

Doctoral (PhD) dissertation

Beti Ivanovska

Budapest

2026



Hungarian University of Agriculture and Life Sciences

Title

Molecular characterization of the *Prunus Falling Stones* transposon and its application in marker development

Beti Ivanovska

Budapest

2026

Doctoral School of Agricultural and Food Sciences

Name: Beti Ivanovska

Discipline:

Crop Production and Horticultural Sciences

Head of the Doctoral School:

Prof. Dr. Melinda Kovács

University Professor, Full Member of the Hungarian Academy of Sciences

Supervisor:

Prof. Dr. Attila Hegedűs

Head of Department, University Professor, DSc

Hungarian University of Agriculture and Life Sciences, Budapest, Hungary

Institute of Genetics and Biotechnology, Group of Horticultural Plant Genetics, Department of Plant Biotechnology

.....
Approval of the Head of Doctoral School

.....
Approval of the Supervisor(s)

(NB: The doctoral dissertation has to be submitted in four copies bound with the original signatures and the ten theses with the photocopied signatures to the Office of the PhD School concerned.)

Table of Contents

ABBREVIATIONS	6
1. INTRODUCTION	9
2. OBJECTIVES TO ACHIEVE	11
3. LITERATURE OVERVIEW	12
3.1 Classification of Transposable Elements (TEs)	13
3.2 Fruit trees genomes	15
3.3 Genetic variability induced by transposition	17
3.4 Phenotypic traits associated with transposable elements	19
3.5 Transposable elements causing the breakdown of self-incompatibility	22
3.6 RNA-Seq Analysis and Read Alignment	24
3.7 Molecular markers developed for genetic analysis of <i>Prunus</i> species	27
4. MATERIALS AND METHODS	31
4.1 Plant material	31
4.1.1 Plant material for transcriptome analyses	31
4.1.2 Plant material for molecular marker development	31
4.1.3 Plant material for RNA-Seq analyses	33
4.2 RNA and DNA extraction	34
4.2.1 RNA extraction and cDNA synthesis for transcriptome analyses	34
4.2.2 Genomic DNA extraction	35
4.3 Sequencing and library preparation for RNA-Seq analyses	35
4.4 PCR analyses	36
4.4.1 PCR parameters and detection of amplified fragments for transcriptome analyses	36
4.4.2 Real-Time PCR	37
4.4.3 Genomic PCR with SCoT and <i>FaSt</i> -specific primers	38
4.5 Cloning and DNA sequencing	39
4.6 Data evaluation and bioinformatics analyses	39
4.6.1 Data evaluation and bioinformatics analyses for molecular marker development	39
4.6.2 Bioinformatic tools for <i>in silico</i> analyses of the presence of <i>FaSt</i> in the promoter region of the <i>CBF</i> gene	41
4.6.3 Bioinformatic tools for <i>in silico</i> analysis of the occurrence <i>FaSt</i> in <i>Prunus</i> genome and transcriptome	42
4.6.4 Data analyses from the qRT-PCR	42
4.7 Analyses of RNA-Seq data	42
4.7.1 Pre-processing of sequencing data	42
4.7.2 PCA and sample-distance analysis (quality control and exploratory data analysis) ...	43
4.7.3 Quantification of gene expression levels	44
4.7.4 Differential expression analysis	44
4.7.5 Functional annotation of differentially expressed genes	45
5. RESULTS AND DISCUSSION	46
5.1 Detection of <i>FaSt</i> -related transcripts in <i>Prunus</i> by RT-PCR	46
5.2 Real-time PCR detection of <i>FaSt</i> -derived transcripts	47
5.3 <i>In silico</i> analysis of the occurrence of <i>Falling Stones (FaSt)</i> transposable elements in <i>Prunus</i> genome and transcript datasets	48
5.4 Insertion of a <i>FaSt</i> in the promoter region of an almond <i>C-repeat-binding factor (CBF)</i> gene	50
5.5 Results for RNA-Seq analysis and read alignment	52
5.5.1 Read alignment	53
5.5.2 Scatterplots of transformed counts	54
5.5.3 Principal Component Analysis	57

