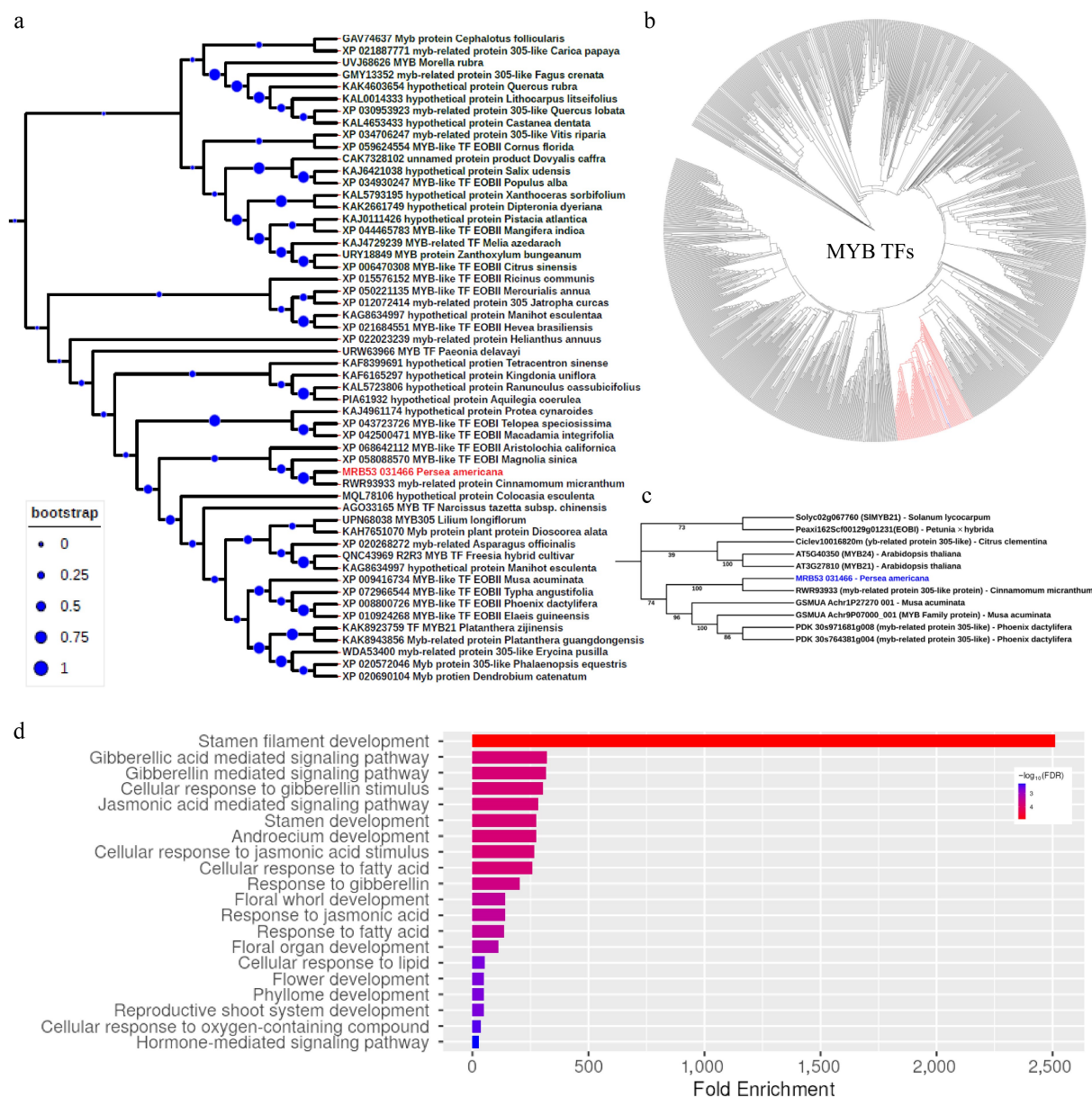
Fig. 4 Phylogenetic analysis of PaDSPD transcription factor. (a) Phylogenetic tree illustrating the evolutionary relationship of PaDSPD proteins from different botanical lineages, revealing it as a homolog of *EMMISSION OF BENZOIDS* (*EOB1*), a MYB transcription factor. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree was inferred from 1,000 replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. (b) Phylogenetic tree of MYB transcription factor protein sequences from *Arabidopsis thaliana*, *Solanum lycopersicum*, *Citrus clementina*, *Petunia axillaris*, *Cinnamomum micranthum*, *Musa acuminata*, and *Phoenix dactylifera*, showing broader evolutionary relationship of EOB1 highlighted in red color. (c) Phylogenetic tree of subgroup 19 within the MYB family, demonstrating the evolutionary placement of PaDSPD highlighted in blue font within this subgroup. (d) Functional enrichment analysis of the *Arabidopsis thaliana* homologs of *PaDSPD*, highlighting key biological processes.

