Received: 30 April 2024 | Accepted: 10 July 2024

DOI: 10.1111/ppl.14482

RESOURCE ARTICLE

Physiologia Plantarum

# Genetic mosaic of the Mediterranean fig: comprehensive genomic insights from a gene bank collection

Christos Bazakos[1,2,3#] | Michail Michailidis[2,4#] | Nikolaos Tourvas[5] | Konstantinos G. Alexiou[6,7] | Ifigeneia Mellidou[1,2] | Chrysanthi Polychroniadou[2,8] | Anastasia Boutsika[1] | Aliki Xanthopoulou[1,2] | Theodoros Moysiadis[1,8,9] | Christina Skodra[4] | Marios-Georgios Kollaros[4] | Evangelos Glavakis[10] | Vasileios Stournaras[11,12] | Evangelos Karagiannis[13] | Filippos Aravanopoulos[5] | Athanassios Molassiotis[4] | Georgia Tanou[2,8] | Ioannis Ganopoulos[1,2]

[1]Institute of Plant Breeding and Genetic Resources, ELGO DIMITRA, Thermi, Greece

[2]Joint Laboratory of Horticulture, ELGO DIMITRA, Thermi, Greece

[3]Max Planck Institute for Plant Breeding Research, Department of Comparative Development and Genetics, Cologne, Germany

[4]Laboratory of Pomology, Department of Horticulture, Aristotle University of Thessaloniki, Thermi, Greece

[5]Laboratory of Forest Genetics, Faculty of Agriculture, Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, Greece

[6]Centre for Research in Agricultural Genomics, CSIC-IRTA-UAB-UB, Barcelona, Spain

[7]Institut de Recerca i Tecnologia Agroalimentàries, Barcelona, Spain

[8]Institute of Soil and Water Resources, ELGO-DIMITRA, Thermi, Greece

[9]Department of Computer Science, School of Sciences and Engineering, University of Nicosia, Nicosia, Cyprus

[10]Glavakis Fruit Trees, Greece

[11]Institute of Olive Tree, Subtropical Crops and Viticulture, ELGO-DIMITRA, Department of Olive and Horticultural Crops, Kalamata, Greece

[12]Department of Agriculture, University of Ioannina, Arta, Greece

[13]Department of Agriculture, University of Western Macedonia, Florina, Greece

**Correspondence**
Christos Bazakos,
Email: cbazakos@elgo.gr

Georgia Tanou,
Email: gtanou@elgo.gr

Ioannis Ganopoulos,
Email: iganopoulos@elgo.gr

**Funding information**
Operational Program Competitiveness, Entrepreneurship and Innovation under the call RESEARCH - CREATE - INNOVATE (2nd cycle), Grant/Award Number: T2EΔK-03933

**Edited by** P. Paajanen

## Abstract

High-depth whole-genome resequencing of 53 diverse fig tree genotypes yielded a rich dataset of genetic variants. We successfully identified 5,501,460 single-nucleotide polymorphisms (SNPs) and 1,228,537 insertions and deletions (InDels), providing a high-density and excellent-quality genetic map of the fig tree. We also performed a detailed population structure analysis, dividing the 53 genotypes into three geographical groups and assessing their genetic diversity and divergence. Analysis of structural variants (SVs) and copy number variations (CNVs) revealed their potential functional impact, particularly in plant-pathogen interaction and secondary metabolism. Metabolomic fingerprinting of fig genotypes uncovered extensive variation in primary metabolites and polyphenolic compounds, highlighting the influence of genotype on fruit quality traits such as nutritional content and bioactive compound

[#] These authors contributed equally to this work.

composition. The genome-wide association study (GWAS) identified critical SNPs associated with fruit quality and morphological features. The discovery of significant candidate genes, such as AGL62, GDSL, and COBRA-like protein 4 genes, offers promising targets for marker-assisted selection and genome editing approaches to improve fig fruit morphological and quality traits. This extensive genomic analysis of fig trees enhances our understanding of the genetic basis of important agronomic traits and provides a rich resource for future research in this economically and nutritionally significant fruit.

## 1 | INTRODUCTION

The fig tree (*Ficus carica* L., Moraceae), an emblematic tree of the Mediterranean region, is a species of high heterozygosity (Mori et al., 2017) that was domesticated about 5500 years ago (Zohary and Spiegel-Roy, 1975). Currently, fig trees are grown throughout the Mediterranean area and in various temperate regions worldwide (Sadder and Ateyyeh, 2006). There are over 6,000 described fig cultivars, several of which are of great commercial importance for consumption, either fresh or dried (Flaishman, Rodov and Stover, 2008).

The fig fruit is a key natural provider of bioactive substances in the nutritious Mediterranean diet (Viuda-Martos et al., 2015). Its main health-enhancing phytochemicals include flavonoid and phenolic acids, trace minerals, fibers and vitamins (Veberic, Colaric and Stampar, 2008).

Interestingly, the classification and description of the different fig cultivars are mainly based on morphological traits of fruit and leaves. However, this approach is not very accurate and reliable since many phenotype features are highly dependent on climatic conditions. A comparative analysis of the genomes of several cultivars can shed light on their genetic background and the influence of artificial selection on their evolution (Li et al., 2019). Progress in genome sequencing technology has enabled the tracking of gene evolution in fruit trees, resulting in the discovery of numerous genes linked to horticultural traits (Zahid et al., 2022). This progress has produced a vast amount of critical information beneficial for plant molecular breeding, enhancing our comprehension of the classification and domestication of cultivated plants at the intraspecific level (Yu et al., 2018). Recently, Usai et al. (2020) unveiled a high-quality reference genome (~333 Mbp) for the most prominent Italian fig cultivar 'Dottato' that serves as a crucial tool for comprehensive genome resequencing analyses, enhancing fig breeding programs aimed at expanding the fruit's global cultivation and distribution.

In this work, we performed whole-genome resequencing of 53 fig varieties to uncover genome-wide variations, including single-nucleotide polymorphisms (SNPs), small insertions and deletions (InDels), copy number variants (CNVs) and structural variants (SVs). These variations were further analyzed to gain insights into the fig genome's characteristics and population structure, which will support ongoing studies and breeding efforts. A key part of our study was a genome-wide association study (GWAS), where we identified specific genetic regions and potential candidate genes connected to phenotypic variations in essential agro-morphological traits, like fruit number and juiciness, lenticel presence in fruit and fig productive type, which are critical for fig tree breeding. Additionally, we determined fruit physiological traits as well as the metabolic profiling of fig cultivars that were producing fruit, exploring metabolic diversity. The comprehensive resequencing of the fig genome and the detailed metabolome data collected from a diverse range of fig genotypes offer a valuable resource for the genetic improvement of this economically significant fruit tree.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant Material

We studied 53 fig tree genotypes belonging to 4 caprifigs (wild), 8 landraces, and 41 cultivars originating from several Mediterranean countries such as Greece (25), Spain (1), Italy (14), Cyprus (11), France (1) and Turkey (1) (Table S1). All cultivars and 3 caprifigs are maintained at the National Fig Tree Germplasm Repository (Institute of Olive Tree, Subtropical Crops and Viticulture, Kalamata).

### 2.2 | Morphological characterization

The 39 selected genotypes were characterized for morphological traits according to the guidelines provided by the International Union for the Protection of New Varieties of Plants-UPOV guidelines (UPOV, 2010), in the fig tree descriptors. The phenotypic observations were conducted for 3 consecutive years (2018–2021). Based on availability, between 2 to 5 trees for each genotype were sampled, and 23 morphological traits were qualitatively assessed according to rating and coding systems (Table S2).

### 2.3 | Whole genome re-sequencing

DNA of high molecular weight and quality was isolated from the leaf tissue of a single tree per genotype, utilizing the DNeasy Plant Pro Kit (Qiagen Inc.) following the guidelines provided by the manufacturer. Illumina libraries were generated with a mean

insertion size of 350 bp and sequenced (PE150) on Illumina NovaSeq 6000 system.

## 2.4 | Variation calling and annotation

Quality control for paired-end Illumina sequences was performed using FastQC (version 0.11.9) and MultiQC (version 1.9). The preprocessing step, which involved trimming sequencing adapters and discarding low-quality ends (<Q20), was executed with fastp (version 0.20). Sequences that passed the quality criteria (average quality ≥ Q20 and minimum length ≥ 18 bp) were aligned to the *Ficus carica* reference genome UNIPI_FiCari_1.0 (Usai et al., 2020) employing the Burrows-Wheeler Alignment Tool "bwa mem" (version 0.7.17). Documentation included mapping quality, sequencing depth, and coverage consistency (Table S1). The process for calling and annotating variants was in line with the methodology described by Bazakos et al. (2023). The functional gene annotation was employed by Usai et al. (2020) and validated by the OrthoFinder tool (Emms and Kelly, 2019).

## 2.5 | SNP variants in genes related to flowering time and phenylpropanoid biosynthesis

The genetic diversity within the candidate genes that are associated with controlling flowering period and phenylpropanoid production was examined among various genotypes, and the possible impact of genetic variations was analyzed using snpEff, which annotated each SNP according to its predicted effect.

## 2.6 | Structural Variant (SV) calling

Copy Number Variants (CNVs) were identified at the genome level using Control-FREEC v11.6 (Boeva et al., 2012). Prior to CNV detection, *minExpectedGC* and *maxExpectedGC* for non-human species were calculated using gccount as suggested by the authors. Following CNV detection, statistical significance was obtained using the "assess_significance.R" file, where both Wilcoxon and Koglomorov-Smirnov test p-values are added in each output file. Filtering was applied using a Kolmogorov–Smirnov P value threshold of 0.05. Nearby variants less than 100 bp apart were merged using *bcftools* and removed sites that were of mixed type (copy number gain and loss) and sites where the variant was not present in at least 3 lines. Finally, we generated a binary table with the CNVs for the 53 lines, representing absence and presence of a variant with "0" and "1", respectively.

Large deletions/insertions (>50 nt) and inversions were called using Manta (Chen et al., 2016; v1.6.0) at the individual level, following the software manual. After variant calling, individual vcf files were merged into a single VCF file using SURVIVOR (Jeffares et al., 2017; v1.07) with the following parameters: "max distance between breakpoints": *1000*, "minimum number of supporting caller": *1*, "take the type into account": *yes*, "take the strands of SVs into account": *yes*, "minimum size of SVs to be taken into account": *50*. The final set of Manta variants was obtained by keeping only those that had a minimum allele frequency (MAF) of 0.01 or higher, SV size of less than 1Mbp, less than 40% of missing genotypes and at least 3 individuals with the variant allele.

