BMC Plant Biology

RESEARCH

Open Access

2b-RAD sequencing of *Malus florentina* populations reveals strong population structure and signals of balancing selection at disease resistance loci

David Papp[1*], Viktor Papp[2], Csaba Kozma[1], Sadia Tabassum Tanni[1], Fabio Conti[3], Tamás Deák[4] and Zoltán Barina[5]

Abstract

Background Crop wild relatives of *Malus domestica* represent exceptional value as breeding resources and ornamental trees, nonetheless their natural populations are under threat due to anthropogenic factors. *Malus florentina* is an apple species endemic to the Balkans and Italy, and our current knowledge on its population genetics and conservation biological status is strongly limited.

Results In the current study we used a 2b-RAD-seq approach to investigate the population genetic structure and diversity of four *M. florentina* populations from three different countries. We detected strong genetic differentiation among populations and a substantial risk of inbreeding within them. Furthermore, we identified candidate genes under selection associated with various stress responses including biotic resistance, salt stress and drought stress tolerance, further supporting local adaptation. Our enrichment analyses based on non-synonymous SNP counts at resistance loci revealed a significant 2.4-fold change, indicating that disease resistance genes are under selection.

Conclusions According to this study *M. florentina* populations face considerable risk, and fragmentation of the studied populations cannot be ruled out. Despite this, the assessed populations exhibit considerable diversity at disease resistance loci, likely maintained by balancing selection. These findings highlight the importance of conserving the genetic diversity of wild apple populations for key agronomic traits such as disease resistance.

Keywords Crop wild relatives, Eriolobus florentinus, Apple scab, Erwinia amylovora, Frequency-based selection

*Correspondence:
David Papp
papp.david@uni-mate.hu
[1]Department of Fruit Growing, Hungarian University of Agriculture and Life Sciences, Budapest, Hungary
[2]Department of Botany, Hungarian University of Agriculture and Life Sciences, Budapest, Hungary
[3]Scuola di Bioscienze e Medicina Veterinaria, Università di Camerino, Barisciano, Italy
[4]Department of Viticulture, Hungarian University of Agriculture and Life Sciences, Budapest, Hungary
[5]World Wild Fund for Nature Hungary, Budapest, Hungary

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Background

Wild relatives of the domestic apple (*Malus domestica*) are important genetic resources and are contributing significantly to biodiversity in their natural habitats. The genus *Malus* comprises approximately 40 recognized species; however, its taxonomy is considered unresolved due to extensive interspecific hybridization, unclear species and genus-level boundaries [1–3]. Apple breeding programs utilize these wild *Malus* spp. primarily to enhance cultivar resistance to biotic stressors and, to a lesser extent, to improve ecological adaptability. Resistance genes and Quantitative trait loci (QTLs) have been identified in various apple species for major apple diseases such as apple scab (*Venturia inaequalis*) and fire blight (*Erwinia amylovora*). Notable examples include *M. floribunda* (*Rvi6*, *Rvi7*, *FB_Mf*), *M. sieversii* (*Rvi8*, *Msv_FB7*), *M. baccata* (*Rvi11*, *Rvi12*, *FB_Mb*), *M. robusta* (*FB_Mr5*), *M. fusca* (*FB_Mfu10*), and *M. arnoldiana* (*FB_Mar12*) [4, 5]. Wild relatives also represent valuable sources of resistance to a range of other diseases and pests, such as powdery mildew, cedar apple rust, Alternaria leaf spot, blue mold, bitter rot, replant disease, Valsa canker, apple blotch, woolly apple aphid, apple maggot fly, etc [6–9]. In addition to their value in crop breeding *Malus* spp. are among the most important ornamental small trees of the Northern Hemisphere, have potential as rootstocks, pollinators, and their crabby fruits can serve as healthy ingredients of processed foods and drinks [10, 11].

Despite their versatility and economic relevance, wild *Malus* species are threatened globally, with natural populations facing extinction [1, 12]. Notably, in Central Asia, the main progenitor of the domestic apple *M. sieversii* is already on Red List (IUCN), and its populations are experiencing rapid declines due to habitat loss from agricultural expansion, overgrazing, aridification, and widespread hybridisation with *M. domestica*. These have led to severe fragmentation of the natural populations [13]. Similarly, *M. sylvestris*, the most common apple species in Europe has already been included in the regional Red List of many countries [14].

The endemic *M. florentina* has a limited distribution to the Balkan Peninsula, Italy, and to a minimal extent to Anatolia, and is considered endangered or vulnerable in most of these countries. It is most common in North Macedonia and Italy, but there are reports from Serbia, Albania, Greece, and Turkey [15, 16], and most recently from Bulgaria [17]. Recently, new localities have been documented in Italy [18], Greece [19], Serbia [20], Albania [21], and Kosovo [22], but data remain scarce regarding its continued presence at historically reported sites, and thus, its exact and up to date distribution remains unclear.

It is occasionally cultivated as an ornamental tree in North America and Europe, valued for its distinctive hawthorn-like leaves, attractive bark, rounded or upright crown shape, and late-season flowering [17, 23, 24]. Although previously considered susceptible to major apple diseases (23), *M. florentina* has not shown susceptibility to *Erwinia amylovora* in Europe (24). A survey of a cultivar collection in Geneva, New York, identified one of its specimens (PI588868) as highly resistant to apple scab (*Venturia inaequalis*) [25]. It also exhibited moderate resistance to apple blotch (*Diplocarpon mali*) [26]. Furthermore, the *M. florentina* accession 'Skopje P2' (PI589385) was reported to be resistant to *Rosellinia necatrix* [27].

Malus florentina has long been the subject of taxonomic uncertainty and conflicting classifications. First described as *Pyrus florentina* in 1806, it since has been proposed to be an interspecific hybrid, it has been recombined or renamed seven times by various taxonomists placed in eight different genera: *Pyrus*, *Crataegus*, *Mespilus*, *Sorbus*, ×*Malosorbus*, *Torminaria*, *Eriolobus* and *Malus* [20]. Currently the most accepted name is *Malus florentina* (*Malus* sect. *Florentinae*) [28, 29], although its designation remains debated in the modern genetic era. Recent molecular analyses have revealed that *M. florentina* is closer to North American apple species (*M. coronaria* and *M. angustifolia*, and *M. ioensis*) and the recently recombined former *Docynia* species (e.g. *Malus indica*) than to Eurasian apples [29, 30]. Most recently a genetic similarity with *Cydonia* has also been proposed, together with the recombination of *M. florentina* and its sister species *M. trilobata* to the genus *Eriolobus* [3]. Despite these uncertainties in the current paper we use the widely accepted name *Malus florentina* (Zucc.) C.K.Schneid.

Although several genetic studies have been conducted on *M. florentina* to clarify its taxonomic status, the population genetics of its natural populations has not yet been studied to date. However, a study using microsatellite markers (SSRs) has been conducted on its similarly threatened sister species *M. trilobata*. The study on a total of 69 individuals of five Greek populations revealed a fragmented population structure and patterns of isolation by distance, typical to peripheral populations [31].

Traditionally, population genetic studies have relied on SSRs, which are effective for assessing genetic structure but offer no, or limited information on the functionality of the variation. In contrast, restriction site associated DNA sequencing (RAD-seq) provides comparable accuracy, and being more powerful is more reliable for populations with low sample numbers, while also allows the identification of single nucleotide polymorphisms (SNPs) linked to functional genes [32, 33]. Among RAD-seq approaches 2b-RAD-seq offers a flexible and cheap solution by using type IIB restriction enzymes to produce fixed-length, short fragments through a relatively

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

simple protocol [34, 35]. Although it may yield significantly lower genome coverage than usual RAD-seq techniques, it requires low amount of DNA-input, and results in controllable and uniform coverage with less bias.

Although 2b-RAD-seq is especially suitable for studying non-model organisms, high-quality reference genomes greatly enhance the potential of in-depth functional analyses. For *M. florentina* no genomic resources were available until recently. The construction of a *Malus* pan-genome, however, now includes a high-quality assembly of *M. florentina*, providing opportunities for more detailed evolutionary and functional studies of the species [36].

Understanding genetic diversity and adaptive potential of the threatened *M. florentina* is crucial from many perspectives. It might support decision-making in management, conservation, and selection work for direct use or breeding. To detect patterns of balancing selection at disease resistance loci (specifically nucleotide-binding leucine-rich repeat receptors or NLRs) reduced representation sequencing has been used in several other species [37–39]. The elevated diversity of resistance genes in natural populations, resulting from the host-pathogen arms race, is crucial for identifying valuable gene sources and strongly supports the case of in situ conservation; particularly in clonally propagated species such as apples.

For these reasons, in the present study we analyzed the population genetic diversity and structure of four *M. florentina* populations from three countries (Albania, Bulgaria, and Italy) using a 2b-RAD-seq approach. In addition, we identified and functionally annotated candidate genes under selection, which may indicate local adaptation. We also performed SNP enrichment analysis on disease resistance loci to detect signatures of balancing selection and to assess the genetic potential of the populations.

Table 1 Summary of *Malus florentina* populations sampled in this study

Pop name	Country	Nearest city	Long (°E)	Lat (°N)	Reference (first report)
ALB1	Albania	Boshanj	20.60189	40.43242	Barina et al. 2010
ALB2	Albania	Leskovik	20.64644	40.17717	Barina et al. 2010
BLG	Bulgaria	Boboshevo	22.96119	42.17706	Aneva and Zhelev, 2023
ITA	Italy	Collelongo	13.59583	41.91250	Conti et al. 2006

Materials and methods

Summary of historical sites of *M. florentina*

Prior sample collection we summarized previously reported sites of *M. florentina*. Georeferenced points were manually extracted from historical distribution maps using the imager R package [40], and a coordinate transformation was applied to generate a consensus map. During our expeditions of the last two decades new sites of *M. florentina* were noted from 2006 to 2025.