segregating for flowering type provided additional evidence for the same genomic region associated with flowering type. Thirdly, the classical genetic analyses demonstrate a one-gene model for avocado flowering type, which agrees with the one-gene model reported previously[14,15] but contradicts the multiple loci model[24]. Regarding results reported by Lavi et al.[24], some of the crosses, such as Type A × B, had sufficient population sizes consistent with these results. However, results from other crosses are not consistent with results and those of Ashworth et al.[14], suggesting that smaller population sizes or misclassification of some progenies, as pointed out by these authors, could be the likely explanation for the discrepancies. Also, there is a possibility of outcrossing of some of the reported selfed crosses, e.g., B × B selfing, which resulted in a 2.7:1 B/A segregation ratio (Table 2, row 4, Lavi et al.[24]) , plausibly possible if outcrossing occurred since results have shown that Type B is always homozygous, which can only yield homozygous progenies and should not segregate. Likewise, Ashworth et al.[14] also

identified several markers linked to a quantitative trait locus (QTL) for flowering type, as determined through QTL mapping in a population derived from a cross between cv. 'Gwen' (Type B) and 'Fuerte' (Type A). The co-localization of GWAS signals with high F_{ST} values in the study also shows that this region could be under positive selection pressure during avocado evolution, perhaps to promote cross-pollination and genetic diversity. According to Alcaraz and Hormaza[55], this drift suggests that positive selection may have been involved in the regulation of flowering organization throughout avocado's evolutionary past to promote outcrossing and increase genetic diversity[33].

Insights into the mendelian inheritance of *PaDSPD* gene

Both genomic and mendelian inheritance patterns suggested that flowering type in avocado is controlled by a single gene. The
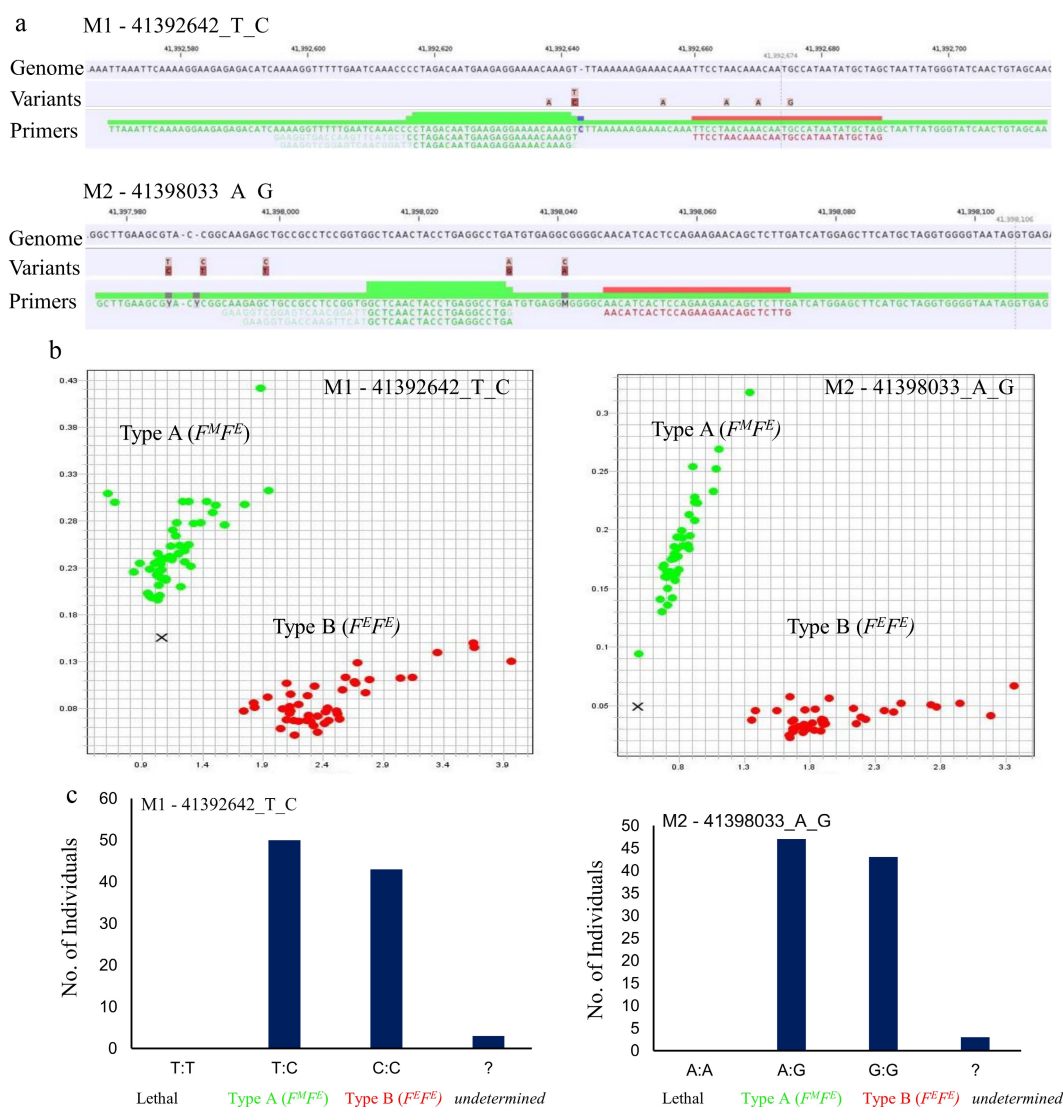