## 2.7 | Relatedness analysis

We used KING v2.3.0 (Manichaikul et al., 2010) with the "–kinship" argument to estimate degrees of relatedness between all samples based on pairwise comparisons of SNP data. The total unpruned genomic data set was used as the input file. Estimated kinship coefficients were used to infer relationship degrees between samples according to KING's manual and these were plotted into a network with the R package igraph (Csardi and Nepusz, 2006).

## 2.8 | Identification of selective sweep signals

Variant filtering was conducted using PLINK 1.9 (Chang et al., 2015) with the "–indep-pairwise 50 10 0.2" parameters, resulting in 291,872 independent variants. The filtered VCF file was converted to a bed format for input into the R package PCAdapt (Privé et al., 2020). PCAdapt detects outlier SNPs based on correlations between variants and PCA principal components which represent the neutral genetic structure in a data set. To avoid bias, only one individual represented each putative duplicate group as detected by the relatedness analysis. The ideal number of K was determined from a scree plot showcasing each principal component's contribution and sample projections on the initial PCA axes. P-values underwent Bonferroni correction for a = 0.01.

## 2.9 | Population genetics analyses

Excluding outlier SNPs identified by PCAdapt, genetic differentiation among samples was explored through PCA in PLINK. A relative dissimilarity distance matrix was generated using bitwise.dist function from poppr v2.9.3 (Kamvar, Tabima and Grünwald, 2014), which supported the construction of a Neighbor-Joining dendrogram with ape v5.7 (Paradis and Schliep, 2019) and ggtree v3.6.2 (Yu et al., 2017). ADMIXTURE (Alexander and Lange, 2011) was used to study admixture patterns among fig tree genotypes using the same dataset of unlinked variants. Ten runs were conducted for K = 1 to K = 5, with results consolidated by the CLUMPAK pipeline (Kopelman et al., 2015), and consistency found up to K = 3, establishing it as the optimal K.

For analyzing genetic structure with SV data, genotypes were encoded as haploid presence-absence data and analyzed in R 4.2.2 using adegenet v2.1.10 (Jombart, 2008). PCAs were performed separately for each SV type based on Euclidean distances using the cmdscale function (R Core Team, 2022), and a combined plot was created after identifying consistent patterns. A dissimilarity distance matrix was calculated using the diss.dist function from poppr to

construct a Neighbor-Joining dendrogram, following the methodology outlined for the SNP dataset by Bazakos et al. (2023).

Fixation indices (FST) between SEMed, WMed, and MEMed gene pools were calculated per Weir and Cockerham (1984) methodology. Nucleotide diversity ($\pi$) and Tajima's D were assessed as previously described by Bazakos et al. (2023).

### 2.10 | Genome-wide association study (GWAS)

For the genome-wide association study (GWAS), the list of phenotypic traits of 39 fig genotypes alongside their SNP data across 13 chromosomes were retained (Tables S1; S2; S3). SNP filtering was carried out based on criteria such as Minor Allele Frequency (MAF) (≥0.1), a minimum count of minor alleles (≥5), and a maximum missing rate (≤0.5), utilizing PLINK v1.9 (Chang et al., 2015). SNPrelate (v1.22.0) (Zheng et al., 2012) was employed to perform quality checks of bi-allelic SNPs. Association tests were conducted using the EMMAX pipeline (Kang et al., 2010). To adjust for the effects of population structure, kinship matrices were generated via the EMMAX-BN (Balding-Nichols) approach, and visual representations of the data were created with Manhattan and QQ plots using R (R Core Team, 2022). The Bonferroni correction was applied to correct for multiple testing.

### 2.11 | Fig fruit quality attributes

Quality determination based on soluble solids concentration (SSC), acidity (TA), respiration rate, ethylene biosynthesis, weight, color indicators, firmness, dry matter, and elements content (N, P, K, Ca), in 33 fig cultivars was performed at commercial harvest stage from 11/8/2021 to 26/8/2021 (Table S3). Fruits were harvested from 8:30–9:30 am. Respiration rate, ethylene biosynthesis, elements content, SSC (%) and TA (%, citric acid), was determined from 3 replicates of 5 fruit, while weight, color indicators, firmness, and dry matter was determined in 15 individual fruit at harvest.

Gas chromatography (Shimadzu GC-2014) determined respiration rate and ethylene production as outlined by Polychroniadou et al. (2022). Fig fruit color parameters ($L^*$, $a^*$ and $b^*$) and firmness were measured using a colorimeter (Minolta CR-200) and a TA.XT2i Texture Analyzer (Stable Micro Systems), respectively. Nutrient element quantification was performed using ICP-OES (Perkin Elmer Optima 2100DV), and nitrogen levels were estimated via the Kjeldahl method with a Vapodest 50s system (Karagiannis et al., 2021).

### 2.12 | Fig primary and secondary metabolic analysis

Primary metabolite analysis in fig edible tissue (3 biological replications of 5 fruit per cultivar) was conducted with a gas chromatography–mass spectrometry system (PerkinElmer Clarus SQ 8S GC/MS), following protocols by Polychroniadou et al. (2022) and detailed analysis by Michailidis et al. (2017). Metabolic peak identification was based on standards or the NIST11 database, with results expressed as adonitol relative abundance (Tables S4 and S5).

Polyphenolic compound analysis was performed using ultra-pressure liquid chromatography mass spectrometry (Agilent Technologies 1260 Infinity II Series LC system). Freeze-dried samples were extracted and subjected to UPLC–MS analysis as described by Valasiadis et al. (2024), with data acquisition and processing through OpenLab LC–MS software (Agilent Technologies).

## 3 | RESULTS AND DISCUSSION

### 3.1 | Comprehensive genomic variation map for fig trees

A comprehensive whole-genome resequencing of 53 fig tree genotypes from a diverse GenBank collection was carried out, achieving an average sequencing depth of 32.97X. The high-quality, filtered data were then aligned with the reference genome of the fig tree, *Ficus carica* UNIPI_FiCari_1.0 (Usai et al., 2020) (Figure 1a; Table S1).

Through the process of detecting variations, filtering genotypes, and imputing data, a comprehensive and high-quality dataset was produced from the 53 fig tree genotypes, uncovering 5,501,460 high-quality single nucleotide polymorphisms (SNPs) (1 SNP every 60 bases). Of these, 351,723 SNPs were located in coding regions (Figure 1b; Table S6.1). The ratio of non-synonymous to synonymous SNPs (Ka/Ks) was 1.35, as calculated by SNPeff tool, a figure in line with that seen in other fruit trees like the olive (1.45; Bazakos et al., 2023) but lower than that of sweet cherry (1.78; Xanthopoulou et al., 2020) and mango (1.52; Wang et al., 2020), which are typically cloned, and significantly higher than peach (1.06; Li et al., 2019).

The analysis of insertions and deletions (InDels) identified 1,228,537 variants, split between 642,059 insertions and 586,478 deletions. Of these, 0.52% were located within gene exons, and 0.09% affected splice sites, while 0.54% of InDels were found in 5′- and 3′-untranslated regions (UTRs), with the majority (55.94%) correspond to a single nucleotide (Figure 1b; Table S7.1), a finding consistent with reports on other Mediterranean perennials such as sweet cherry and olive tree (Bazakos et al., 2023; Xanthopoulou et al., 2020). Additionally, SVs and CNVs were analyzed, revealing 549 SVs, including insertions, deletions, and inversions, averaging 10.35 SVs per genotype (Figure 1c; Table S8). Furthermore, we identified genes that overlap with these SVs to highlight their potential functional importance. We found 157 DELs that overlapped completely (full gene overlap) with genes. Due to the absence of TE annotation, the gene annotation that was available from Ensembl Plants was only used however, no overlap was detected between genes and the identified insertions. We found only a single gene affected by an insertion when we looked for partial overlap (20% of the gene sequence). The functional significance of these SVs was further underscored through Gene Ontology (GO) enrichment analysis, revealing their roles in processes like stress response, defense, and auxin response, as well as binding functions (Figure 1d; Table S9). For CNVs, 202 were