Sample collection

We sampled 41 trees from four previously reported *M. florentina* populations in August 2024. The sampling sites were located in Albania (ALB1 and ALB2, n = 9 and n = 12, respectively) [21], Bulgaria (BLG, n = 9) [17], and Italy (ITA, n = 11) [18]. Leaf samples (3–4 per tree) were collected from wild trees (with minimum 10 m between trees) and kept under humid conditions during transport, typically in an electric cooler. *Malus florentina* is not listed by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, www.cites.org), nor was it subject to local legislation. Additionally, sampling was conducted outside protected areas (Table 1).

Species identification relied on morphological traits described in previous works [17, 28] and was done by the authors (by Csaba Kozma and David Papp, or in case of ITA samples by Fabio Conti). Key distinguishing features included an upright, apple-like growth habit with hawthorn-like leaves, and patchy bark. (Fig. 1).

DNA extraction

When the leaf samples arrived at the laboratory, they were frozen to −80 °C. Frozen samples were ground in a mortar, and then DNA was extracted using the E.Z.N.A. Plant DNA Kit (Norcross, GA, USA) following the manufacturer's instructions. The concentrations and integrity of the extracted DNA samples were determined by a NanoDrop spectrophotometer (NanoDrop ND-1000; Thermo Fisher Scientific, Waltham, MA, USA), and by electrophoresis on 1% agarose (Sigma-Aldrich, St. Louis, MO, USA) gel, using TAE buffer.

Library preparation, sequencing, and quality control

Library preparation and raw read processing followed the 2b-RAD-seq protocol described by Wang et al. [35]. Genomic DNA of 100 ng was digested with the Type IIB restriction enzyme BsaXI, which recognizes the sequence 5′-N_9ACN$_5$CTCCN$_{10}$–3′ and cleaves on both sides to generate fragments of 32–36 bp with 3-base 3′ overhangs. Adaptors were ligated to the cohesive ends, and five RAD tags were concatenated per library, each corresponding to an individual sample. Barcoded libraries were amplified by PCR and underwent size selection to

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

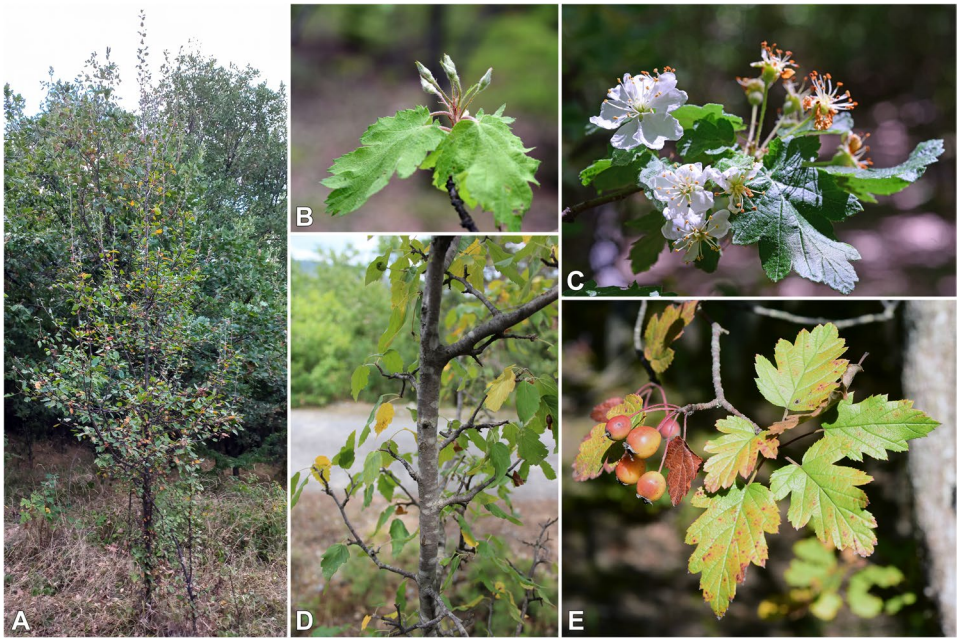


Fig. 1 Morphology of *M. florentina* in the studied populations: upright crown shape of soliter trees (A), Hawthorn-like leaves and small, 2–3 cm large generative organs (B, C, E), and typical patchy bark (D). Photo credit: D. Papp (A, D), F. Conti (B, C, E)

enrich fragments of approx. 320 bp, followed by sequencing on an Illumina NovaSeq platform using 150 bp paired-end reads (PE150).

Raw reads went through demultiplexing, trimming, and filtering using built-in Perl scripts. The five concatenated RAD tags were separated and assigned to each sample. Then reads containing the restriction enzyme recognition sites were extracted. Finally, low-quality sequences, adapter contamination, and other artifacts were removed by trimming the adapters, and by removing reads with more than 8% undetermined bases (N) or more than 15% of bases with a Phred quality score below 30. Sequencing and QC was performed at CD Genomics (Shirley, NY, USA, https://www.cd-genomics.com).

Read alignment and variant calling

Cleaned reads (approx. 28 bp long) were aligned to the *M. florentina* consensus genome assembly (accession number: GWHCAWS00000000, bioproject: PRJCA015452, referred to as *Eriolobus florentinus*) using Bowtie v1.3.1 and the Rbowtie package in R v4.3.3 [36, 41]. Resulting SAM files were converted to BAM format, sorted, and indexed using Samtools v1.21 and the Rsamtools package [42]. For quality control, BAM files were used to calculate mapping ratio, genome coverage, and sequencing depth.

To identify low-quality samples, cutoff values were determined using the interquartile range (IQR) method. For each metric, the first quartile (Q1) and third quartile (Q3) were calculated, and the IQR was defined as Q3-Q1. Samples with values below Q1-1.5×IQR (mapping ratio < 66.48%, and genome coverage < 0.38%) were classified as outliers. No meaningful, negative cutoff value was determined for depth by the IQR method.

Variant calling was performed using FreeBayes v1.3.9 [43]. To obtain high-quality variants, only sites with a minimum coverage of 5 reads, a minimum alternate allele counts of 3, and an alternate allele frequency of at least 0.1 were kept. Complex haplotype calling was disabled, and only the top-ranking alternate allele per site was retained. Variant calling was run with the --report-monomorphic flag to include invariant sites. Where not specified otherwise, variants were filtered to include only those with a base quality ≥ 20 and a mapping quality ≥ 30. Furthermore, the raw variant calls were filtered using BCFtools v1.9 ("bcftools view -v snps -m2 -M2 -e 'TYPE!="snp"' input_file.vcf -o output_file.vcf -O v") and as a final step with VCFtools v0.1.16 ("vcftools --vcf input_file.vcf --max-missing 0.9 --min-alleles 2 --max-alleles 2 --recode --out output_file") [44, 45]. All indels, complex errors, and multiallelic sites were excluded to retain only biallelic SNPs, maximum 5% rare alleles and 10% missing data. Minor allele frequency (MAF) and linkage disequilibrium (LD) decay were calculated from the VCF file using VCFtools.

Population diversity and structure

Based on the variants called, whole-population and within-population genetic diversity indices were calculated using VCFtools. The following indices were determined: observed and expected heterozygosity (Ho, He), inbreeding coefficient (FI_S), Tajima's D, and nucleotide diversity (π). To calculate Tajima's D and nucleotide

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

diversity (π) we used 50 kb window size (inferred from the LD decay plot), and we first defined callable regions based on the raw VCF file.

Mean relatedness coefficients (P_HAT), estimating the proportion of alleles that two individuals share due to common ancestry, were calculated for all within-population pairwise comparisons using PLINK v1.9 [46]. In addition, we identified private alleles (alleles unique to a single population) using the private_alleles function in the poppr R package.

Population structure was analysed using ADMIXTURE v1.3.0 [47], DAPC (Discriminant Analysis of Principal Components) from the adegenet R package v2.1.11 [48], and pcadapt v4.3.3 [49]. For these analyses a stricter filtering was applied and outlier samples with low mapping ratio or depth were excluded. Variants were recalled with FreeBayes, and missing data were filtered using VCFtools (adding "--max-missing 1 and "--maf 0.05").

Pairwise FST values [50] were computed by VCFtools across all SNPs, giving six pairwise comparisons among the four populations. Isolation by distance (IBD) was tested by correlating genome-wide FST values and geographic distances among populations. Geographic distances were computed from the coordinates of the four populations using the Haversine formula through the geosphere v1.5–20 R package [51]. The significance of the correlation was determined by the Pearson method in a Mantel test with 999 permutations using the vegan package v2.6–10 in R [52].

Functional annotation of SNPs

For functional annotation SNPs were mapped to gene regions of the reference genome (gff3 file) using the GenomicRanges and rtracklayer R packages [53, 54]. Protein sequences containing one or more SNPs were functionally annotated using InterProScan v5.73 with all available databases [55]. In addition, SNPs were annotated for functional impact using SnpEff v5.0e [56]. Variants were classified into four impact classes (HIGH, MODERATE, LOW, and MODIFIER) according to default SnpEff terminology. To summarize functional annotations, we reported only Pfam signature descriptions of candidate genes containing SNPs under selection. Genes were considered candidates under selection if met al.l the following criteria: PCAdapt significance ($p \leq 10^{-5}$), ranked in the top 5% of mean pairwise FST values, and were classified as having MODERATE or HIGH impact by SnpEff.