Fig. 5 Development and validation of competitive allele-specific PCR markers for avocado flowering type. (a) Genomic context of the PACE allele-specific primers (F^M 'female in the morning' and F^E, 'female in the evening' highlighted in green, and the common reverse primer, highlighted in red) targeting two significant SNPs; M1 - 41392642_T_C and M2 - 41398033_A_G. (b) Scatter plots highlighting distinct grouping of accessions into two genetic clusters representing Type A (F^MF^E) and Type B (F^EF^E). Red dots represent homozygous alleles and green dots represent heterozygous alleles. X represents undetermined genotypes with ambiguous calls that could not be confidently assigned to either cluster. (c) Bar graphs displaying the calling rates for markers M1 - 41392642_T_C and M2 - 41398033_A_G. The data represent genotype distributions across three categories: Type A (F^MF^E), Type B (F^EF^E), and the lethal genotype (F^MF^M).

mendelian inheritance analyses in the 'Simmonds' selfed progeny population revealed a modified 2:1 (Type A : Type B) ratio—in contrast to the expected 3:1 mendelian ratio—which agrees with a single gene model where one of the rare allele homozygous genotypes is lethal ($F^M F^M$). Likewise, in both reciprocal avocado populations – 'Tonnage' × 'Simmonds' and 'Simmonds' × 'Tonnage', where 'Simmonds' is heterozygous ($F^M F^E$) and 'Tonnage' is homozygous ($F^E F^E$), the expected 1:1 segregation ratio was observed. This type of inheritance is similar to what has been documented in other heterodichogamous plant families like walnut (*Juglans nigra* L.) and pecan (*Carya illinoinensis*), whose dichogamy is normally controlled by one locus with dominance interactions[56,57]. Observation of a 2:1 segregation ratio is noteworthy as it provides evidence for the single-gene inheritance pattern with homozygous lethality, which has previously not been reported for avocado flowering type. Although analyses suggest that $F^M F^M$ is lethal, future in-depth investigations, such as pollen/embryonic lethality, seedling mortality, gametophytic selection, and self-incompatibility, are required to determine the cause and mechanism of lethality and to rule out environmental factors. This discovery is important because it has broad implications for breeding programs for developing new cultivars with a preferred flowering type[58].