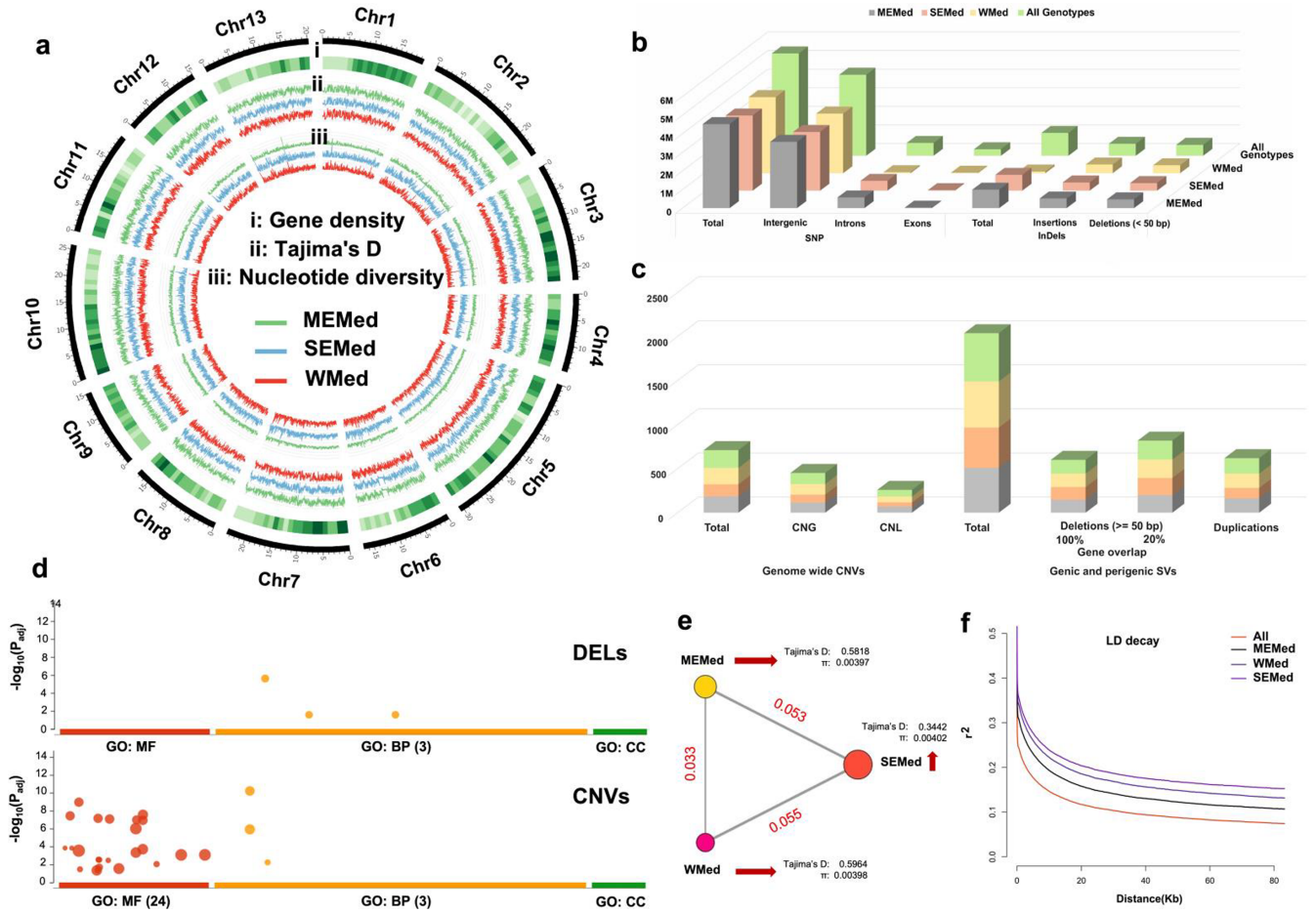


**FIGURE 1** Genomic variation map and divergence among the different fig tree groups. a) A circos plot representing genome-wide nucleotide diversity. Layers from the outer to inner are: i, Gene density; ii, Tajima's D; iii, Nucleotide diversity. The color coding for each group is green for Middle Mediterranean (MEMed), blue for South-East Mediterranean (SEMed), and dark red for West Mediterranean (WMed). b) The total number of SNP and InDel variants detected across 53 fig tree varieties and their corresponding group, classified by intergenic, introns and exons. c) The total number of identified Genome-wide Copy Number Variants (CNVs) classified by CNG (copy number gain), CNL (copy number loss) and genic/perigenic SV (structural variant). d) Enrichment analysis (Biological Process (BP), Molecular Function (MF), Cellular Component (CC)) of DELs and CNVs. e) Overview of nucleotide diversity (π and Tajima's D) and population differentiation (Fixation index-FST) among the three specified groups. The number inside each circle denotes the nucleotide diversity of that group, and the numbers between circles reflect the population divergence (FST). f) Analysis of linkage disequilibrium (LD) decay in fig trees across different groups.

identified, averaging 3.81 CNVs per genotype, with 176 genes overlapping with CNVs showing significant enrichment in pathways like plant-pathogen interaction and secondary metabolism, including isoflavonoid biosynthesis, highlighting the CNVs' functional relevance (Tables S8; S10).

These varied types of genomic sequence variations offer a valuable resource, and subsequent research into these structural variations and associated genes will enhance our comprehension of the molecular processes behind fig tree adaptation and cultivation.

Overall, the re-sequencing of 53 fig tree genotypes reveals in a comprehensive manner a wide population genetic diversity of this emblematic perennial fruit tree species. To date, research on this species, that is characterized by long flowering cycle, high sexual reproduction rate and unpredictable flowering period, mainly focuses on cultivation, while only a limited number of studies are conducted on fig tree selection and breeding. Hence, having access to representative genotypes as germplasm, along with their thoroughly detailed genomic variations as digital germplasm, will significantly boost research and breeding initiatives in the future (Zhao et al., 2021).

## 3.2 | Population structure analysis

To understand the population structure, we constructed a neighbor-joining (NJ) tree and a principal component analysis (PCA) and performed model-based clustering using ADMIXTURE (Figure 2a-c). Based on those, we categorized the 53 genotypes into three geographic clusters: 19 genotypes from the West Mediterranean (WMed; mainly Italy, Spain, and France), 23 genotypes from the Middle Mediterranean (MEMed; coastline of Turkey to Greece), and 11 genotypes from the South-East Mediterranean (SEMed; mainly Cyprus)

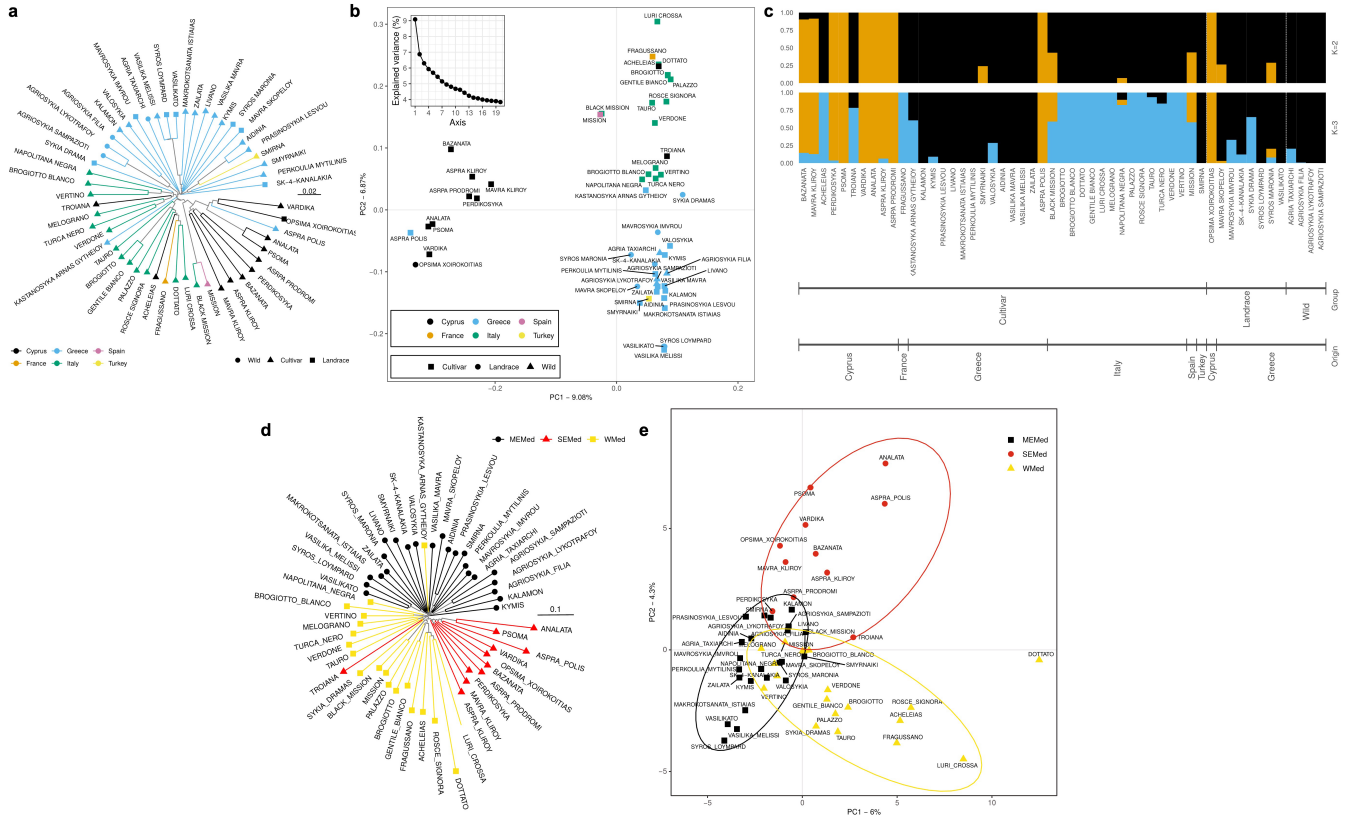


**FIGURE 2** Population structure analysis of 53 fig tree genotypes. Based on SNPs of the 53 fig tree genotypes: a) Neighbor-joining (NJ) dendrogram and b) Principal component analysis (PCA). c) Structure analysis. The results are depicted for K = 2 and K = 3. Based on SVs of the 53 fig tree genotypes, clustered in 3 groups: d) NJ dendrogram, e) PCA.

(Figure 2a). We have also used SVs as a measure of genetic diversity for generating the NJ phylogenetic tree and PCA to investigate whether the clustering of the three geographical groups is distinct (Figure 2d-e). Although the PCA and clustering heatmap of SVs are not sufficient to clearly discriminate the MEMed group from the other 2 (SEMed and WMed), the NJ phylogenetic tree offers a distinct clustering of the 3 groups.

To analyze the genetic variation and divergence among the three geographical groups, we calculated the nucleotide diversity ($\pi$) and Tajima's D values for each group and conducted a pairwise evaluation of genetic distances using fixation index values (FST) (Figure 1f). The SEMed group exhibited the highest nucleotide diversity ($\pi = 4.02 \times 10^{-3}$), coherent with a previous study that used Simple Sequence Repeats (SSRs) molecular markers (Ganopoulos et al., 2015). Compared with the SEMed group, the MEMed ($3.97 \times 10^{-3}$) and the WMed groups ($3.98 \times 10^{-3}$) showed a decreased nucleotide diversity (Figure 1f). The FST value between the SEMed and WMed groups was 0.055, decreasing slightly to 0.053 when comparing SEMed to MEMed, suggesting increased population differentiation with eastward expansion (Figure 1e). Additionally, Tajima's D values in the SEMed (0.344) were substantially lower than those in the MEMed (0.581) and WMed (0.596) groups, possibly reflecting a recent population bottleneck in the central and western Mediterranean regions due to allele fixation driven by genetic drift (Bazakos et al., 2023).

Furthermore, linkage disequilibrium (LD) decay for the three fig tree groups was rapid at 6 kb for the entire population (Figure 1f). The SEMed population showed a slower decay rate (15 kb) compared to the MEMed group (8 kb). Cultivated genotypes from the western Mediterranean demonstrated a quicker LD decay than those from the eastern region, likely due to the inclusion of genotypes from diverse geographic origins, such as Greece, Spain, and Italy, in the western group. The relatively fast LD decay in these fig tree genotypes could potentially improve the precision of association studies in identifying narrow intervals for candidate Quantitative Trait Loci (QTL).