Enrichment around resistance loci

We conducted an SNP frequency enrichment to assess signatures of selection near loci potentially involved in disease resistance. To assess whether regions around disease resistance loci have overrepresented diversity, SNP counts in genomic windows of ± 450 kb flanking NLRs (nucleotide-binding leucine-rich repeat) and known resistance genes (R-genes) were compared to randomly sampled control windows [37]. The nucleotide-binding leucine-rich repeat (NLR) genes of the reference genome were annotated using NLR-Annotator v2.1b [57]. To map known resistance genes towards major apple diseases (apple scab and fire blight) 16 flanking markers and gene sequences of characterized genes (Table S1) were blasted to the reference genome [38]. Alongside the full quality-filtered SNP set, we also run the analyses with SNPs filtered for having moderate or high impact based on the SnpEff annotation. The number of SNPs overlapping each NLR or *R*-gene flanking window was counted using the GenomicRanges and rtracklayer packages. To test for enrichment, a matched set of randomly placed windows (equal in number and size, sampled across chromosomes) was generated. A one-sided Wilcoxon rank-sum test and a permutation test (1000 iterations) were used to compare mean SNP counts between NLR or R-gene flanking and random windows. The circos plot was drawn with the circlize package [58], while all other visualization of the results was performed with ggplot2 v3.5.2 [59].

Results

Distribution summary

We have identified altogether 27 new localities of *M. florentina* from 2006 to 2025 in Albania (n = 23), Kosovo (n = 3), and in North Macedonia (n = 1). The total number of reports according to our new findings and our literature summary is 307 (Fig. 2, Table S2) [16–18, 20–22, 60–62].

Statistics of sequencing, QC and variant calling

In total, 41 samples were sequenced using the 2b-RAD method. A mean of 5.1 million raw reads were generated per sample, of which 3.0 million (59%) matched the expected recognition site of the BsaXI restriction enzyme. After quality filtering, a mean of 1.9 million clean reads per sample was retained, corresponding to 37.3% of the initial raw reads. Alignment statistics indicated sufficient quality for downstream analyses (Table S3). The mean mapping ratio to the reference genome was 85.9%, with a sequencing depth of 0.057× and genome coverage of 0.58% (per-locus depth ≈ 10).

Variant calling returned 6,903,982 unique variant sites across all samples. After filtering with BCFtools we have got 40,951 SNPs, reduced to a final 4,812 by the VCFtools filter. In this SNP set, the mean MAF across all SNPs was 0.25 ± 0.17; while LD was high at short distances ($r^2 \approx$ 0.50 at 1 kb) but decreases rapidly with distance, reaching near-background levels ($r^2 \approx 0.10$) by 50 kb (Fig. 3).

For population structure analyses we used an alternative, stronger filtering, and three outlier samples (with

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

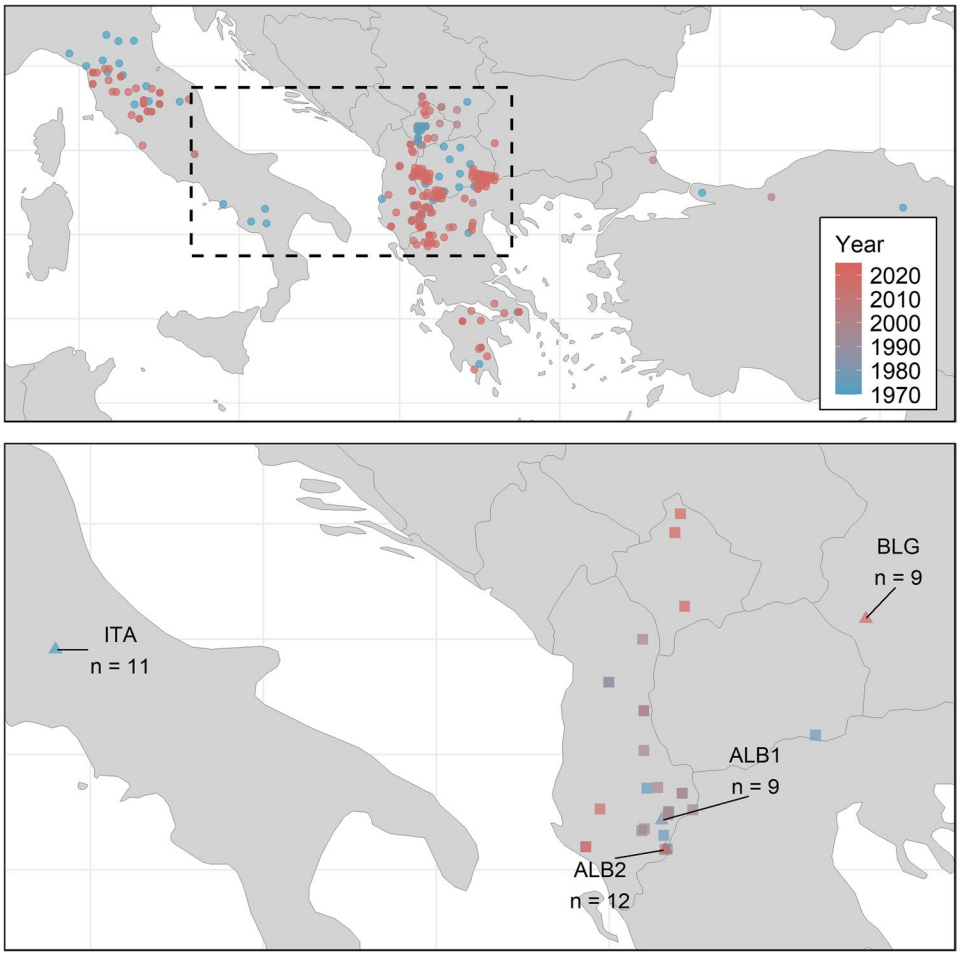


Fig. 2 Consensus map of *Malus florentina* populations summarizing the reported sites in the last half century (top) [16–18, 20–22, 60–62]; new locations reported in this study, and the four sampled historic sites (bottom, square: new locations, triangle: sampled historic sites); colors indicate the year in which the site was reported

low mapping ratio or low coverage) were excluded based on the cutoff values determined by the IQR method. Furthermore, variants were recalled with no monomorphic sites included resulting in 42,253 variants. BCFtools filter returned 40,801 SNPs and finally, a stronger VCFtools filter (only keeping variants with no missing data and with minor allele frequency ≥ 5%) returned a final 1,257 high-quality SNPs suitable for analysing population structure.

Genetic diversity indices

Genetic diversity analyses revealed moderate variation across all 41 individuals, with observed heterozygosity (Ho) of 0.35 ± 0.04 and expected heterozygosity (He) of 0.38, yielding a low but positive inbreeding coefficient (FI_S = 0.09 ± 0.12) that reflects deficit of heterozygotes. Genome-wide nucleotide diversity (π) averaged 0.34 ± 0.17, and Tajima's D in 50 kb windows was strongly positive (1.14 ± 0.85), indicating an excess of alleles at intermediate frequencies, consistent with demographic structure or balancing selection.

At the population level, all four populations displayed comparable patterns, with π ranging from 0.26 to 0.28 and Ho from 0.34 to 0.36, while He remained constant at 0.38 (Table 2). FI_S values were positive across populations (0.05–0.11), lowest in Bulgaria, again suggesting a general deficit of heterozygotes. Tajima's D values were slightly positive (0.04–0.07), further supporting an over-representation of intermediate-frequency alleles relative to neutrality expectations.

Although no clonal individuals were detected, population level PI_HAT values were high, especially in ALB2 population, where the mean relatedness (0.41) approaches that expected between full siblings (0.5).

A total of 431 private alleles were detected across populations, with the highest number in Italy (358), followed by Bulgaria (42) and Albania_1 (22), while Albania_3 exhibited only 9 private alleles.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

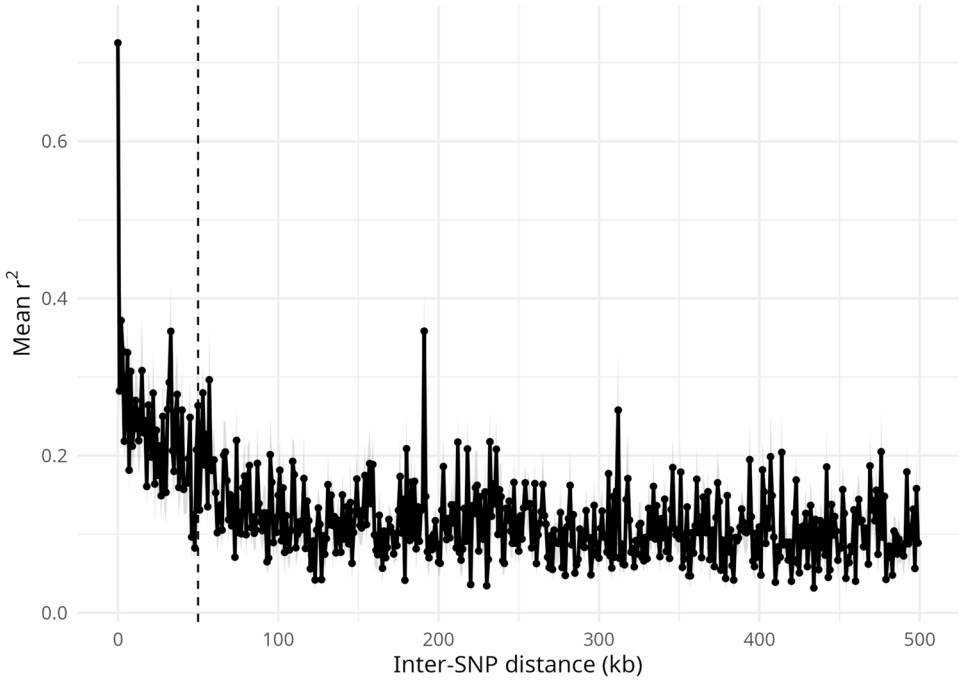


Fig. 3 Genome-wide linkage disequilibrium (LD) decay across 4,812 SNPs. Mean pairwise r^2 values (± SE) are plotted against physical distance between SNPs, binned in 1 kb intervals up to 500 kb. The vertical dashed line marks the 50 kb threshold, indicating the distance at which LD has largely decayed