5.5.4 Sample Distance Analysis.....	57
5.5.5 Independent filtering.....	60
5.5.6 Differential expression analysis in anther tissues	60
5.5.7 Discussion on the RNA-Seq analysis of apricot samples	64
5.6 Development of a new marker system based on <i>Falling Stones</i> and SCoT primers	67
5.6.1 Amplification by the SCoT and FaSt-R primers and their combinations	67
5.6.2 Phylogenetic analysis of <i>Prunus</i> using SCoT-FaSt markers.....	73
5.6.3 Sequence analysis of the amplified fragments.....	75
5.6.4 Discussion of the SCoT-FaSt marker system.....	80
5.6.5 Discussion on functional and evolutionary aspects of the genomic landscape of <i>FaSt82</i>	
6. CONCLUSIONS AND RECOMMENDATIONS.....	86
7. NEW SCIENTIFIC RESULTS	88
8. SUMMARY.....	89
9. PUBLICATIONS SERVING AS THE BASIS OF THE DISSERTATION:	91
9.1 Papers related to the thesis:	91
9.2 Other papers:.....	91
9.3 Publications:	92
10. APPENDICES.....	93
Appendix A1. Bibliography.....	93
Appendix A2. Bioinformatics Pipeline Script (from <i>REPORT.ipynb</i>)	113
Appendix A3. R Script for RNA-Seq Analysis	119
Appendix A4. PCR Amplification Patterns	126
Appendix A5. Sequence Data	127
Appendix A6. Declaration of Students and Doctoral Candidates on the Use of Artificial Intelligence (AI)”.....	134
11. ACKNOWLEDGMENT	136

ABBREVIATIONS

ASE – Allele-specific expression
AFLP – Amplified fragment length polymorphism
AMS – Aborted microspores
bHLH – Basic helix-loop-helix transcription factor
BAM – Binary alignment/map
BLASTn – Basic local alignment search tool
bp – Base pairs
CBF – C-repeat-binding factor
CCCH - Cys-Cys-Cys-His zinc finger motif – a class of zinc finger proteins implicated in RNA metabolism and regulation
CCHC - Cys-Cys-His-Cys zinc finger motif – a structural motif in RNA-binding proteins involved in transcriptional or post-transcriptional regulation
cDNA – Complementary Deoxyribonucleic Acid
CNS – Conserved non-coding sequences
CTAB – Cetyltrimethylammonium bromide
CYP71AN24 – Cytochrome P450, family 71, subfamily AN, polypeptide 24
DEGs – Differently expressed genes
DFRA – Dihydroflavonol 4-Reductase
DIRS-like - Dictyostelium Intermediate Repeat Sequence-like
DMRs – Differently methylated regions
DNA – Deoxyribonucleic acid
dNTPs – Deoxynucleotide Triphosphates
Ds-like element – Dissociation-like element
evg – Evergrowing
EST-SSRs - Expressed Sequence Tag–Simple Sequence Repeats
EtBr – Ethidium bromide
FaSt – *Falling Stones* transposon
FPKM – Fragments per kilobase of transcript per million fragments
GDR – Genome Database of Rosaceae
GDSL - GDSL esterase/lipase
GENCODE – Genome ENCyclopedia Of DNA Elements
GMAP – Genomic Mapping and Alignment Program
GSI – Gametophytic self-incompatibility
h – Nei's gene diversity
HISAT – Hierarchical Indexing for Spliced Alignment of Transcripts
HTSeq – High-Throughput Sequencing
I – Shannon's information index
iPBS – Inter-primer binding site