R2R3 MYB-like transcription factor as a regulator of flowering type in avocado

In this study, *PaDSPD* (MRB53_031466) was identified as the most likely candidate gene controlling flowering type in avocados. The localization of all significant markers exclusively in the promoter and coding region of *PaDSPD* strongly supports its role in regulating floral opening timing. Homology search analysis (Fig. 4a) confirmed that *PaDSPD* encodes an R2R3-type MYB-like transcription factor, closely related to *EOB1* and *EOB11* in several plant lineages, including the distantly related *Arabidopsis MYB21* (At3g27810) and *MYB24* (At5g40350), and the closely related *Cinnamomum micranthum* and *Magnolia sinica EOB1*. These MYB proteins belong to subgroup 19 of the R2R3 MYB family (Stracke et al.) and have established roles in floral development, influencing processes such as floral organ identity, pollen maturation, and flowering time[27,59,60]. Specifically, *MYB21* and *MYB24* regulate jasmonate-mediated stamen development[61], suggesting that *PaDSPD* plays a crucial role in the temporal control of stamen function and contributes to the distinct diurnal pattern of protogynous dichogamy observed in avocado. Functional enrichment analysis (Fig. 4d; Supplementary Fig. S6a) further underscores the dominant role of *PaDSPD* in floral development. Previous studies by Chopy et al.[27] have demonstrated that MYB transcription factors regulate diverse floral processes. Mutants of *MYB21* and *MYB24*, as well as *MYB21 MYB24* double mutants, exhibit disruptions in floral development, including flower opening, petal expansion, anther filament elongation, anther sac dehiscence, inhibition of lateral vascular development in unfertilized carpels, and abscission of sepals, petals, and stamens. The conserved roles of MYB proteins in flowering time and floral architecture suggest that *PaDSPD* may similarly govern flowering type in avocado. Likewise, comparative genomics analysis identified homologs of *PaDSPD* in multiple plant species, indicating an evolutionarily conserved function in reproductive development (Fig. 4b, c). Its placement within the MYB subclade for floral regulation supports its functional capacity to influence flowering behavior in avocado. Furthermore, previous studies have demonstrated that MYB transcription factors orchestrate flowering and floral organ maturation across diverse crop species[62], reinforcing the conserved role of MYB proteins in

reproductive development. The current findings, along with previous reports, strongly suggest that *PaDSPD* controls flowering types by directly influencing stamen and pistil development. *PaDSPD* is proposed to be under the strong influence of the circadian clock and light, which collectively regulate flowering type in avocado[27,63]. In the context of circadian rhythm, it is interesting to note that the morning phase circadian genes, CCA1/LHY, are also MYB-like TFs[64]. Since MYB TFs function as dimers[65], *PaDSPD* might be functioning by making heterodimers with these TFs in controlling flower opening in response to the circadian rhythm. Furthermore, genetic differences in the promoter and/or coding regions of *PaDSPD* may drive variations in flowering type by differential allelic expression and/or protein structural variation. Regarding variation in protein structure, one amino acid change in the loop domain and several in or in the vicinity of the TAD domain were observed, which, given the importance of these domains in the function of these TFs, are likely to contribute to regulating flowering type by affecting making homo/heterodimers and/or interaction with other proteins. It is interesting to note that the F^M allele did not occur as a homozygote in the analyses, indicating that it likely interferes with an essential function resulting in lethality.

As shown in Supplementary Fig. S6a, *Arabidopsis* homologs of the *PaDSPD* gene are expressed in both stamens and petals, with predominant expression in the stamens. This pattern aligns with its previously suggested role in flower development. The *PaDSPD* gene is co-expressed with genes such as JAZ protein-coding genes and basic helix-loop-helix (bHLH) DNA-binding family proteins, which are involved in the jasmonic acid-mediated signaling pathway. For instance, *At4G17880* (*MYC4*; Basic helix-loop-helix (bHLH) DNA-binding family protein) and *At1G32640* (*MYC2*) are JAZ-interacting TF that act together with *MYC2* and *MYC3* to activate JA-responses[66]. Co-expression network suggests that *PaDSPD* interacts with *bHLH transcription factors*, including *MYC2*, *MYC3*, and *MYC4*, which are direct targets of JAZ proteins. This interaction leads to the formation of a bHLH-MYB complex that plays a crucial role in the regulation of flower opening and closing[63]. It has been reported that MYB/bHLH complexes regulate distinct cellular processes, including circadian rhythm in plants[67]. However, further extensive experimental work, such as transgenic complementation or gene knockout studies, is required to explore this aspect in detail, which is beyond the scope of this study. Floral organ development is tightly controlled by circadian rhythms, phytohormones, and transcription factors in a complex interplay[61,68]. Genetic variations within the *PaDSPD* locus may modulate these regulatory mechanisms, shaping the distinct flowering habits of Type A and Type B cultivars.