Table 2 Population-level genetic diversity indices of four *Malus florentina* populations

Population	Albania 1	Albania 2	Bulgaria	Italy
Nr. of individuals	9.00	12.00	9.00	11.00
Ho mean	0.35	0.35	0.36	0.34
Ho SD	0.03	0.02	0.02	0.08
He mean	0.38	0.38	0.38	0.38
He SD	0.00	0.00	0.00	0.00
F_IS mean	0.10	0.09	0.05	0.11
FI_S SD	0.08	0.06	0.06	0.21
π mean	0.26	0.26	0.28	0.28
π SD	0.22	0.22	0.21	0.21
TajimaD mean[1]	0.06	0.07	0.07	0.04
TajimaD SD[1]	0.35	0.37	0.37	0.27
PI_HAT mean	0.36	0.41	0.29	0.20
PI_HAT SD	0.10	0.09	0.12	0.20

Ho: observed heterozygosity; He: expected heterozygosity; F_IS: inbreeding coefficient; π: genome wide nucleotide diversity; PI_HAT: pairwise relatedness coefficient; [1]TajimaD was calculated with 50 kb window size

Population structure

The PCA graph, showed strong distinction between all populations, except the two Albanian ones (Fig. 4A). The scree plot of PCAdapt suggested K = 2 (Fig. S1) and the strongest distinction could be observed alongside PC1 explaining 42.6% variance, distinguishing the Italian population from the Balkan ones. Similarly, the Balkan samples were clearly separated along PC2 (5.8%), whose distribution suggests a clinal variation.

A similar population structure was suggested by the DAPC analyses. With best K = 4 (inferred from the Bayesian Information Criterion or BIC plot) the analyses showed that all populations have individual clusters with the exception of the two admixed Albanian populations located 30 km-s from each other (Fig. 4B and C).

We explored the relationship between genetic and geographic distance (IBD) among populations using a Mantel test (Fig. S2). Although the result was not statistically significant ($p = 0.125$), likely due to the low number of populations ($n = 4$) and number of permutations ($n = 23$), a positive, strong correlation was observed (Mantel $r = 0.923$).

The lowest cross-validation error (0.53) in the ADMIXTURE analysis suggested two ancestral populations. Ancestry proportions at K = 2 appeared to correlate with geographic distance, supporting isolation by distance, except in the ALB2 population (Fig. 5). We additionally examined K = 3 and K = 4. K = 3 revealed an ancestral population largely restricted to the Balkan populations, and highlighted differences between populations of the two major geographic area. ALB2 again showed a notable deviation, lacking an ancestry common to all other populations. At K = 4 BGR population was revealed to be the most admixed population, while ALB2 was still lacking a common ancestry.

Functional annotation of candidate genes under selection

To identify candidate genes with SNPs under selection, we filtered for high PCAdapt significance ($p \leq 10^{-5}$), pairwise mean FST value (top 5%), and MODERATE or HIGH SnpEff impact (Table 3). All four candidate

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

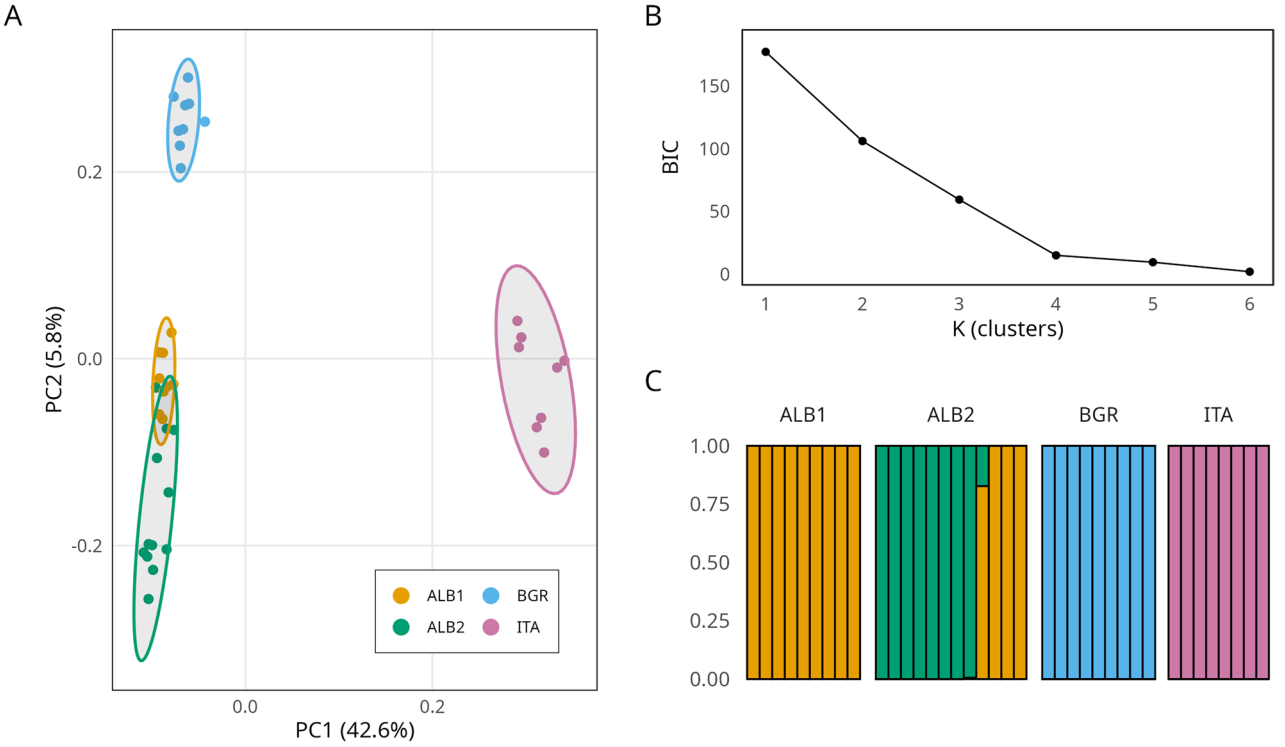


Fig. 4 Genetic structure of four *M. florentina* populations based on (**A**) a PCAdapt plot (with 90% confidence ellipses), (**B**) a Bayesian Information Criterion (BIC) plot to infer best K, and (**C**) a DAPC barplot with K = 4 genetic clusters

genes fulfilling these criteria had MODERATE impact class putatively causing missense variation. Almost half (47.8%) of the SNPs which were considered selected based on PCAdapt also had high FST values, providing strong support for divergent selection among populations. Four candidate genes we identified were (1) a ubiquitin-conjugating enzyme (IPR000608), (2) a gene with a PWWP domain (IPR000313), (3) a gene consisting of a protein kinase domain (IPR000719) and a NAF domain (IPR004041), likely encoding a CIPK-type serine/threonine protein kinase [63], and (4) an aspartyl-tRNA synthetase (IPR004364). A full annotation of all SNPs is presented in the supplementary files (Table S4).

For all four identified genes, the alternative allele is fixed in the Albanian populations, nearly fixed in the Bulgarian population (alternate allele frequency: 0.94 to 1), but shows variable frequencies in the Italian population ranging from 0 to 0.45, indicating population-specific differentiation.

SNP enrichment near disease resistance associated loci

To detect signatures of selection near disease resistance loci, we carried out an SNP frequency enrichment analysis by comparing ± 450 kb windows around resistance loci to randomly selected genomic regions with the same window sizes (Fig. 6).

To first detect NLR genes, we run NLR-Annotator on the reference genome, which returned 827 NLRs. The most common class was TIR-NB-LRR (40.9%), followed by CC-NB-LRR (27.6%) and NB-LRR (12.3%). The remaining 19.2% were distributed among other less common categories, such as the CC-TIR-NB-LRR and various incomplete NLR structures. To locate the position of *R*-genes/QTLs acting against the two major diseases of apple (apple scab and fire blight), we have blasted flanking marker regions and *R*-gene sequences to the reference genome. Highest quality BLAST results were in accordance with the expected chromosomes and positions for all the 16 genes.