ISSR – Inter-simple sequence repeat
LINEs – Long interspersed elements
LRR - Leucine-rich repeat
LTR – Long terminal repeat
MADS - MCM1, AGAMOUS, DEFICIENS, SRF (serum response factor) – a family of transcription factors regulating floral development
MAPK - Mitogen-activated protein kinase
MdPI – *Malus domestica* PISTILLATA
MDS - Multidimensional Scaling
MgCl₂ – Magnesium Chloride
miRNA - MicroRNA
MITEs – Miniature inverted-repeat transposable elements
μL – Microliter
mM – Millimolar
μM – Micromolar
MS1 - Male sterility 1 – a gene required for tapetum development and pollen viability
Mya – Million years ago
MYB – Myeloblastosis transcription factor
NDB – Number of differently sized bands
NCBI – National Center for Biotechnology Information
PA ≤ 1 kb – Percentage of amplicons with size ≤ 1000 bp
ParMDO – *Prunus armeniaca* M-locus DsbA-like oxidoreductase
phasiRNA - Phased small interfering RNA
PpCCD4 – *Prunus persica* Carotenoid Cleavage Dioxygenase 4
PpMYB25 – *Prunus persica* MYB transcription factor 25
PpYUC11 – *Prunus persica* YUCCA 11 (flavin monooxygenase-like enzyme)
PCR – Polymerase chain reaction
PIC – Polymorphism information content
PPB – Percentage of the polymorphic bands
qRT-PCR – Quantitative reverse transcription PCR
RAPD – Random amplified polymorphic dna
RefSeq – Reference Sequence Database
RNA – Ribonucleic acid
RNA-seq – RNA sequencing
RRM – Reiterated reproductive meristem
Rp – Resolving power
RSEM – RNA-Seq by Expectation-Maximization
RT – Red petal tissues
S-locus - Self-incompatibility locus
S-RNase – S-ribonuclease
SAM – Sequence Alignment/Map
SBP - SQUAMOSA promoter-binding protein

SC - Self-compatible
SCoT – Start codon targeted
SFB – S-haplotype-specific F-box
SI – Self-incompatibility
SINEs – Short interspersed nuclear elements
SRA – Size range of amplicons
SSR – Simple sequence repeat
STAR – Spliced Transcripts Alignment to a Reference
TBE – Tris-Borate-EDTA (buffer used in electrophoresis)
TDs – Transposon-display
TEs – Transposable elements
TFBSs – Transcription factors binding sites
TIRs – Terminal inverted repeats
TNB – Total number of bands
TPase – Transposase
TSD – Target site duplications
UGFGT – UDP-Glucose:Flavonoid 3-O-Glucosyltransferase
UGT79B6 – UDP-Glycosyltransferase 79B6
UDP-Glycosyltransferases – uridine diphosphate-glycosyltransferases
VvMYBA2 – *Vitis vinifera* Myeloblastosis A2 transcription factor
VvTFL1A – *Vitis vinifera* Terminal flower 1A
WD – Tryptophan-aspartic acid (W-D) repeat proteins
WRKY - WRKY transcription factor family
WT – White petal tissues
XTH - Xyloglucan endotransglucosylase/hydrolase
YUCCA8 – YUCCA flavin monooxygenase 8

1. INTRODUCTION

The global production of stone fruits reached approximately 48 million tonnes in 2022, with peaches and nectarines accounting for more than half of this yield. These fruits are not only economically significant but also nutritionally valuable, contributing essential vitamins, antioxidants, and dietary fiber to human diets worldwide. However, the genetic improvement of *Prunus* species remains a formidable challenge due to several biological and technical limitations. The most important factor among these is the long juvenile periods, ranging from 3 to 5 years in stone fruits, alongside the complex inheritance of polygenic traits and the low heritability of key quality attributes such as fruit size, sweetness, firmness, and tolerance to biotic and abiotic stresses.

Modern breeding approaches, notably marker-assisted selection (MAS) and genomic selection (GS), can accelerate genetic gain in *Prunus*, but their success depends on sufficiently dense, informative, and population-appropriate marker sets rather than strictly “species-specific” markers. The transposable element (TE) derived markers are widely used in *Prunus* species such as peach and apricot and deserve broader attention. Overall, transposable elements make up a large fraction of eukaryotic (including *Prunus*) genomes and remain a valuable, under-exploited source of variation for marker development and functional studies.

Although TE insertions are well documented and functional roles have been shown in several fruit crops (e.g., apple, peach, grape), comprehensive functional characterization of many TE families in *Prunus* remains incomplete. The *Falling Stones* (*FaSt*) non-autonomous element, for example, preferentially accumulates in gene-rich euchromatin. This pattern suggests possible regulatory influence through proximity to genes, indicating that *FaSt* is transcriptionally active or that its insertions alter host gene expression is limited. Therefore, while TEs are plausible contributors to epigenetic and regulatory variation affecting traits and stress responses, targeted expression and functional assays are required to confirm their roles for elements such as *FaSt*.

Recent advances in transcriptomics provide new opportunities to investigate TE dynamics and their interactions with host genes. Understanding these relationships could identify novel pathways for trait improvement in *Prunus*. Moreover, the integration of TE

biology with molecular marker development may lead to the development of innovative tools for breeding programs.