## 3.3 | Cultivation history of fig tree

The precise tracing of fig tree ancestry using genomic information is challenging because of the simultaneous occurrence of asexual propagation and sexual reproduction (Ganopoulos et al., 2015). Though, the patterns of identity-by-descent (IBD) relationships were analyzed among the 53 fig tree genotypes (Figure 3). The pedigree network showed that most of the first-degree relationships were among genotypes in the same group. The analysis of interrelationship order placed the genotype 'F6' ('Analata') from the SEMed group in a monozygotic twin relationship with 'F36' ('Psoma'), and identified 'F32' ('Opsima Xoirokoitias') as third-degree relatives (Figure 3; Table S11).

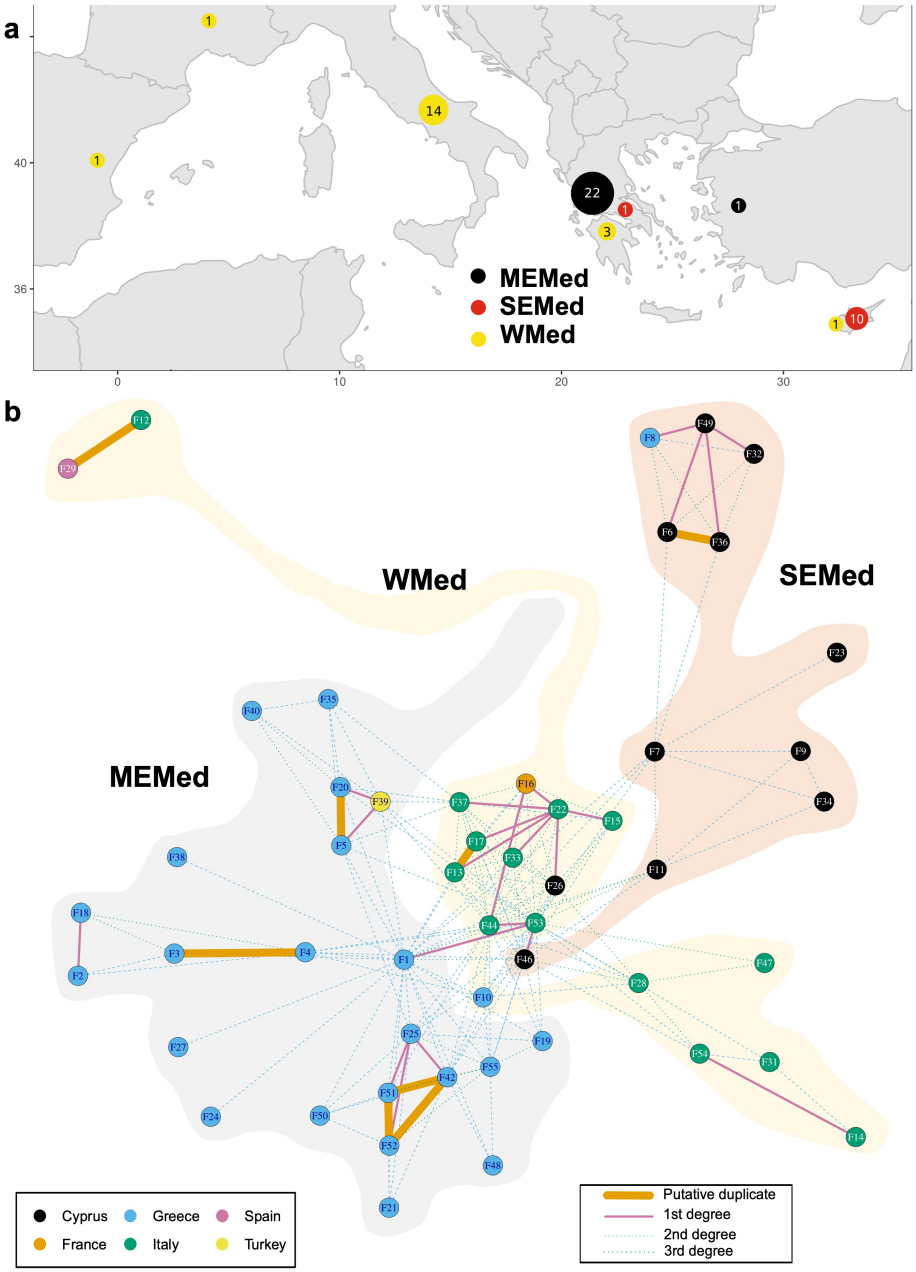


**FIGURE 3** a) Origin by geography of 53 fig tree genotypes. b) Kinship relationships were determined among various fig tree genotypes, mapping each one to its nearest kin based on the kinship coefficient. The nature of these relationships (e.g., putative duplicate samples, first, second, or third degree) is depicted using varied line styles. The degree of relationship was calculated using KING software, indicating the extent of shared heterozygous or homozygous haplotype blocks among the genotypes (Table S1). The nodes representing genotypes are colored according to their country of origin, visually distinguishing their geographical backgrounds. Colors encompassing genotype nodes correspond to populations: Grey, MEMed; light pink, SEMed; light yellow, WMed.

Furthermore, MEMed genotypes 'F5' ('Aidinia') and 'F20' ('Prasinosykia Lesvou') showed first-degree relationships with 'F39' ('Smyrna type') (Figure 3; Table S11), suggesting they originate from a distinct breeding event. Within the WMed group, 'Luri Crossa' ('F22') is likely linked through the same breeding event as numerous other WMed genotypes, given their first-degree kinship (Figure 3; Table S11).

## 3.4 | Signatures of selection in fig tree genomic profiles

Mapping individuals to the principal components using a score plot and analyzing the variance explained helped ascertain that three PCs represent the optimal number for the SNP data. The PCadapt analysis highlighted SNPs under selective pressure across all chromosomes of the fig tree, identifying 168 regions across 11 chromosomes as potential candidate regions for selection sweeps (Figure 4; Table S12).

A small selection of these SNPs is found within genes of known function. For example, a locus on chromosome 1, an endonuclease gene, was flagged as under selection. This gene (EG14) is part of the endo-β-1,4-glucanases (EGs) family under the glycosyl hydrolase family 9 (GH9), which is involved in cell wall synthesis, remodeling, and degradation (Jara et al., 2019). EGs have been previously demonstrated to play crucial roles in the ripening process of various fruits, including figs (Cui et al., 2022). Furthermore, a selection region on chromosome 10 contains the glutathione S-transferase (GST) gene (FCD_00000376), which plays a role in developing fig fruit peel color (Liu et al., 2023). Additionally, on chromosome 13, a selected area

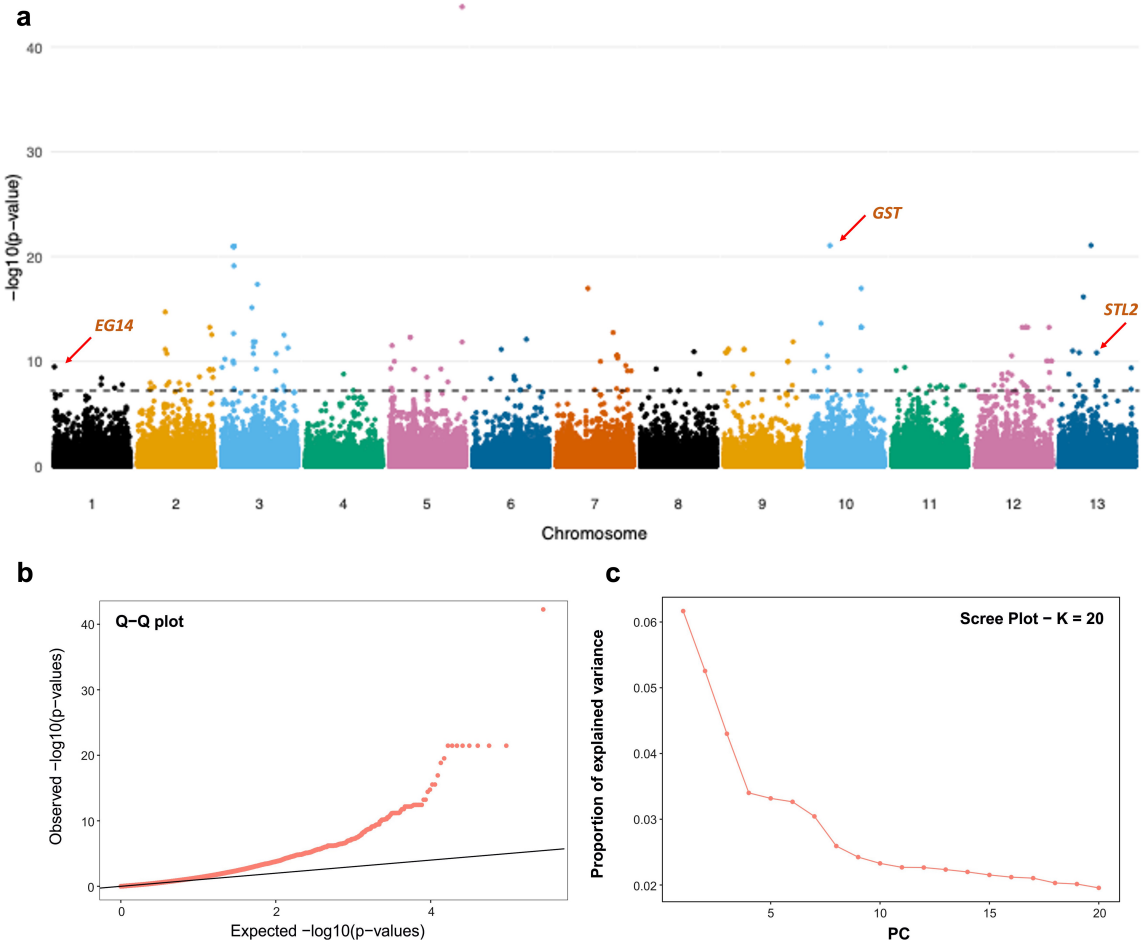


**FIGURE 4** Identification of signatures of selection in fig tree genomes through pcadapt. a) A Manhattan plot illustrating adjusted P-values highlights genomic loci containing SNPs subject to selection pressure. b) A quantile-quantile plot comparing the observed to the expected transformed P-values. c) A scree plot showing the variance explained by the first 20 principal component analysis axes, encompassing both adaptive and neutral SNPs.

encompasses STELLO2 (STL2) (FCD_00021095), identified as a Golgi-residing glycosyltransferase. This gene is intimately connected to cellulose synthesis in the cell wall as it orchestrates the formation and transport of cellulose synthase complexes (Zhang et al., 2016) (Figure 4; Table S12).