Enrichment analyses were carried out with all SNPs and with SNPs filtered to MODERATE and HIGH impact, and with NLRs and R-genes/QTLs in all four combinations. The Wilcoxon test was significant in all cases ($p \leq 0.05$) except for the *R*-genes/QTLs tested with all SNPs ($p = 0.16$). However, the permutation test with 1000 random window sets was highly significant in all combinations ($p < 1 \times 10^{-4}$). Calculating with all SNPs we observed a 1.3 fold change of SNP count at NLR windows compared to random window sets. The fold change value increased to 1.6 for *R*-genes/QTLs, many of which is well characterized and known to have strong phenotypic effect. Filtering SNPs to have MODERATE or HIGH impact (SnpEff) further increased the fold changes. Windows of NLRs showed 1.7 fold change, while *R*-genes/QTLs had 2.4 times more SNPs in the ± 450 kb windows

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

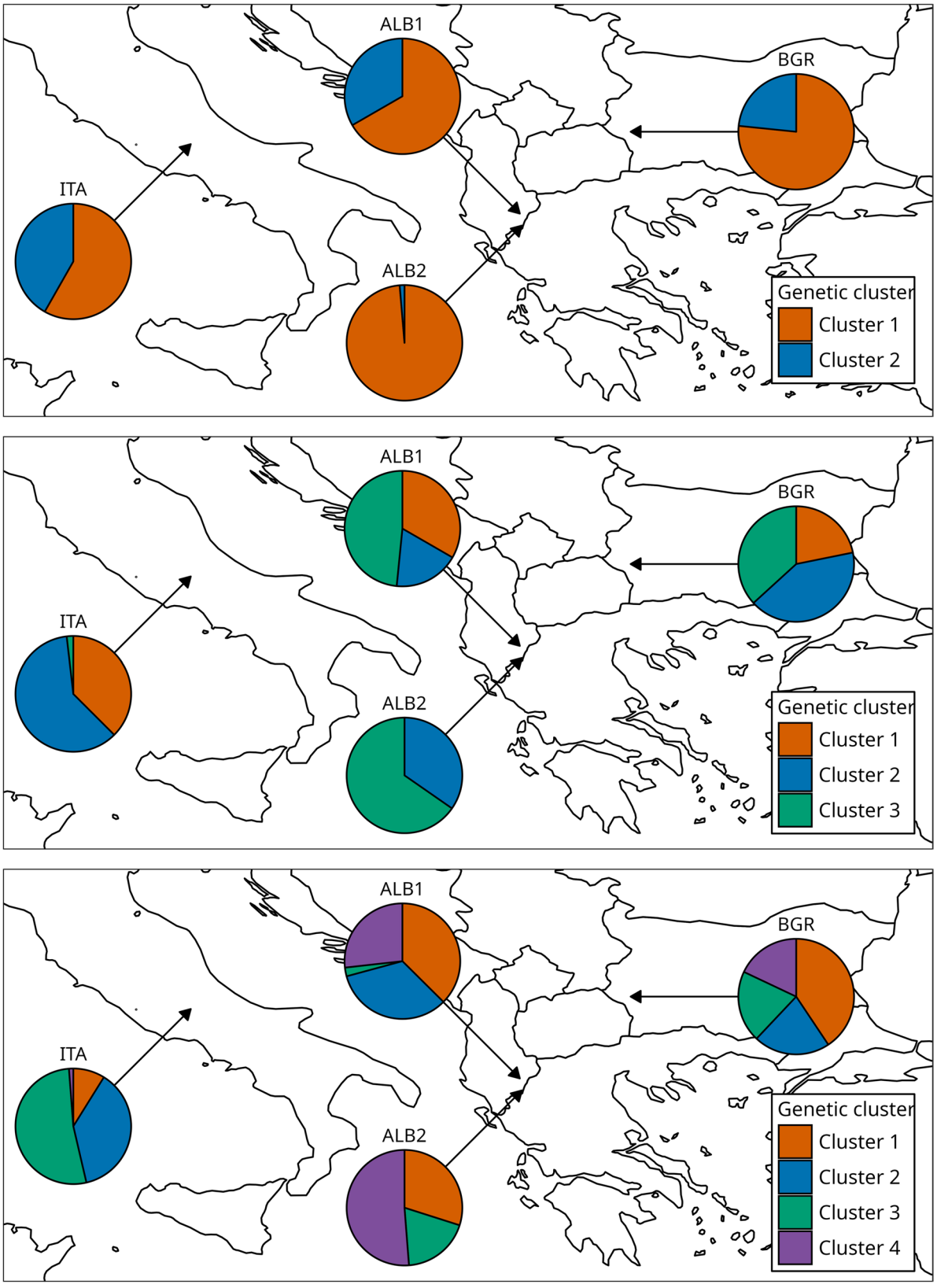


Fig. 5 Map of sampling locations and admixture plots for K = 2 (top), K = 3 (middle), and K = 4 (bottom) ancestral populations

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 3 Candidate genes containing SNPs under selection, identified by PCAdapt ($p \leq 10^{-5}$), high FST values (top 5%), and by MODERATE SnpEff impact class

Gene	Chr	Position	Ref	Alt	Signature desc.[1]	FST[2]	InterPro
g21807.t1	chr3	283,234	T	C	tRNA synthetases class II; OB-fold nucleic acid binding domain	0.5918	IPR004364;IPR004365
g22274.t1	chr3	4,690,237	A	G	Protein kinase domain; NAF domain	0.9451	IPR000719;IPR004041
g698.t1	chr6	36,176,126	T	C	PWWP domain	0.9404	IPR000313
g32667.t1	chr16	12,106,779	G	A	Ubiquitin-conjugating enzyme	0.5288	IPR000608

[1]Pfam database; [2]mean pairwise FST

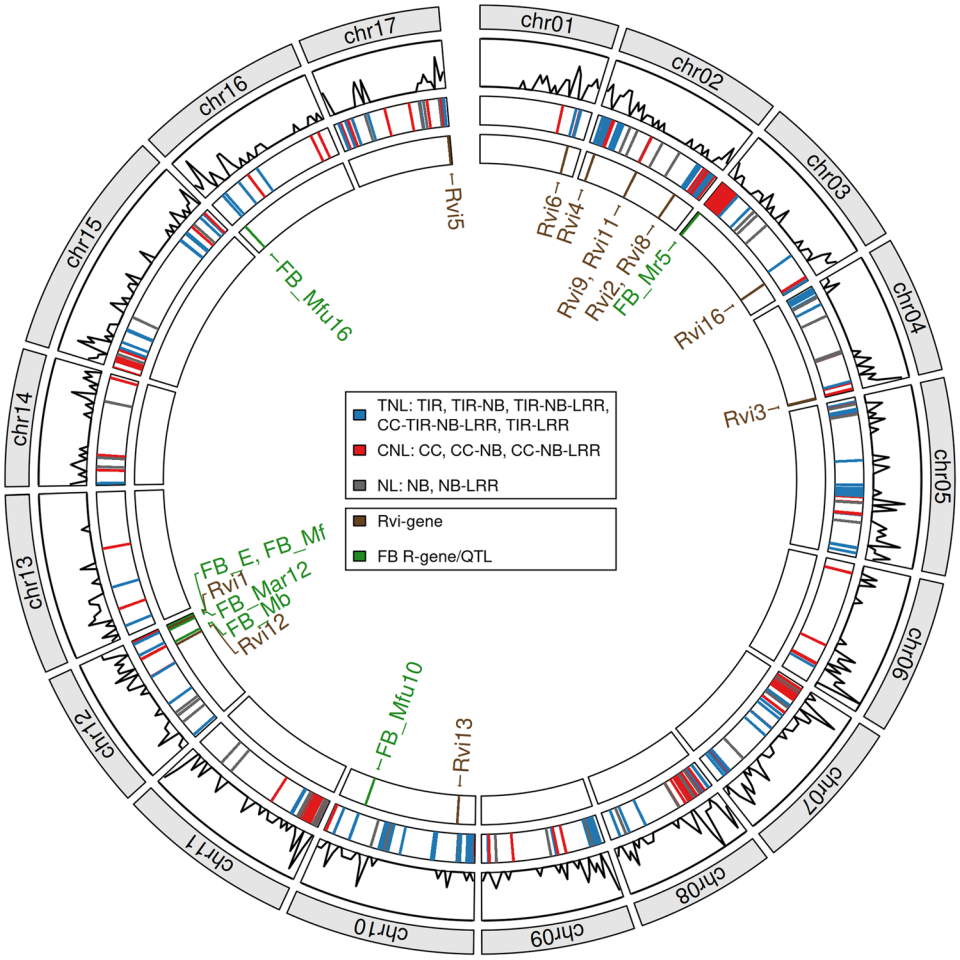


Fig. 6 Circos plot summarizing genetic variation at resistance loci in 41 *Malus florentina* samples; the outer ring shows the frequency of non-synonymous SNPs; the middle ring displays NLR types in ±450 kb windows (colors represent NLR types); the inner ring marks the positions of 16 flanking markers or functionally characterized R-genes/QTLs for apple scab (brown) and fire blight (green)

around them as the randomly selected windows of the same sizes.

Discussion

Habitat fragmentation caused by human activities can erode genetic diversity, reduce gene flow, and compromise the adaptive potential of wild plant populations, increasing the risk of extinction [64]. Wild *Malus* populations are also affected by habitat loss, and their conservation is a serious global concern [1, 12, 65]. The picture on their population genetic diversity and structure, however, is complex. Various studies, on different wild *Malus* species such as *M. sieversii*, *M. sylvestris*, *M. orientalis*, and *M. baccata* demonstrate weak population genetic structure and no strong isolation by distance, suggesting considerable gene flow among populations, and high within population genetic diversity; with encouraging cases of well-conserved wild apple populations in relatively intact forests [65–69]. Meanwhile some recent studies report that many natural wild apple populations are severely affected by habitat fragmentation, which poses a serious threat to their survival [13, 70]. One of the greatest