The identification of *PaDSPD* as a flowering-type regulator in avocado extends its known functional role beyond scent emission in *Petunia axillaris*[27]. Prior research has shown that *EOB1* modulates floral volatiles, influencing pollinator attraction and reproductive success[27–29]. However, findings suggest a broader regulatory function in avocado flowering, linked to phenological traits in perennial crops. Established flowering regulators such as *FT*, *AP1*, and *LFY* govern floral induction and differentiation[69], and the discovery of *EOB1* as a critical player in avocado suggests that it may interact with these genes within a transcriptional network governing flowering phenology. Comparative studies in *Arabidopsis* and *Petunia* indicate that floral timing and scent emission are interconnected via hormonal and genetic regulation. In *Petunia*, *EOB1* modulates jasmonic acid-responsive pathways, influencing flower maturation and volatile biosynthesis[27]. If a similar mechanism exists in avocado, *EOB1* may regulate floral opening or timing via interactions with jasmonic acid-mediated signaling pathway. Additionally,

its role may extend to pollinator preference, indirectly shaping reproductive success and fruit development.

PACE molecular markers development and validation for avocado flowering type breeding

For accelerating breeding of avocados, the development of competitive allele-specific markers for important agronomical traits is important. For instance, the flowering type markers developed and validated in the current study make early selection of flowering type in breeding programs possible with high efficiency and precision. It will have a significant impact since these markers could identify flowering type in the seedling and therefore facilitate interplanting of avocado with type A and type B flowering types, which has been shown to result in increased yield[16−19]. It will accelerate breeding cycles by allowing flowering type prediction at the seedling stage, rather than waiting for reproductive maturity, which can take several years in avocado[31]. Four markers were found, distinguishing Type A and Type B cultivars with high confidence. Remarkably, the identification of four markers within the 20 Kb QTL containing the *PaDSPD* gene provides a robust approach for distinguishing between Type A and Type B flowering patterns in avocado. These markers allow comprehensive tracking of the entire genomic region in segregating populations, thereby enabling haplotype-assisted breeding for flowering type. Compared to single-marker analysis, which can sometimes yield uncertain phenotype predictions due to the absence of a complete haplotype, the use of multiple markers enhances accuracy and confidence in selection strategies.

The applicability of these markers was validated in three genetically diverse hybrid populations resulting from crosses 'Hass' × 'Bacon', 'Simmonds' × 'Tonnage', and reciprocal 'Tonnage' × 'Simmonds' with contrasting flowering types and genetic backgrounds with high success rate. There is always a potential risk of allele dropout for an individual marker, even if it is highly significant. In the studies, multiple markers were developed and validated as a haplotype block, which minimized the potential risks of allele dropout or mis-scoring. The three hybrid populations originated from four genetically distant parents, all of which exhibited distinct binary flowering types (Type A or Type B). This, together with the 47 genetically diverse accessions, broadly representing the three ecotypes of *P. americana*, further suggest that the chances of allele dropout or mis-scoring are very low. Nevertheless, it would be good to test these markers in other populations involving other parents.

Conclusions

This research is an important contribution to determining the genetic underpinnings of synchronous dichogamy of avocado. Applying population genomics, association mapping using high-density markers, and comparative functional annotation analysis, a single gene, named *PaDSPD*, encoding an R2R3-type MYB transcription factor that controls flowering type in avocado, was uncovered. The markers developed in this study have direct applications in marker-assisted selection for flowering type in avocado breeding programs. Moreover, these findings enhance basic evolutionary knowledge of flowering type behavior and provide a foundation for studying such synchronization systems in other dichogamous species. Further research would require functional verification and validation of the *PaDSPD* gene using transgenic techniques or gene editing approaches. Investigation of upstream and downstream regulatory cascades and proteins that interact with this MYB

transcription factor would also reveal the molecular mechanism of the temporal segregation of male and female floral functions.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Ali GS; project management: Ali GS; experimental work and data analysis: Khan A, Eltaher S, Freeman B, Ali GS; flowering type data collection: Freeman B, Ali GS, Khan A; PACE marker design: Ali GS; PACE marker experiments and data analysis: Eltaher S, Khan A, Ali GS; draft manuscript preparation: Khan A, Eltaher S, Ali GS. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The data underlying this article are available in the article and in its online supplementary material. DNA sequencing data are available in the NCBI SRA archive under Project No. PRJNA1295960.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Hammami AM, Huang KM, Guan Z. 2024. An overview of the avocado market in the United States. *Edis* 2024:FE1150