## 3.5 | SNP variants in genes involved in the phenylpropanoid-flavonoid (PF) pathway and flowering

Flower bud differentiation is essential for fruit production and occurs during the shoot elongation period in figs. During fig fruit development, secondary metabolites, such as polyphenolic compounds, which enhance the nutritional value of figs, are accumulated (Wang et al., 2021). Herein, we selected candidate genes related to flowering (Hanke et al., 2007) and secondary metabolisms such as phenylpropanoid and flavonoid biosynthesis (Hoang et al., 2015) to identify annotated SNP variants with a high impact on those genes' function (Table S13).

Our analysis identified important high-impact mutations in flowering genes of several fig genotypes. Specifically, two stop-gain mutations were revealed (T/C and C/G) in the *FRIGIDA (FRI)* gene (FCD_00001095), the principal activator of *FLC* (*Flowering Locus C*) that acts as a repressor of floral transition in *Arabidopsis* (Jung and Müller, 2009). Additionally, a stop-gain mutation (G/A; FCD_00009609) was revealed in the transcription factor *Agamous-like* gene (*AGL61*) that regulates the expression of downstream genes during central cell development (Steffen et al., 2008) (Table S13).

Concerning the phenylpropanoid-flavonoid genes, one stop gained mutation was identified in the 2-hydroxyisoflavanone dehydratase (*2-HID*) gene (FCD_00009376; C/T) that catalyzes the isoflavone biosynthesis (He et al., 2011) (Table S13). Furthermore, a significant mutation was discovered in the chalcone isomerase (CHI) gene (G/A; FCD_00011570), a key enzyme that acts as the second rate-limiting step in the flavonoid biosynthesis pathway (Yin et al., 2019) (Table S13). Recent research by Li et al. (2020) highlighted *FcCHI1*'s crucial contribution to anthocyanin production in figs. In addition, a critical mutation (G/A; FCD_00011868) was found in the gene

encoding 4-coumarate-CoA ligase (4CL), an enzyme essential for directing the flow within various phenylpropanoid biosynthesis pathways in plants (Lavhale, Kalunke and Giri, 2018) (Table S13). This enzyme plays a key role in the metabolic engineering of pathways leading to the synthesis of compounds such as curcuminoids, resveratrol, biofuels and in the enhancement of nutritional profiles. (Lavhale, Kalunke and Giri, 2018).

### 3.6 | Fruit quality and metabolome variation within the selected fig collection

Since the existing genotypic variation within the initial fig tree collection, a total of 33 cultivars and landraces from the three clusters, exhibiting significant genetic differences as well as contrasting pericarp and pulp coloration, were further evaluated for several fruit physiological traits (Figure 5; Table S3). The analysis of skin color parameters – a discriminant morphometric traits for fig screening – indicated that 'Vasilika Mavra' and 'Valosykia' were more diverse in coloration, having a dark-purple and a light-yellow skin color, respectively (Figure 5; Table S3). Previously, the dark peel of fig accessions was assumed to be genetically closer to the wild type, while genotypes with green peel represent a color mutation related to the MYB family transcription factors (Wang et al., 2017). Fresh weight of fruits between the tested genotypes ranged from 18.2 g to 53.9 g, whereas dry weight (%) ranged from 20.1% to 39.0% (Table S3). Other standard parameters for determining fruit quality at harvest include soluble solid content (SSC) and titration acidity (TA) (Karagiannis et al., 2021), both showing remarkable differences among the fig genotypes (Figure 5; Table S10). The top SSC was recorded in the MEMed and WMed groups, while the top TA was in the SEMed and WMed groups (Table S3). The values for SSC were in line with previous reports, whereas TA values were lower than in other *ex-situ* fig collections (Hssaini et al., 2020). Regarding fruit firmness, skin penetration and fruit deformation forces are considered major factors in fruit textural properties (Karagiannis et al., 2021). In this study,

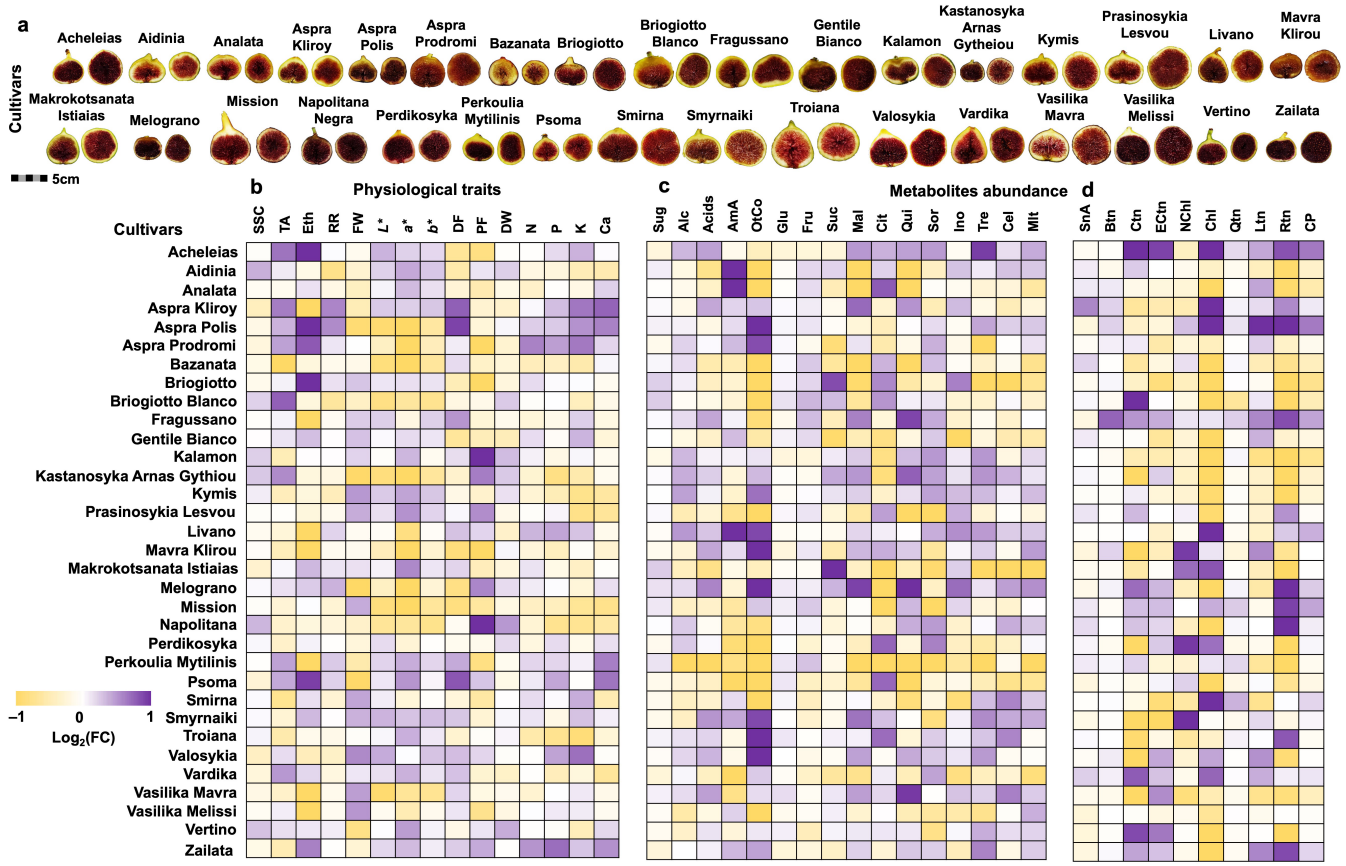


**FIGURE 5** Metabolomic fingerprinting of 33 fig fruit genotypes in transversely and horizontal section. a) Phenotype of 33 fig fruit genotypes, b) Heatmap of 15 physiological traits and nutrient content, c) abundances of 5 categories of primary metabolites and of 11 primary metabolic abundances, and d) abundances of 9 polyphenolic compounds along with cumulative polyphenols. Abbreviation list is provided, SSC: soluble solid concentration, TA: titratable acidity, Eth: Ethylene production, RR: respiration rate, FW: fresh weight, Color indicators L*, a,* b*, DF: deformation force, PF: penetration force, DW: dry weight, N: nitrogen, P: phosphorus, K: potassium, Ca: calcium, Sug: total sugars, Alc: total alcohols, AmA: total amino acids, OtCo: other compounds, Glu: glucose, Fru: fructose, Suc: sucrose, Mal: malic acid, Cit: citric acid, Qui: quinic acid, Sor: sorbitol, Ino: inositol, Tre: trehalose, Cel: cellobiose, Mlt: maltose, SnA: sinapic acid, Btn: biotin, Ctn: catechin, ECtn: epicatechin, NChl: neochlorogenic acid, Chl: chlorogenic acid, Qtn: quercetin, Ltn: luteolin, Rtn: rutin, CP: cumulative polyphenols. Numeric data are provided in Tables S10, S11 and S12.

'Mavra Klirou' and 'Aspra Polis', both assigned to the SEMed cluster, had the softest and the firmest fruit, respectively, within the collection (Table S3). This is partly in accordance with the results of penetration force analysis, demonstrating that the genetic clustering of these genotypes in the three groups (Figure 2) was not associated with fruit quality properties. A typical characteristic burst in both ethylene and respiration rate occurs at the onset of fig ripening, but unlike other climacteric fruits, this ethylene-dependent ripening process seems to be ABA-dependent (Qiao et al., 2021). In this study, a wide range in both ethylene and respiration rates was revealed (Figure 5c), but there was no obvious correlation between these traits. Interestingly, the respiration rate was negatively correlated with fruit weight (Figure 5b), suggesting that the smaller the fruit is, the higher the respiration rate it has, thus shortening the fruit shelf-file, as previously demonstrated in tomato (Ustun et al., 2023).

Fig fruits serve as valuable sources of macronutrients such as P, K, and Ca (Mendoza-Castillo et al., 2019), with their concentrations being variable across different genotypes, geographical locations, types of soil, and climates in general (George et al., 2023). Our results follow previous studies, proposing that nutrient status is highly dependent on genotype. It was evident that some genotypes, such as 'Aspra Klirou', 'Aspra Polis', 'Aspra Prodromi' (from the SEMed group) and 'Zailata' (from the MEMed group) were rich in all three macronutrients, while others, including 'Mission', 'Napolitana Negra' (WMed group), and 'Kymis' (MEMed group) were relatively poor (Figure 5; Table S3).