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

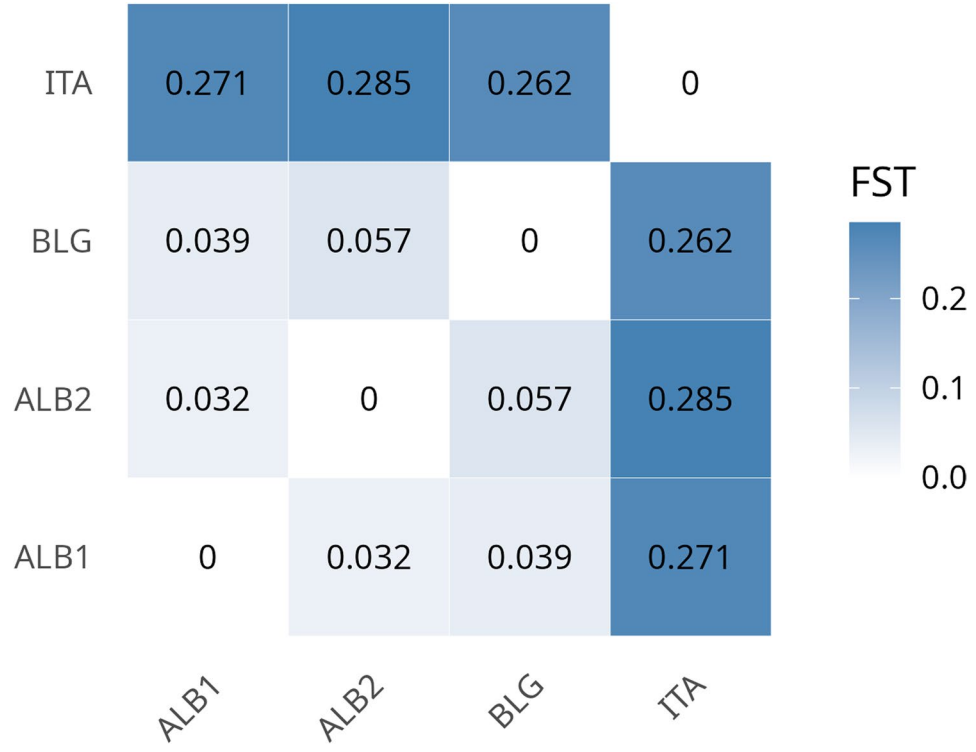


Fig. 7 Heatmap of pairwise FST values among the four studied populations; blue shading indicates higher genetic differentiation

concerns in protecting wild *Malus* germplasm is extensive hybridization with the domesticated apple [71, 72]. Crop-to-wild introgression, however, has smaller risk for *M. florentina*, for the larger genetic distance resulting in strong reproductive barriers [27].

The closest relative, *M. trilobata* has small, scattered populations similar in size and distribution to *M. florentina* and was shown to have fragmented population structure in Greece with strong isolation by distance effect. In addition to various threats, wildfires have recently decimated some of its areas suggesting that conservation efforts should account for fire management and climate change mitigation [31]. Besides this example, a predictive study demonstrated habitat loss for wild apples under future climate scenarios [13].

In this study we observed strong population structure for *M. florentina*, similarly what was shown for its sister species [31]. The analyses revealed pronounced genetic differentiation and limited gene flow between populations from different countries. In contrast, the two Albanian populations located 30 km from each other had substantial gene flow between them. Genetic differentiation among populations was positively correlated with geographic distance. As expected, pronounced differences were observed between the Italian and Balkan samples by various analyses in terms of genetic diversity, population structure, and pairwise genetic distances (Fig. 7).

When comparing the nucleotide diversity observed in our study with data on non-structured *Malus* populations [66], we found much lower π values (0.26 to 0.28 vs. 0.74 to 0.90), which may indicate a recent evolutionary bottleneck in the population history of *M. florentina*. This scenario is further supported by the positive Tajima's D values we observed, which may also reflect the action of balancing selection in these populations.

We analysed private allele counts which reflected the population structure, with the majority of private alleles identified in the Italian population, and the second largest number in the Bulgarian. The lowest number of private alleles and the highest pairwise relatedness values were found in the ALB2 population, which may reflect inbreeding, population fragmentation or localized reproduction.

Gametophytic self-incompatibility (SI) is a major factor driving high genetic diversity and heterozygosity in *Malus*, and its breakdown could lead to highly structured selfing populations. S-alleles underlying SI have been identified in most wild apple species, but not yet in *M. florentina* or its sister species *M. trilobata* [73]. Higher ploidy levels can sometimes relax SI, but cytogenetic studies have confirmed the diploid status of *M. florentina* [74]. While self-compatibility has been reported in some *Malus* accessions, including ornamental forms of wild species, it has not been documented at the population level in natural habitats [75]. Although widespread

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

self-compatibility in *M. florentina* appears unlikely, further controlled crossing experiments are necessary to exclude the possibility.

We identified four candidate genes under selection, which were associated with diverse range of functions in model species. The ubiquitin-conjugating enzyme (IPR000608) has many functions including DNA repair, cell cycle regulation, selective protein degradation, and stress response [76], making it a strong candidate for stress-related adaptation. The PWWP domain (IPR000313), possibly coding a DNA-binding protein functioning as a transcription factor, is associated with meristem development and flowering time; mutations leading to early flowering in *Arabidopsis* [77]. The CIPK-type serine/threonine protein kinase (IPR000719; IPR004041) is involved in calcium signalling pathways related to salt stress response in *A. thaliana* [63, 78]. Finally, the aspartyl-tRNA synthetase gene (IPR004364) has been shown to be involved in both BABA-induced disease immunity in *Arabidopsis* [79] and drought tolerance in poplar, where it co-occurs with a DNA-binding domain (IPR004365), similarly as in our case [80]. In line with the strong population structure these candidates suggest potential involvement in local adaptation. Other gene families, such as the MdAGO Argonaute family and MdMYB5 genes, have been extensively studied in apple and linked to drought and salt stress resistance, as well as local adaptation. In contrast, to the best of our knowledge, the functions of our candidate genes have not yet been characterized in apple, and our findings should therefore be considered indicative [36, 81].

Although, our results support a structured population, it remains unclear whether this population structure is the result of recent fragmentation or if it reflects a historically small and isolated distribution. This distinction has important implications for conservation biology, particularly in assessing extinction risk and designing appropriate management strategies [82]. Our sampling for genetic analyses focused on the main distribution area of *M. florentina* in Italy and parts of the Balkans, however, we acknowledge that some relevant regions where the species has also been reported, such as central Balkan areas, North Macedonia, and peripheral populations in Greece and Turkey were not included. Future studies that incorporate these populations with finer-resolution sampling may uncover additional patterns of genetic differentiation.

Although no formal survey was conducted to update the current distribution of *M. florentina*, we reported 27 new locations, in addition to summarizing the literature [16–18, 20–22, 60–62]. While more rigorous analyses are required to draw firm conclusions, the summarized data suggest that recently published reports originate from more southern regions compared to historical records.

Decline of abiotic and biotic resistance can be a sign of genetic erosion [83]. We have analysed evolutionary processes acting on the disease resistance of the populations. Our analysis revealed a pronounced enrichment of SNP variation in proximity to known disease resistance loci (NLRs and R-gene/QTL regions), with 2.4 fold change non-synonymous (HIGH or MODERATE) SNP density. This pattern is consistent with previous studies showing that NLRs carry disproportionately high diversity. For example, Buckley et al. [37] reported that genomic regions of elevated polymorphism in *A. lyrata* contained a significantly higher density of predicted NLRs than other regions. Gladieux et al. [39] found balancing selection at NLRs in traditional rice landraces, which harboured many ancient NLR haplotypes absent from modern varieties, with higher NLR diversity. In Rosaceae, many studies have reported signatures of both positive and balancing selection on S-alleles, which govern gametophytic incompatibility and are known to evolve in a manner similar to *R*-genes under frequency-dependent selection [84, 85]. It has been demonstrated in apple that NLRs show signatures of balancing selection and evolved under positive (Ka/Ks >1) selection [38]. Our observed excess of non-synonymous SNPs (MODERATE or HIGH) near resistance loci likely reflects such diversifying selection.

We detected these signals using 2b-RAD, a reduced-representation genotyping approach. Although RAD-seq was used before to detect this signal [37], it is worthwhile to note that more specifically 2b-RAD has significantly lower genome coverage than usual RAD-seq techniques. Despite of this, with a more simplistic approach using SNP count instead of indices (nucleotide diversity or TajimaD), we have shown similar selection patterns as previous studies [37–39].

Although our results support the presence of balancing selection at disease resistance loci in the studied populations, some potential confounding factors underlying elevated SNP diversity should be considered. Disease resistance genes are often organized in clusters of highly similar paralogous sequences, which can lead to misalignment of short reads, such as those generated by 2b-RAD, and thereby inflate SNP counts [86]. In addition, NLR loci are typically located near the ends of chromosomes, regions that are generally more polymorphic compared to central chromosomal regions, which may further contribute to the observed excess of SNPs [87].

Conclusion

In this study, we demonstrated that the sampled *M. florentina* populations exhibit strong genetic structure, like their sister species *M. trilobata*. The pronounced differentiation between Italian and Balkan samples should be considered for conservation efforts. Candidate genes

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

under selection include one potentially involved in salt stress tolerance and another linked to drought tolerance, suggesting possible adaptation to aridification, which might be linked to the changing climate conditions. We summarized and expanded the known distribution of *M. florentina*. Our data support that *M. florentina* populations are under threat; nonetheless, we observed high diversity at resistance loci. The elevated SNP counts in resistance loci compared to random windows might indicate balancing selection. These findings support that natural genetic reserves are particularly valuable for apple resistance breeding programs. In situ conservation or expanding the number of ex situ preserved accessions is especially important for *M. florentina*, since, to the best of our knowledge, germplasm collections currently maintain only a few accessions of this species.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07855-2.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7. Table S1 Resistance gene and linked marker sequences used for detecting functional disease resistant loci for apple scab and fire blightTable S2 Summarized data on locations of Malus florentinaTable S3 QC summary of the 41 samples after alignment: mapping ratio, depth, and coverageTable S4 Summary of annotation of all SNPs detected in the study

Acknowledgements
The authors are thankful for Enikő Tóth, for helping in the sample collection.

Authors' contributions
D.P. designed the work. Z.B., C.K., and D.P., summarized and expanded known locations of *M. florentina*. D.P., C.K., and F.C. collected the samples. S.T.T. and DP. carried out the lab work of the genetic analyses. D.P. performed the bioinformatic analyses. V.P. and D.P. prepared figures and tables. D.P. wrote the manuscript. F.C, Z.B., C.K., T.D., and V.P. revised the manuscript. All authors approved the final manuscript.