[2] Huang KM, Guan Z, Blare T, Hammami AM. 2023. Global avocado boom. *Choices* 38:1−9

[3] Sagwe RN. 2022. *Pollinator diversity, pollination deficits, and pollination efficiency in avocado (Persea americana) production across different landscapes in Murang'a county, Kenya*. Doctoral Thesis. Universität Würzburg, Fakultät für Biologie, Germany. 148 pp. doi: 10.25972/OPUS-26920

[4] Alcaraz ML, Hormaza JI. 2024. Inadequate pollination is a key factor determining low fruit-to-flower ratios in avocado. *Horticulturae* 10:140

[5] Nirody BS. 1922. *Investigations in avocado breeding*. Master's thesis. Massachusetts Agricultural College, USA. 90 pp. doi: 10.7275/6083348

[6] Schnell RJ, Tondo CL, Brown JS, Kuhn DN, Ayala-Silva T, et al. 2009. Outcrossing between 'Bacon' pollinizers and adjacent 'Hass' avocado trees and the description of two new lethal mutants. *HortScience* 44:1522−1526

[7] Davenport T. 2011. Avocado Flowering. In *Horticultural Reviews*, ed. Janick J. Hoboken: John Wiley & Sons, Inc. 406 pp. www.wiley.com/en-us/Horticultural+Reviews%2C+Volume+8-p-9781118060933

[8] Ish-Am G. Avocado pollination: a review. *Proc. New Zealand and Australia avocado grower's conference, Tauranga, New Zealand, 20–22 September, 2005*. New Zealand: avocadosource.com. 9 pp. http://avocadosource.com/Journals/AUSNZ/AUSNZ_2005/IshAmGad2005.pdf

[9] Stout AB. 1933. *The pollination of avocados*. Gainesville, Florida, USA: University of Florida Agricultural Experiment Station. Bulletin No.257. 44 pp. www.avocadosource.com/papers/Research_Articles/StoutAB1933.pdf

[10] Sedgley M. 1985. Some effects of daylength and flower manipulation on the floral cycle of two cultivars of avocado (*Persea americana* Mill., Lauraceae), a species showing protogynous dichogamy. *Journal of experimental botany* 36:823–832

[11] Stout AB. 1923. A study in cross-pollination of avocados in southern California. *California Avocado Association Annual Report 1922-1923*. New York: New York Botanical Garden. pp. 29–45 www.avocadosource.com/CAS_Yearbooks/CAS_08_1922/CAS_1922-23_PG_29-45.pdf

[12] Hormaza I. 2014. Factors influencing avocado fruit set and yield. *From the Grove* 2014:34–36

[13] Kamble V, Narayana C. 2024. Floral phenology and maturity indicesin avocado. *Indian Horticulture* 69:22–25

[14] Ashworth VETM, Chen H, Calderón-Vázquez CL, Arpaia ML, Kuhn DN, et al. 2019. Quantitative trait locus analysis in avocado: the challenge of a slow-maturing horticultural tree crop. *Journal of the American Society for Horticultural Science* 144:352–362

[15] Solares E, Morales-Cruz A, Balderas RF, Focht E, Ashworth VETM, et al. 2023. Insights into the domestication of avocado and potential genetic contributors to heterodichogamy. *G3: Genes, Genomes, Genetics* 13:jkac323

[16] Bekey R. 1986. Pollination of avocado Some new insights with special reference to the 'Hass' variety. *California Avocado Society Yearbook* 70:91–98

[17] Gustafson CD, Bergh BO. 1966. History and review of studies on cross-pollination of avocados. *California Avocado Society Yearbook* 50:39–49

[18] Robinson TR. 1933. Pollination and other factors influencing the production of avocados. *Proceedings of the Florida State Horticultural Society* 46:109–114

[19] Degani C, Goldring A, Gazit S, Lavi U. 1986. Genetic selection during the abscission of avocado fruitlets. *HortScience* 21:1187–1188

[20] Kobayashi M, Lin JZ, Davis J, Francis L, Clegg MT. 2000. Quantitative analysis of avocado outcrossing and yield in California using RAPD markers. *Scientia Horticulturae* 86:135–149