It is more than common ground that the genotype has a strong impact on the metabolomic composition of various horticultural species. To explore the extent of fruit metabolic variation within our fig collection, we quantified 27 metabolites of the primary metabolism, corresponding to 14 sugars, 3 organic acids, 4 amino acids and 6 others (Figure 5; Table S4). Among sugars, glucose and fructose were the most abundant carbohydrates, followed by sucrose, sorbitol (sugar alcohol), and inositol, but without showing any group-specific pattern of accumulation. Glucose and fructose have been previously reported to be among the main reduced sugars that form the flavor of fig fruit (Wang et al., 2021). Additionally, up to five- or six-fold variation was evident concerning organic acid and amino acid content, respectively, suggesting a broad diversity within the collection that could be valuable for breeding purposes towards enhancing fruit flavor and taste (Figure 5; Table S4). The most abundant organic acid was malate, followed by citrate, which is in line with previous reports in other tree species (Karagiannis et al., 2021) (Figure 5; Table S4). Previously, differences in amino acid composition and richness have been correlated with different flavor matter (Wang et al., 2021).

Genotype-specific variations in both composition and abundance of polyphenolic compounds were also revealed in this study, ranging from 105 to 275 μg $g^{-1}$ DW in total content (Figure 5; Table S5). For example, 'Acheleias', a cultivar from the WMed group, was characterized by markedly higher contents of individual polyphenolic compounds, such as catechins, epicatechins, chlorogenic acid and rutin and therefore could be recognized as a rich source of key bioactive health benefiting compounds. Among the monomeric flavonoids, epicatechin and catechin are usually highly abundant in the fig fruit (Wang et al., 2017). Other good sources of chlorogenic acid were 'Aspra Klirou', 'Aspra Polis', 'Livano' and 'Smirna', whereas 'Melograno', 'Mission', 'Napolitana Negra' from the WMed group contained high rutin levels. It is worth mentioned that no metabolic-based grouping of the examined cultivars was observed, suggesting that metabolomic fingerprinting cannot reflect the genetic relationship among the different fig genotypes.

## 3.7 | Genome-wide association study on fig plant and fruit agro-morphological traits

The National Fig Tree Germplasm Repository (Institute of Olive Tree, Subtropical Crops, and Viticulture, Kalamata, Greece) maintains an extensive collection of fig cultivars. This diverse collection presents a unique chance to carry out detailed phenotypic observations over several years, thereby reducing environmental impacts on the phenotype for application in GWAS. Following the International Union for the Protection of New Varieties of Plants-UPOV guidelines, we qualitatively assessed 23 morphological and fruit traits using the specified fig tree descriptors (Table S2).

The Pearson correlation analysis of 23 phenotypic traits (Figure 6A) revealed that the number of leaves on a two-year-old shoot has a positive correlation with three other plant characteristics, namely growth habit, vigor and branching density. Notably, branching density also shows a positive relationship with fruit quantity and maturation timing while inversely relating to fruit size and juiciness, two attributes of significant horticultural importance. The classification of the productive type (one crop/year, two crops/year, Wild) is logically linked to key fruit characteristics such as quantity, pulp color, and juiciness. The principal component analysis (PCA) conducted on the phenotypic traits of 39 fig genotypes did not exhibit clear clustering based on their geographical groups (MEMed, SEMed, and WMed) (Figure 6B). However, the principal components PC1 and PC2 were effective in distinguishing several genotypes within the same group (Figure 6B).

Leveraging the extensive phenotypic and genomic data acquired in this research, the challenge of a GWAS was undertaken to identify associations between genetic variations and observed traits. The study panel included 39 of the 53 genotypes, focusing on 23 agro-morphological characteristics, as detailed in Table S2. The GWAS was performed with a total of 87.223 SNP variants after filtering for minor allele frequency (MAF) (≥0.1), minor allele number (≥5), and missing rate (≤0.5). Bonferroni correction was applied to minimize the identification of false positive alleles and to determine significant peaks. Manhattan plots were created for each characteristic to illustrate the SNPs that showed a significant association with the agro-morphological traits throughout the *F. carica* genome (Figure 6C; Figure S1; Figure S2).

The present GWAS study revealed significant associations at several loci in 3 fruit traits (fruit number, fruit juiciness and fruit lenticel presence) and the productive type (Figure 6, Figure S1, Figure S2).

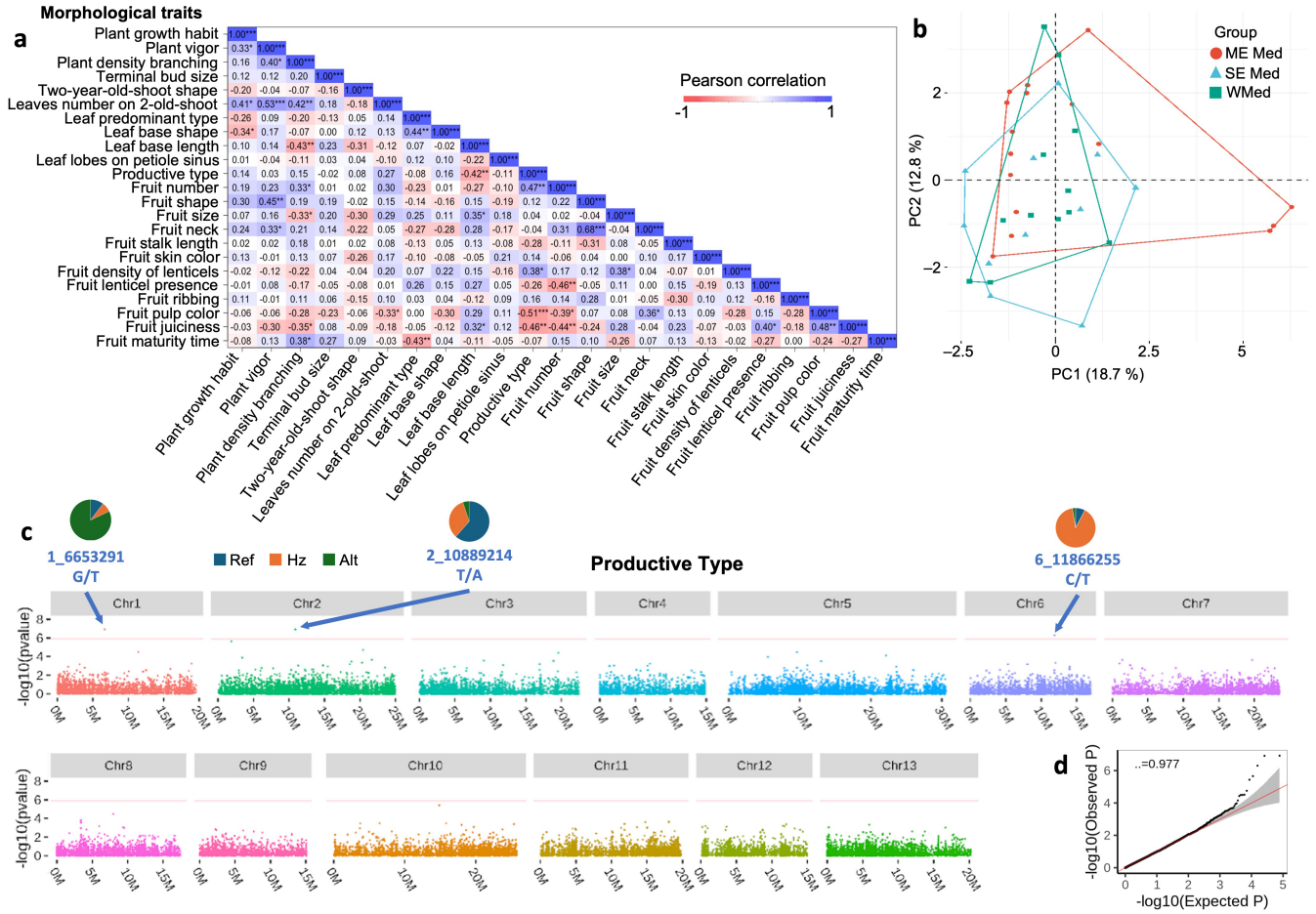


**FIGURE 6** a) Analysis of Pearson correlation among 23 morphological characteristics. b) Principal component analysis (PCA) focusing on the phenotypic characteristics of 39 fig varieties and their distribution into three population groups as shown in Figure 3 (SEMed, MEMed, WMed). c) Manhattan and d) quantile-quantile (QQ) plots from GWAS on the productivity type in 39 fig tree varieties. The plots display log-transformed P-values in relation to their chromosomal positions for 87,223 single nucleotide polymorphisms characterized by a minor allele frequency (MAF) (≥0.1), a count of minor alleles (≥5), and a rate of absence (≤0.5). The Bonferroni threshold is depicted by a horizontal red line. The pie charts depict the distribution of SNP alleles among the 39 fig tree genotypes, categorizing them into homozygous reference alleles (Ref), heterozygous alleles (Hz), and homozygous alternative alleles (Alt).

In our GWAS, we have identified several loci that are associated with fruit number (yield), fruit juiciness (quality) and fruit lenticel presence (morphology) (Figure S1; Figure S2). Based on the position of the significant SNPs, we scanned within a 100-kb interval (LD) the *F. carica* genome (Usai et al., 2020) for annotated genes (Table S14). Apart from SNPs that were linked to previously annotated genes, numerous loci were found to encompass a wide array of genes with unknown functions. Then, we further identified all the high-impact SNPs that are annotated as 'stop gained'; 'stop lost'; 'start lost'; 'splice acceptor variant'; and 'splice donor variant' in those genes, by performing the functional annotation (SNPeff) (Table S14).

Fruit juiciness is a commercially important trait that is strongly associated with fig fruit quality. In our GWAS, the significant peak is found in chromosome 11 (11_3738288) (Table S14; Figure S1). Several genes are annotated within the LD interval; however, the most interesting is the gene 'FCD_00013907' that encodes the flavoprotein subunit, succinate dehydrogenase subunit 4 (SDH) (Table S14). A previous study of the *Ilex paraguariensis* St. Hil. (yerba mate) tree revealed that SDH is upregulated in response to drought, serving as a physiological mechanism to prevent tissue desiccation by inducing stomatal closure during the initial phases of dehydration (Acevedo et al., 2013). Consequently, this suggests that SDH may also influence fig fruit juiciness by contributing to the avoidance of water loss.