Funding
This research received no external funding.

Data availability
The sequencing datasets generated and analysed in this study are available in the National Library of Medicine database (www.ncbi.nlm.nih.gov) under BioProject PRJNA1286818, with accession numbers SRX29538978–SRX29539018.

Declarations

Ethics approval and consent to participate
No specific permits were required, sample collection was conducted in accordance with local conservation regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 June 2025 / Accepted: 26 November 2025
Published online: 15 December 2025

References

1. Volk GM, Chao CT, Norelli J, Brown SK, Fazio G, Peace C, et al. The vulnerability of US Apple (*Malus*) genetic resources. Genet Resour Crop Evol. 2015. https://doi.org/10.1007/s10722-014-0194-2.
2. Liu BB, Ren C, Kwak M, Hodel RG, Xu C, He J, et al. Phylogenomic conflict analyses in the apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving diversification, with insights into the complex biogeographic history in the Northern Hemisphere. J Integr Plant Biol. 2022. https://doi.org/10.1111/jipb.13246.
3. Sun J, Zhao D, Qiao P, Wang Y, Wu P, Wang K, et al. Phylogeny of genera in *Maleae* (Rosaceae) based on chloroplast genome analysis. Front Plant Sci. 2024. https://doi.org/10.3389/fpls.2024.1367645.
4. Bus VG, Rikkerink EH, Caffier V, Durel CE, Plummer KM. Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*. Annu Rev Phytopathol. 2011. https://doi.org/10.1146/annurev-phyto-072910-095339.
5. Peil A, Emeriewen OF, Khan A, Kostick S, Malnoy M. Status of fire blight resistance breeding in *Malus*. J Plant Pathol. 2021. https://doi.org/10.1007/s42161-020-00581-8.
6. Wöhner T, Emeriewen OF, Höfer M. Evidence of Apple blotch resistance in wild Apple germplasm (*Malus* spp.) accessions. Eur J Plant Pathol. 2021. https://doi.org/10.1007/s10658-020-02156-w.
7. Wen X, Yuan J, Bozorov TA, Waheed A, Kahar G, Haxim Y, et al. An efficient screening system of disease-resistant genes from wild apple, *Malus sieversii* in response to *Valsa Mali* pathogenic fungus. Plant Methods. 2023. https://doi.org/10.1186/s13007-023-01115-w.
8. Yu X, Hu K, Geng X, Cao L, Zhou T, Lin X, et al. The Mh-miR393a-TIR1 module regulates *Alternaria alternata* resistance of *Malus hupehensis* mainly by modulating the auxin signaling. Plant Sci. 2024. https://doi.org/10.1016/j.plantsci.2024.112008.
9. Tegtmeier R, Švara A, Gritsenko D, Khan A. *Malus sieversii*: a historical, genetic, and conservational perspective of the primary progenitor species of domesticated apples. Hortic Res. 2025. https://doi.org/10.1093/hr/uhae244.
10. Goncharovska I, Vladimyr K, Antonyuk G, Dan C, Sestras AF. Flower and fruit morphological characteristics of different crabapple genotypes of ornamental value. Not Sci Biol. 2022. https://doi.org/10.15835/nsb14110684.
11. Zelený L, Stryhalová G, Kozáková T. Usage of three Malus genotypes as pollinizers for chosen market Apple varieties. Vedecké Práce Ovocnárské. 2022;28(2):16–24.
12. Kell S, Qin H, Chen B, Ford-Lloyd B, Wei W, Kang D, Maxted N. China's crop wild relatives: diversity for agriculture and food security. Agric Ecosyst Environ. 2015. https://doi.org/10.1016/j.agee.2015.02.012.
13. Tian Z, Song H, Wang Y, Li J, Maimaiti M, Liu Z, et al. Wild apples are not that wild: conservation status and potential threats of *Malus sieversii* in the mountains of central Asia biodiversity hotspot. Diversity. 2022. https://doi.org/10.3390/d14060489.
14. Jacques D, Vandermijnsbrugge K, Lemaire S, Antofie A, Lateur M. Natural distribution and variability of wild Apple (*Malus sylvestris*) in Belgium. Belg J Bot. 2009;142(1):39–49.
15. Terpó A. *Malus* miller. In: Tutin TG, et al. editors. Flora Europaea. Cambridge: Cambridge Univ. Press; 1968.
16. Browicz K. „*Malus florentina*" - its history, systematic position and geographical distribution. Fragm florist, geobot. 1970;16(1):61–83.
17. Aneva I, Zhelev P. *Malus florentina* (Rosaceae): a new species for the Bulgarian dendroflora. Phytol Balcan. 2023;29(2):179–84.
18. Conti F, Bartolucci F, Catonica C, D'orazio G, Londrillo I, Manzi A, et al. Aggiunte alla flora d'Abruzzo. II contributo. Inf Bot Ital. 2006;38(1):113–6.
19. Christensen KI. ×*Malosorbus florentina* (Zuccagni). In: Phitos D, Strid A, Snogerup S, Greuter W, editors. The red data book of rare and threatened plants of Greece. Financed by the world wide fund for nature (WWF). Athens: Hellenic Botanical Society; 1995.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

20. Tomovic G, Niketic M, Stevanovic V. *Malosorbus florentina* (Rosaceae-Maloideae).-distribution, synecology and threatened status in Serbia. Phyton. 2003;43(2):295–306.
21. Barina Z, Pifkó D, Mesterházy A. Contributions to the flora of Albania. Willdenowia. 2010. https://doi.org/10.3372/wi.41.41118.
22. Prodanović D, Krivošej Z, Amidžić L, Bartula M. Floristic and chorological news from Northern Kosovo, in the Ibar river Valley. Nat Montenegrina. 2013;12(2):257–69.
23. Jacobson AL. North American landscape trees. Berkeley: Ten Speed; 1996.
24. Sutton J, Dunn N. '*Malus florentina*' from the website Trees and Shrubs Online. 2021. treesandshrubsonline.org/articles/malus/malus-florentina/. Accessed 24 Jun 2025.
25. Papp D, Gao L, Thapa R, Olmstead D, Khan A. Field apple scab susceptibility of a diverse *Malus* germplasm collection identifies potential sources of resistance for apple breeding. CABI Agric Biosci. 2020. https://doi.org/10.1186/s43170-020-00017-4.
26. Abe K, Moriya S, Iwanami H, Kanematsu S. Evaluation of resistance to apple blotch (*Diplocarpon mali*) in *Malus* species and varieties. Horticultural Research (Japan). 2020. https://doi.org/10.2503/hrj.19.151.
27. Lee S, Ko K, Aldwinckle HS. Resistance of selected *Malus* germplasm to *Rosellinia necatrix*. J Am Pomol. 2000;54(4):219–28.
28. Qian G, Liu L, Hong D, Tang G. Taxonomic study of <Emphasis Type="Italic">Malus</Emphasis>section <Emphasis Type="Italic">Florentinae</Emphasis>(Rosaceae). Bot J Linn Soc. 2008. https://doi.org/10.1111/j.1095-8339.2008.00841.x.
29. Liu GN, Ma DK, Xu C, Huang J, Ge BJ, Luo Q, et al. <Emphasis Type="Italic">Malus</Emphasis>includes <Emphasis Type="Italic">Docynia</Emphasis>(<Emphasis Type="Italic">Maleae</Emphasis>, Rosaceae): evidence from phylogenomics and morphology. PhytoKeys. 2023. https://doi.org/10.3897/phytokeys.229.103888.
30. Chen P, Li Z, Zhang D, Shen W, Xie Y, Zhang J, et al. Insights into the effect of human civilization on *Malus* evolution and domestication. Plant Biotechnol J. 2021. https://doi.org/10.1111/pbi.13648.
31. Balaska K, Adamidis GC, Varsamis G, Papalazarou K, Karanikola P, Manolis A, et al. Genetic diversity and spatial structure of the imperiled European population of *Malus trilobata* in Greece. Genetika. 2021. https://doi.org/10.2298/GENSR2101103B.
32. Lemopoulos A, Prokkola JM, Uusi-Heikkilä S, Vasemägi A, Huusko A, Hyvärinen P, et al. Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—implications for brown trout conservation. Ecol Evol. 2019. https://doi.org/10.1002/ece3.4905.
33. Sunde J, Yıldırım Y, Tibblin P, Forsman A. Comparing the performance of microsatellites and RADseq in population genetic studies: analysis of data for Pike (*Esox lucius*) and a synthesis of previous studies. Front Genet. 2020. https://doi.org/10.3389/fgene.2020.00218.
34. Wang S, Meyer E, McKay JK, Matz MV. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Methods. 2012. https://doi.org/10.1038/nmeth.2023.
35. Wang S, Liu P, Lv J, Li Y, Cheng T, Zhang L, et al. Serial sequencing of isolength RAD tags for cost-efficient genome-wide profiling of genetic and epigenetic variations. Nat Protoc. 2016. https://doi.org/10.1038/nprot.2016.133.
36. Li W, Chu C, Zhang T, Sun H, Wang S, Liu Z. Pan-genome analysis reveals the evolution and diversity of *Malus*. Nat Genet. 2025. https://doi.org/10.1038/s41588-025-02166-6.
37. Buckley J, Holub EB, Koch MA, Vergeer P, Mable BK. Restriction associated DNA-genotyping at multiple spatial scales in *Arabidopsis lyrata* reveals signatures of pathogen-mediated selection. BMC Genomics. 2018. https://doi.org/10.1186/s12864-018-4806-7.
38. Singh J, Sun M, Cannon SB, Wu J, Khan A. An accumulation of genetic variation and selection across the disease-related genes during Apple domestication. Tree Genet Genomes. 2021. https://doi.org/10.1007/s11295-021-01510-1.
39. Gladieux P, van Oosterhout C, Fairhead S, Jouet A, Ortiz D, Ravel S, et al. Extensive immune receptor repertoire diversity in disease-resistant rice landraces. Curr Biol. 2024. https://doi.org/10.1016/j.cub.2024.07.061.
40. Barthelme S. Imager: an R package for image processing based on CImg. J Open Source Softw. 2025. https://doi.org/10.32614/CRAN.package.imager.
41. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009. https://doi.org/10.1186/gb-2009-10-3-r25.
42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. Bioinformatics. 2009. https://doi.org/10.1093/bioinformatics/btp352.
43. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. https://arxiv.org/abs/1207.3907. Accessed 24 Jun 2025.
44. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA. The variant call format and vcftools. Bioinformatics. 2011. https://doi.org/10.1093/bioinformatics/btr330.
45. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021. https://doi.org/10.1093/gigascience/giab008.
46. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015. https://doi.org/10.1186/s13742-015-0047-8.
47. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009. https://doi.org/10.1101/gr.094052.109.
48. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008. https://doi.org/10.1093/bioinformatics/btn129.
49. Luu K, Bazin E, Blum MGB. Pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 2017. https://doi.org/10.1111/1755-0998.12592.
50. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984. https://doi.org/10.2307/2408641.
51. Hijmans R, Karney C, Williams E, Vennes C. geosphere: spherical trigonometry. R package version 1.5–18. 2023. https://rdrr.io/cran/geosphere/. Accessed 24 Jun 2025.
52. Oksanen J. Vegan: community ecology package. R package version, 2, 3. 2015. https://rdrr.io/rforge/vegan/. Accessed 24 Jun 2025.
53. Lawrence M, Gentleman R, Carey V. Rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009. https://doi.org/10.1093/bioinformatics/btp328.
54. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013. https://doi.org/10.1371/journal.pcbi.1003118.
55. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014. https://doi.org/10.1093/bioinformatics/btu031.
56. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly. 2012. https://doi.org/10.4161/fly.19695.
57. Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek HJ, Yu G, et al. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. Plant Physiol. 2020. https://doi.org/10.1104/pp.19.01273.
58. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. 2014. https://opus.bibliothek.uni-augsburg.de/opus4/files/79292/79292(2).pdf. Accessed 24 Jun 2025.
59. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New-York: Springer-Verlag; 2016.
60. Maccaroni L. Studio del popolamento di *Malus florentina* e dei principali driver ecologici responsabili del suo assetto all'interno della riserva naturale dell'Abbadia di Fiastra (MC). 2018. https://hdl.handle.net/20.500.12075/5947. Accessed 24 Jun 2025.
61. Monnet AC, Cilleros K, Médail F, Albassatneh MC, Arroyo J, Bacchetta G, et al. WOODIV, a database of occurrences, functional traits, and phylogenetic data for all Euro-Mediterranean trees. Sci Data. 2021. https://doi.org/10.1038/s41597-021-00911-0.
62. Conti F, Cancellieri L, Cangelmi G, Filibeck G, Rosati L, Bartolucci F. New records of native and alien vascular plants from Abruzzo, Lazio and molise National park (Italy) – and additions to the flora of Abruzzo and molise administrative regions. Ann Bot. 2022. https://doi.org/10.13133/2239-3129/17631.
63. Albrecht V, Ritz O, Linder S, Harter K, Kudla J. The NAF domain defines a novel protein-protein interaction module conserved in Ca2+-regulated kinases. EMBO J. 2001. https://doi.org/10.1093/emboj/20.5.1051.
64. Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Mol Ecol. 2008. https://doi.org/10.1111/j.1365-294X.2008.03971.x.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

65. Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P. The domestication and evolutionary ecology of apples. Trends Genet. 2014. https://doi.org/10.1016/j.tig.2013.10.002.
66. Schnitzler A, Arnold C, Cornille A, Bachmann O, Schnitzler C. Wild European apple (<Emphasis Type="Italic">Malus sylvestris</Emphasis>(L.) Mill.) population dynamics: insight from genetics and ecology in the Rhine Valley. Priorities for a future conservation programme. PLoS One. 2014. https://doi.org/10.1371/journal.pone.0096596.
67. Gao Y, Wang K, Wang D, Zhao J, Zhang C, Cong P. The genetic diversity and population structure analysis on *Malus baccata* (L.) borkh from 7 Sources. Sci Agric Sin. 2018;51(19):3766–77.
68. Kumar C, Singh SK, Singh R, Pramanick KK, Verma MK, Srivastav M et al. Genetic diversity and population structure analysis of wild *Malus* genotypes including the crabapples (*M. baccata* (L.) Borkh. & *M. sikkimensis* (Wenzig). Genet. Resour. Crop Evol. 2019. https://doi.org/10.1007/s10722-019-00780-y
69. Ha YH, Oh SH, Lee SR. Genetic admixture in the population of wild Apple (*Malus sieversii*) from the Tien Shan Mountains, Kazakhstan. Genes. 2021. https://doi.org/10.3390/genes12010104.
70. Bitz L, Heinonen M, Moisander S, Tanhuanpää P, Sarvarinne J. Differential allelic richness between *Malus sylvestris* L. and *Malus* × *domestica* Borkh. from Finland as a measure of genetic loss. Sustainability. 2019. https://doi.org/10.3390/su11246949.
71. Feurtey A, Cornille A, Shykoff JA, Snirc A, Giraud T. Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: towards a comprehensive conservation strategy of the wild Apple in Europe. Evol Appl. 2017. https://doi.org/10.1111/eva.12441.
72. Ruhsam M, Jessop W, Cornille A, Renny J, Worrell R. Crop-to-wild introgression in the European wild Apple *Malus sylvestris* in Northern Britain. Forestry. 2019. https://doi.org/10.1093/forestry/cpy033.
73. Liu Z, Gao Y, Wang K, Feng J, Sun S, Lu X, et al. Identification of S-RNase genotype and analysis of its origin and evolutionary patterns in *Malus* plants. J Integr Agric. 2024;23(4):1205–21. https://doi.org/10.1016/j.jia.2024.01.014.
74. Höfer M, Meister A. Genome size variation in <Emphasis Type="Italic">Malus</Emphasis>species. J Bot. 2010;1:480873. https://doi.org/10.1155/2010/480873.
75. Wu R, Chen X, Xia B, Yang Y, Molloy C, Wang R, et al. Analysis on unveiling the natural dynamics of parthenocarpy and self-compatibility in Apple trees. Horticulturae. 2024;10(12):1261. https://doi.org/10.3390/horticulturae10121261.
76. Jentsch S, Seufert W, Sommer T, Reins HA. Ubiquitin-conjugating enzymes: novel regulators of eukaryotic cells. Trends Biochem Sci. 1990. https://doi.org/10.1016/0968-0004(90)90161-4.
77. Hohenstatt ML, Mikulski P, Komarynets O, Klose C, Kycia I, Jeltsch A, et al. PWWP-domain interactor of polycombs1 interacts with Polycomb-Group proteins and histones and regulates *Arabidopsis* flowering and development. Plant Cell. 2018. https://doi.org/10.1105/tpc.17.00117.
78. Gandhi A, Oelmüller R. Emerging roles of receptor-like protein kinases in plant response to abiotic stresses. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms241914762.
79. Luna E, Van Hulten M, Zhang Y, Berkowitz O, López A, Pétriacq P. Plant perception of β-aminobutyric acid is mediated by an aspartyl-tRNA synthetase. Nat Chem Biol. 2014. https://doi.org/10.1038/nchembio.1520.
80. Feng CH, Niu MX, Zhao S, Guo S, Yin W, Xia X, Su Y. Aspartyl tRNA-synthetase (AspRS) gene family enhances drought tolerance in Poplar through BABA-PtrIBIs-PtrVOZ signaling module. BMC Genom. 2023. https://doi.org/10.1186/s12864-023-09556-2.
81. Zhou S, Ma S, Li M, Li C, Gong X, Guan Q, et al. Comprehensive genomic analysis and expression profiling of argonaute gene family and examination of their regulatory roles in water-use efficiency and abiotic stress responses in Apple. Acta Physiol Plant. 2016. https://doi.org/10.1007/s11738-016-2245-0.
82. Young A, Boyle T, Brown T. The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol. 1996;11(10):413–8.
83. Abebe S. A review on genetic diversity and suitability of *Ex situ* germplasm conservation for Re-introduction. J Plant Sci Agri Res. 2025;9(1). https://doi.org/10.36648/ipjpsar.9.1.170.
84. Vieira J, Pimenta J, Gomes A, Laia J, Rocha S, Heitzler P, et al. The identification of the *Rosa S*-locus and implications on the evolution of the rosaceae gametophytic self-incompatibility systems. Sci Rep. 2021;11:3710. https://doi.org/10.1038/s41598-021-83243-8.
85. Ma X, Cai Z, Liu W, Ge S, Tang L. Identification, genealogical structure and population genetics of *S*-alleles in *Malus sieversii*, the wild ancestor of domesticated apple. Heredity. 2017;119:185–96. https://doi.org/10.1038/hdy.2017.28.
86. McKinney GJ, Waples RK, Seeb LW, Seeb JE. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Mol Ecol Resour. 2017;17(4):656–69. https://doi.org/10.1111/1755-0998.12613.
87. Begun D, Aquadro C. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature. 1992;356:519–20. https://doi.org/10.1038/356519a0.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com