[21] Vrecenar-Gadus M, Ellstrand NC. 1985. The effect of planting design on out-crossing rate and yield in the 'Hass' avocado. *Scientia Horticulturae* 27:215–221

[22] Garner LC, Ashworth VETM, Clegg MT, Lovatt CJ. 2008. The impact of outcrossing on yields of 'Hass' avocado. *Journal of the American Society for Horticultural Science* 133:648–652

[23] Davenport TL. 2019. Cross- vs. self-pollination in 'Hass' avocados growing in coastal and inland orchards of Southern California. *Scientia Horticulturae* 246:307–316

[24] Lavi U, Lahav E, Degani C, Gazit S. 1993. Genetics of skin color, flowering group, and anise scent in avocado. *Journal of Heredity* 84:82–84

[25] Chen S, Peng X, Xie Z, Zhang M, Huang A, et al. 2025. Genetic and genomic insights into dichogamy in Zingiberaceae. *Plant communications* 6:101352

[26] Endress PK. 2020. Structural and temporal modes of heterodichogamy and similar patterns across angiosperms. *Botanical Journal of the Linnean Society* 193:5–18

[27] Chopy M, Binaghi M, Cannarozzi G, Halitschke R, Boachon B, et al. 2023. A single MYB transcription factor with multiple functions during flower development. *New Phytologist* 239:2007–2025

[28] Liu F, Xiao Z, Yang L, Chen Q, Shao L, et al. 2017. PhERF6, interacting with EOBI, negatively regulates fragrance biosynthesis in petunia flowers. *New Phytologist* 215:1490–1502

[29] Spitzer-Rimon B, Farhi M, Albo B, Cna'ani A, Ben Zvi MM, et al. 2013. The R2R3-MYB–like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating *ODO1* and structural scent-related genes in petunia. *The Plant Cell* 24:5089–5105

[30] Boza EJ, Tondo CL, Ledesma N, Campbell RJ, Bost J, et al. 2018. Genetic differentiation, races and interracial admixture in avocado (*Persea americana* Mill.), and *Persea* spp. evaluated using SSR markers. *Genetic Resources and Crop Evolution* 65:1195–1215

[31] Li J, Eltaher S, Freeman B, Singh S, Ali GS. 2024. Comprehensive genetic diversity and genome-wide association studies revealed the genetic basis of avocado fruit quality traits. *Frontiers in Plant Science* 15:1433436

[32] Schnell RJ, Brown JS, Olano CT, Power EJ, Krol CA, et al. 2003. Evaluation of avocado germplasm using microsatellite markers. *Journal of the American Society for Horticultural Science* 128:881–889

[33] Borrone JW, Olano CT, Kuhn DN, Brown JS, Schnell RJ, et al. 2008. Outcrossing in Florida avocados as measured using microsatellite markers. *Journal of the American Society for Horticultural Science* 133:255–261

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

[35] Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120

[36] Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760

[37] Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008

[38] Lin MF, Rodeh O, Penn J, Bai X, Reid JG, et al. 2018. GLnexus: joint variant calling for large cohort sequencing. *BioRxiv* :1–9

[39] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81:559–575

[40] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158

[41] Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197:573–589

[42] Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635

[43] Zhang L, Duan Y, Zhang Z, Zhang L, Chen S, et al. 2024. OcBSA: an NGS-based bulk segregant analysis tool for outcross populations. *Molecular Plant* 17:648–657

[44] Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution* 38:3022–3027

[45] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948

[46] Wu Z, Li T, Liu X, Yuan G, Hou H, et al. 2021. A novel R2R3-MYB transcription factor LlMYB305 from *Lilium longiflorum* plays a positive role in thermotolerance via activating heat-protective genes. *Environmental and Experimental Botany* 184:104399

[47] Whiley AW, Schaffer B, Wolstenholme BN. 2002. *The Avocado: Botany, Production and Uses*. Wallingford, UK, Cambridge, MA: CABI Publishing. 416 pp. www.cabidigitallibrary.org/doi/book/10.1079/9780851993577.0000#

[48] Chen H, Morrell PL, Ashworth VETM, de la Cruz M, Clegg MT. 2009. Tracing the geographic origins of major avocado cultivars. *Journal of Heredity* 100:56–65