The presence of lenticels in fruit is a morphological trait with agro-commercial significance since they reduce the fruit appearance quality. GWAS identified 3 loci with significant peaks, two ('1_599366' and '1_4564788') and one ('3_6457765') in chromosomes 1 and 3, respectively (Table S14; Figure S1). Among the genes that are annotated within the LD intervals of those 3 loci, we have identified the gene 'FCD_00028979', in chromosome 1, which encodes LOB domain-containing protein 36-like (Table S14). LBD proteins form a unique group of transcription factors exclusive to plants, which are crucial for controlling the development of plant organs and

are linked to critical developmental processes in fruits (Grimplet et al., 2017). Thus, the gene 'FCD_00028979' could be implicated in lenticel development and its occurrence on the fig fruits across various genotypes.

Apart from the fruit traits, we have performed GWAS on the productive type of trait. Fig varieties are classified into one of three productive types: a) wild (caprifig); b) Smyrna (non-parthenocarpic and requires pollination from the caprifig's profichi to develop its main crop) or Unifera, and c) San Pedro (produces its main crop following caprification and its second (breba) crop parthenocarpically) or Bifera, (Ferrara et al., 2016). Association analysis of the present dataset identified three loci on chromosomes 1,2 and 6 (Figure 6; Table S14). The most significant SNP (1_6653291) is located in an intron of the gene 'FCD_00032790' that encodes the GDSL esterases/lipases (GELP). Interestingly, the GDSL-type esterase/lipase gene, GELP77, causes male sterility in *Arabidopsis* and it is also needed for microspore nucleus development and pollen dissociation (Tsugama et al., 2020). Therefore, GELP could be a potential candidate gene that is related to the reproductive type in fig trees. Notably, the investigation for other candidate genes within the LD interval of "1_6653291", identified 7 tandem genes ('FCD_00032793';'FCD_00032794';'FCD_00032795';'FCD_00032796';'FCD_00032797';'FCD_00032798';'FCD_00032799') that encode the agamous-like MADS-box protein (AGL62) (Table S14). In plants, floral morphogenesis is affected by homeotic transcription factor families such as the AGAMOUS (AG) (Koornneef, de Bruine and Goettsch, 1980). AG participates in the induction of reproductive organ development in *Arabidopsis* through the formation of complexes with other MADS-box proteins (Homma and Goto, 2001). Furthermore, AGL62 is essential for inducing auxin synthesis in the endosperm of both strawberry and *Arabidopsis*, thereby facilitating successful fertilization (Guo et al., 2022). Therefore, the identification of AGL62 tandem genes within the LD interval makes it a very strong candidate for contributing to the productive type trait in fig trees and consequently to the future breeding efforts of *F.carica* L.

A significant finding within the linkage disequilibrium (LD) region for SNP 6_11866255 on chromosome 6 was the identification of a stop-gained mutation in the COBRA-like protein 4 gene at Chr6:11816120 (C/G) (refer to Figure 6; Table S14). This gene is crucial for cell wall synthesis and modification, processes that play a key role in determining fruit texture and firmness (Cao et al., 2012). The mutation in the COBRA-like protein 4 gene indicates possible changes in the firmness of the fruit, which could greatly affect the post-harvest longevity of figs and, consequently the consumer preferences.

Our findings contribute to the understanding of the genetic basis of important fruit and agronomic traits in fig trees while offering promising candidate targets for future breeding efforts. Given the economic and nutritional importance of figs, these genetic markers hold significant potential for the agricultural industry. Further studies are required to validate the functional consequences of these SNPs and to explore their potential use in marker-assisted selection programs.

## 4 | CONCLUSIONS

The current comprehensive genomic analysis of 53 fig tree genotypes has yielded a rich repository of genetic information. A high-density variation map of 5,501,460 high-quality SNPs and 1,228,537 InDels, along with the SVs and CNVs not only provides an extensive overview of the genetic diversity within the fig tree species but also establishes a foundational framework for future research and breeding efforts.

Additionally, the in-depth analysis of genetic variations, coupled with insights into the fig tree's cultivation history and signatures of selection in its genome, provides valuable information for future breeding programs. The elucidation of variation in genes involved in flowering and the phenylpropanoid-flavonoid pathway, as well as the assessment of fruit quality and metabolome variation, offers robust candidate genes for future functional analysis and breeding efforts. The present GWAS analysis identified critical SNPs associated with fruit quality and morphological traits. The discovery of significant candidate genes, such as *AGL62*, *GDSL*, and *COBRA-like protein 4* genes, offers promising targets for marker-assisted selection and genome editing approaches to improve fig fruit morphological and quality traits. This extensive genomic analysis of fig trees enhances our understanding of the genetic basis of important agronomic traits and provides a rich resource for future research in this economically and nutritionally significant fruit.

### AUTHOR CONTRIBUTIONS

I.G., C.B. and G.T. designed the idea and participated in revision and discussion; C.B., M.M., I.M., K.G.A., T.M., and N.T. executed data analysis; V.S., C.P., A.B., E.K., C.S., M.G.K., A.X., E.G., F.A., G.T. and A.M. helped in data collection and method; C.B., N.T. and M.M. conducted experiments; C.B., M.M., I.M., G.T. and I.G. drafted the manuscript. All author(s) read and approved the final manuscript.

### ACKNOWLEDGEMENTS

We would like to thank Ms A. Kardimaki and Ms P. Mpouna for their contribution to phenotypic observations, the bachelor student Pantelis Tirkos for helping us during the sampling process and the determination of fig quality traits and Dr. Sebastián Ramos-Onsins for the helpful comments and advice on population genetic analysis.

### FUNDING INFORMATION

This research was co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation under the call RESEARCH – CREATE – INNOVATE (2nd cycle) (Project Code: T2EΔK-03933).

### DATA AVAILABILITY STATEMENT

The raw sequencing data, in FASTQ format, was deposited in the National Centre for Biotechnology Information's (NCBI) Short Read Archive (SRA) database under the accession number PRJNA951283.

## ORCID

*Christos Bazakos* https://orcid.org/0000-0002-9536-9147
*Athanassios Molassiotis* https://orcid.org/0000-0001-5118-2244
*Georgia Tanou* https://orcid.org/0000-0002-9282-2794

## REFERENCES

Acevedo, R.M., Maiale, S.J., Pessino, S.C., Bottini, R., Ruiz, O.A., Sansberro, P.A., 2013. A succinate dehydrogenase flavoprotein subunit-like transcript is upregulated in Ilex paraguariensis leaves in response to water deficit and abscisic acid. *Plant Physiol. Biochem.* 65, 48–54.

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12, 1–6.

Bazakos, C., Alexiou, K.G., Ramos-Onsins, S., Koubouris, G., Tourvas, N., Xanthopoulou, A., Mellidou, I., Moysiadis, T., Vourlaki, I.T., Metzidakis, I., Sergentani, C., Manolikaki, I., Michailidis, M., Pistikoudi, A., Polidoros, A., Kostelenos, G., Aravanopoulos, F., Molassiotis, A., Ganopoulos, I., 2023. Whole genome scanning of a Mediterranean basin hotspot collection provides new insights into olive tree biodiversity and biology. *Plant J.* 116, 303–319.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., Barillot, E., 2012. Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28, 423–425.

Cao, Y., Tang, X., Giovannoni, J., Xiao, F., Liu, Y., 2012. Functional characterization of a tomato COBRA-like gene functioning in fruit development and ripening. *BMC Plant Biol.* 12, 1–15.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., Saunders, C.T., 2016. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222.

Csardi, G., Nepusz, T., 2006. *The igraph software package for complex network research*. InterJournal, Complex Systems, 1695.

Cui, Y., Zhai, Y., He, J., Song, M., Flaishman, M.A., Ma, H., 2022. AP2/ERF genes associated with superfast fig (*Ficus carica* L.) fruit ripening. *Front. Plant Sci.* 13, 1040796.

Emms DM, Kelly S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20(1):238.

Ferrara, G., Mazzeo, A., Pacucci, C., Matarrese, A.M.S., Tarantino, A., Crisosto, C., Incerti, O., Marcotuli, I., Nigro, D., Blanco, A., Gadaleta, A., 2016. Characterization of edible fig germplasm from Puglia, southeastern Italy: Is the distinction of three fig types (Smyrna, San Pedro and Common) still valid? *Sci. Hortic. (Amsterdam).* 205, 52–58.

Flaishman, M.A., Rodov, V., Stover, E., 2008. The fig: botany, horticulture, and breeding, in: *Horticultural Reviews*. pp. 113–196.

Ganopoulos, I., Xanthopoulou, A., Molassiotis, A., Karagiannis, E., Moysiadis, T., Katsaris, P., Aravanopoulos, F., Tsaftaris, A., Kalivas, A., Madesis, P., 2015. Mediterranean basin *Ficus carica* L.: from genetic diversity and structure to authentication of a Protected Designation of Origin cultivar using microsatellite markers. *Trees - Struct. Funct.* 29, 1959–1971.

George, T.T., Oyenihi, A.B., Oyenihi, O.R., Obilana, A.O., 2023. Composition and Health-Promoting Effects of Fig (*Ficus carica*) Extracts. *Fig (Ficus carica) Prod. Process.* Prop. 561–578.

Grimplet, J., Pimentel, D., Agudelo-Romero, P., Martinez-Zapater, J.M., Fortes, A.M., 2017. The LATERAL ORGAN BOUNDARIES Domain gene family in grapevine: Genome-wide characterization and expression analyses during developmental processes and stress responses. *Sci. Rep.* 7, 1–18.

Guo, L., Luo, X., Li, M. et al. Mechanism of fertilization-induced auxin synthesis in the endosperm for seed and fruit development. *Nat Commun* 13, 3985 (2022).

Hanke, M.V., Flachowsky, H., Peil, A., Hättasch, C., 2007. No Flower no Fruit – Genetic Potentials to Trigger Flowering in Fruit Trees. *Genes, Genomes and Genomics* 1, 1–20.

He, X.Z., Blount, J.W., Ge, S., Tang, Y., Dixon, R.A., 2011. A genomic approach to isoflavone biosynthesis in kudzu (*Pueraria lobata*). *Planta* 233, 843–855.

Hoang, V.L.T., Innes, D.J., Shaw, P.N., Monteith, G.R., Gidley, M.J., Dietzgen, R.G., 2015. Sequence diversity and differential expression of major phenylpropanoid-flavonoid biosynthetic genes among three mango varieties. *BMC Genomics* 16.

Honma, T., Goto, K., 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409(6819), 525–529.

Hssaini, L., Hanine, H., Razouk, R., Ennahli, S., Mekaoui, A., Ejjilani, A., Charafi, J., 2020. Assessment of genetic diversity in Moroccan fig (*Ficus carica* L.) collection by combining morphological and physico-chemical descriptors. *Genet. Resour. Crop Evol.* 67, 457–474.

Jara, K., Castro, R.I., Ramos, P., Parra-Palma, C., Valenzuela-Riffo, F., Morales-Quintana, L., 2019. Molecular Insights into FaEG1, a Strawberry Endoglucanase Enzyme Expressed during Strawberry Fruit Ripening. *Plants* 8, 140.

Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., Sedlazeck, F.J., 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* 8, 1–11.

Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.

Jung, C., Müller, A.E., 2009. Flowering time control and applications in plant breeding. *Trends Plant Sci.* 14, 563–573.

Kamvar, Z.N., Tabima, J.F., Gŕunwald, N.J., 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., Eskin, E., 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354.

Karagiannis, E., Sarrou, E., Michailidis, M., Tanou, G., Ganopoulos, I., Bazakos, C., Kazantzis, K., Martens, S., Xanthopoulou, A., Molassiotis, A., 2021. Fruit quality trait discovery and metabolic profiling in sweet cherry genebank collection in Greece. *Food Chem.* 342, 128315.

Koornneef, M. de Bruine. J. H., Goettsch, P., 1980. Arabidopsis Int. *Serv.* 17, 11–18.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191.

Lavhale, S.G., Kalunke, R.M., Giri, A.P., 2018. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta* 248, 1063–1078.

Li, J., An, Y., Wang, L., 2020. *Transcriptomic analysis of ficus carica peels with a focus on the key genes for anthocyanin biosynthesis*. Int. J. Mol. Sci. 21.

Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., Zhao, P., Guo, J., Ding, T., Guan, L., Zhang, Q., Guo, W., Fei, Z., Wang, L., 2019. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biol.* 20, 1–18.

Liu, L., Zheng, S., Yang, D., Zheng, J., 2023. Genome-wide in silico identification of glutathione S-transferase (GST) gene family members in fig (Ficus carica L.) and expression characteristics during fruit color development. *PeerJ* 11, e14406.

Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.M., 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873.

Mendoza-Castillo, V.M., Pineda-Pineda, J., Vargas-Canales, J.M., Hernández-Arguello, E., 2019. Nutrition of fig (*Ficus carica* L.) under hydroponics and greenhouse conditions. *J. Plant Nutr.* 42, 1350–1365.

Michailidis, M., Karagiannis, E., Tanou, G., Karamanoli, K., Lazaridou, A., Matsi, T., Molassiotis, A., 2017. Metabolomic and physico-chemical approach unravel dynamic regulation of calcium in sweet cherry fruit physiology. *Plant Physiol. Biochem.* 116, 68–79.

Mori, K., Shirasawa, K., Nogata, H., Hirata, C., Tashiro, K., Habu, T., Kim, S., Himeno, S., Kuhara, S., Ikegami, H., 2017. Identification of RAN1 orthologue associated with sex determination through whole genome sequencing analysis in fig (*Ficus carica* L.). *Sci. Rep.* 7.

Paradis, E., Schliep, K., 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528.

Polychroniadou, C., Karagiannis, E., Michailidis, M., Adamakis, I.D.S., Ganopoulos, I., Tanou, G., Bazakos, C., Molassiotis, A., 2022. Identification of genes and metabolic pathways involved in wounding-induced kiwifruit ripening. *Plant Physiol. Biochem.* 179, 179–190.

Privé, F., Luu, K., Vilhjálmsson, B.J., Blum, M.G.B., Rosenberg, M., 2020. Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. *Mol. Biol. Evol.* 37, 2153–2154.

Qiao, H., Zhang, H., Wang, Z., Shen, Y., 2021. Fig fruit ripening is regulated by the interaction between ethylene and abscisic acid. *J. Integr. Plant Biol.* 63, 553–569.

R Core Team, 2022. *R: A language and environment for statistical computing. R foundation for statistical computing.Vienna*, Austria [WWW Document]. https://www.R-project.org/.

Sadder, M.T., Ateyyeh, A.F., 2006. Molecular assessment of polymorphism among local Jordanian genotypes of the common fig (*Ficus carica* L.). *Sci. Hortic. (Amsterdam).* 107, 347–351.

Steffen, J.G., Kang, I.H., Portereiko, M.F., Lloyd, A., Drews, G.N., 2008. AGL61 interacts with AGL80 and is required for central cell development in *Arabidopsis*. *Plant Physiol.* 148, 259–268.

Tsugama, D., Fujino, K., Liu, S., Takano, T., 2020. A GDSL-type esterase/lipase gene, GELP77, is necessary for pollen dissociation and fertility in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* 526, 1036–1041.

Usai, G., Mascagni, F., Giordani, T., Vangelisti, A., Bosi, E., Zuccolo, A., Ceccarelli, M., King, R., Hassani-Pak, K., Zambrano, L.S., Cavallini, A., Natali, L., 2020. Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *Plant J.* 102, 600–614.

Ustun, H., Dogan, A., Peker, B., Ural, C., Cetin, M., Ozyigit, Y., Erkan, M., 2023. Determination of the relationship between respiration rate and ethylene production by fruit sizes of different tomato types. *J. Sci. Food Agric.* 103, 176–184.

Valasiadis, D., Georgios, M., Michailidis, M., Polychroniadou, C., Tanou, G., Bazakos, C., Molassiotis, A., 2024. Postharvest Biology and Technology Wide-characterization of high and low dry matter kiwifruit through spatiotemporal multi-omic approach. *Postharvest Biol. Technol.* 209, 112727.

Veberic, R., Colaric, M., Stampar, F., 2008. Phenolic acids and flavonoids of fig fruit (*Ficus carica* L.) in the northern Mediterranean region. *Food Chem.* 106, 153–157.

Viuda-Martos, M., Barber, X., Pérez-Álvarez, J.A., Fernández-López, J., 2015. Assessment of chemical, physico-chemical, techno-functional and antioxidant properties of fig (*Ficus carica* L.) powder co-products. *Ind. Crops Prod.* 69, 472–479.

Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., Gai, J., Yang, M., Zhu, M., Zhang, H., Ye, X., Gao, A., Tan, X., Wang, S., Wu, S., Cahoon, E.B., Bai, B., Zhao, Z., Li, Q., Wei, J., Chen, H., Luo, R., Gong, D., Tang, K., Zhang, B., Ni, Z., Huang, G., Hu, S., Chen, Y., 2020. The genome evolution and domestication of tropical fruit mango. *Genome Biol.* 21, 1–17.

Wang, Z., Cui, Y., Vainstein, A., Chen, S., Ma, H., 2017. Regulation of fig (*Ficus carica* L.) fruit color: Metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front. Plant Sci.* 8, 308052.

Wang, Ziran, Song, M., Wang, Zhe, Chen, S., Ma, H., 2021. Metabolome and transcriptome analysis of flavor components and flavonoid biosynthesis in fig female flower tissues (*Ficus carica* L.) after bagging. *BMC Plant Biol.* 21.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* (N. Y). 38, 1358–1370.

Xanthopoulou, A., Manioudaki, M., Bazakos, C., Kissoudis, C., Farsakoglou, A.M., Karagiannis, E., Michailidis, M., Polychroniadou, C., Zambounis, A., Kazantzis, K., Tsaftaris, A., Madesis, P., Aravanopoulos, F., Molassiotis, A., Ganopoulos, I., 2020. Whole genome re-sequencing of sweet cherry (*Prunus avium* L.) yields insights into genomic diversity of a fruit species. *Hortic. Res.* 7.

Yin, Y. chao, Zhang, X. Dong, Gao, Z. Qiang, Hu, T., Liu, Y., 2019. The Research Progress of Chalcone Isomerase (CHI) in Plants. *Mol. Biotechnol.* 61, 32–52.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.Y., 2017. Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. *Methods Ecol. Evol.* 8, 28–36.

Yu, Y., Fu, J., Xu, Y., Zhang, J., Ren, F., Zhao, H., Tian, S., Guo, W., Tu, X., Zhao, Jing, Jiang, D., Zhao, Jianbo, Wu, W., Wang, G., Ma, R., Jiang, Q., Wei, J., Xie, H., 2018. Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nat. Commun.* 9, 1–13.

Zahid, G., Aka Kaçar, Y., Dönmez, D., Küden, A., Giordani, T., 2022. Perspectives and recent progress of genome-wide association studies (GWAS) in fruits. *Mol. Biol. Rep.* 49, 5341–5352.

Zhang, Y., Nikolovski, N., Sorieul, M., Vellosillo, T., McFarlane, H.E., Dupree, R., Kesten, C., Schneider, R., Driemeier, C., Lathe, R., Lampugnani, E., Yu, X., Ivakov, A., Doblin, M.S., Mortimer, J.C., Brown, S.P., Persson, S., Dupree, P., 2016. Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in *Arabidopsis*. *Nat. Commun.* 7, 1–14.

Zhao, H., Sun, S., Ding, Y., Wang, Yue, Yue, X., Du, X., Wei, Q., Fan, G., Sun, H., Lou, Y., Yang, H., Wang, Jian, Xu, X., Li, L., Yang, K., Xu, H., Wang, Jiongliang, Zhu, C., Wang, S., Shan, X., Hou, Y., Wang, Yu, Fei, B., Liu, X., Jiang, Z., Gao, Z., 2021. Analysis of 427 genomes reveals moso bamboo population structure and genetic basis of property traits. *Nat. Commun.* 12, 1–12.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328.

Zohary, D., Spiegel-Roy, P., 1975. Beginnings of fruit growing in the Old World. *Science* 187, 319–327.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Bazakos, C., Michailidis, M., Tourvas, N., Alexiou, K.G., Mellidou, I., Polychroniadou, C. et al. (2024) Genetic mosaic of the Mediterranean fig: comprehensive genomic insights from a gene bank collection. *Physiologia Plantarum*, 176(4), e14482. Available from: https://doi.org/10.1111/ppl.14482