[49] Ashworth VETM, Chen H, Clegg MT. 2011. Persea. In *Wild Crop Relatives: Genomic and Breeding Resources: Tropical and Subtropical Fruits*, ed. Kole C. Berlin, Heidelberg: Springer. pp. 173–189 doi: 10.1007/978-3-642-20447-0_8

[50] Borrone JW, Brown JS, Tondo CL, Mauro-Herrera M, Kuhn DN, et al. 2009. An EST-SSR-based linkage map for *Persea americana* Mill. (avocado). *Tree Genetics & Genomes* 5:553–560

[51] Kuhn DN, Livingstone DS, Richards JH, Manosalva P, Van den Berg N, et al. 2019. Application of genomic tools to avocado (*Persea americana*) breeding: SNP discovery for genotyping and germplasm characterization. *Scientia Horticulturae* 246:1−11

[52] Rendón-Anaya M, Ibarra-Laclette E, Méndez-Bravo A, Lan T, Zheng C, et al. 2019. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 116:17081−17089

[53] Saxena RK, Edwards D, Varshney RK. 2014. Structural variations in plant genomes. *Briefings in Functional Genomics* 13:296−307

[54] Alonge M, Wang X, Benoit M, Soyk S, Pereira L, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182:145−161.E23

[55] Alcaraz ML, Hormaza JI. 2014. Optimization of controlled pollination in avocado (*Persea americana* Mill., Lauraceae). *Scientia Horticulturae* 180:79−85

[56] Grauke LJ, Mendoza-Herrera MA, Miller AJ, Wood BW. 2011. Geographic patterns of genetic variation in native pecans. *Tree Genetics & Genomes* 7:917−932

[57] Chatwin W, Shirley D, Lopez J, Sarro J, Carlson J, et al. 2023. Female flowers first: QTL mapping in eastern black walnut (*Juglans nigra* L.) identifies a dominant locus for heterodichogamy syntenic with that in Persian walnut (*J. regia* L.). *Tree Genetics & Genomes* 19:4

[58] Neale DB, Martínez-García PJ, De La Torre AR, Montanari S, Wei XX. 2017. Novel insights into tree biology and genome evolution as revealed through genomics. *Annual Review of Plant Biology* 68:457−483

[59] Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, et al. 2010. MYB transcription factors in Arabidopsis. *Trends in Plant Science* 15:573−581

[60] Ambawat S, Sharma P, Yadav NR, Yadav RC. 2013. MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants* 19:307−321

[61] Song S, Qi T, Huang H, Ren Q, Wu D, et al. 2011. The jasmonate-ZIM domain proteins interact with the R2R3-MYB transcription factors MYB21 and MYB24 to affect jasmonate-regulated stamen development in *Arabidopsis*. *The Plant Cell* 23:1000−1013

[62] Yin J, Chang X, Kasuga T, Bui M, Reid MS, et al. 2015. A basic helix-loop-helix transcription factor, *PhFBH4*, regulates flower senescence by modulating ethylene biosynthesis pathway in petunia. *Horticulture Research* 2:15059

[63] Qi T, Huang H, Song S, Xie D. 2015. Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in Arabidopsis. *The Plant Cell* 27:1620−1633

[64] Harmer SL. 2025. The time machine: feedback loops, post‐transcriptional regulation, and environmental integration in the plant circadian oscillator. *The Plant Journal* 122:e70275

[65] Millard PS, Kragelund BB, Burow M. 2019. R2R3 MYB transcription factors–functions outside the DNA-binding domain. *Trends in Plant Science* 24:934−946

[66] Zhang C, Lei Y, Lu C, Wang L, Wu J. 2020. MYC2, MYC3, and MYC4 function additively in wounding-induced jasmonic acid biosynthesis and catabolism. *Journal of Integrative Plant Biology* 62:1159−1175

[67] Seo PJ, Mas P. 2014. Multiple layers of posttranslational regulation refine circadian clock activity in *Arabidopsis*. *The Plant Cell* 26:79−87

[68] Mizoguchi T, Wright L, Fujiwara S, Cremer F, Lee K, et al. 2005. Distinct roles of *GIGANTEA* in promoting flowering and regulating circadian rhythms in Arabidopsis. *The Plant Cell* 17:2255−2270

[69] Latif Ahmad P, Mohd Yaqub B, Nusrat A, Bilal Ahmad M. 2021. Floral induction pathways: decision making and determination in plants to flower-a comprehensive review. *Journal of Applied Biology & Biotechnology* 9:7−17