Recognizing this need, the present study aims to bridge the gap between transposon biology and applied genetics by establishing *FaSt* as a novel genetic tool for *Prunus* research. Through a combination of bioinformatic analyses, transcriptomic profiling, and marker validation, we explore the potential of *FaSt* to serve both as a molecular marker and as a candidate for functional studies. This integrative approach not only enhances our understanding of TE dynamics in *Prunus* but also contributes to the development of more efficient and targeted breeding strategies.

2. OBJECTIVES TO ACHIEVE

1. Structural and functional characterization of a recently discovered transposon (*FaSt*) in the *Prunus* genome.
2. Analysis of the transcriptome for the expression of *FaSt* elements in diploid and polyploid *Prunus* species.
3. Quantitative analysis of *FaSt* expression in different tissues and across different phenological stages.
4. Identification of genes whose expression might correlate with *FaSt* elements.
5. Development of a new molecular marker system for *Prunus* species based on *FaSt* transposon.

3. LITERATURE OVERVIEW

The impact of the movement of transposable elements (TEs) in plant genomes has been studied from many aspects. Speciation and phylogenomic analysis can infer the evolutionary history of related species. Furthermore, the mutations caused by transposition events contribute to the diversity and if beneficial and domesticated, can increase the fitness of the organism. Although TEs are present at high copy number in *Prunus* genomes, abundance alone does not imply ongoing activity. When TEs insert into coding sequences or regulatory regions, they can influence gene expression or give rise to novel gene variants. However, associations between TEs and specific traits, such as the breakdown of self-incompatibility, require locus-specific and functional evidence. Their abundant presence in *Prunus* genomes is consistent with the earlier hypothesis originally formulated from allele-frequency patterns in Moroccan apricots that *FaSt* is, or has recently been, active. Whether inserted in a coding or regulatory region of a gene, they can alter gene function and induce new phenotypic traits, e.g. self-compatibility.

The first glimpses into the discovery of transposable elements (TEs) in plants, dates back in the middle of the twentieth century (McClintock, 1950). Barbara McClintock has noticed the presence of so-called Ds (dissociator, a non-autonomous hAT superfamily transposon) and Ac (activator, an autonomous hAT superfamily transposon) elements during her analysis on chromosome breakage, gene instability and kernel variegation in maize. Her observations have shown the impact of the movement of genes on the phenotype (McClintock, 1950).

The everlasting impact that TEs impose on the adaptation and the flexibility in transformation of the organisms during longer periods is illustrated by the colonization of the flowering plants (Angiosperms) in the Cretaceous period (Baud *et al.*, 2019). Indeed, the overlaps of TEs with transcription binding sites (TFBSs) involved in flowering in conserved non-coding sequences (CNS) have led to the long-lasting effect on plant evolution and adaptation (Baud *et al.*, 2019).

An expansion of research on TEs in the following decades clarified their effect on plant genomes. From an evolutionary point of view, the transposons demonstrate certain adaptive impact on the host genome. Namely, the insertions of the transposons close to a regulatory region might remain in the genome, which results in exaptation (Joly-Lopez *et al.*, 2012; Barteri, 2019). Through insertions in close proximity to gene regulatory elements, or genes, the transposons are

able to influence the gene expression as well as the genome evolution of *Prunus* species (Wang *et al.*, 2016). Therefore, considering their genome shaping power, the host organism develops mechanisms to lower their activity (Kejnovsky *et al.*, 2012). In order to hinder the transpositional activity, several strategies are involved, such as host silencing on an epigenetic and transcriptomic level, or by acquiring mutations, so that they are unrecognizable for the transposase (TPase) enzyme (Kejnovsky *et al.*, 2012). However, the abundance of mutations generated by the activity of TEs, illustrates their ability to escape silencing imposed by the host genomes (Foster & Aranzana, 2018; Wang *et al.*, 2010).

3.1 Classification of Transposable Elements (TEs)

The vast majority of the eukaryotic organisms studied to date, contain multiple copies of TEs (Kejnovsky *et al.*, 2012). Two main classes of TEs exist according to the classification made by Wicker *et al.* (2007). Class I consists of TEs that can move using an RNA intermediate. They exhibit reverse transcription before integration in the genome and are commonly called retrotransposons. Furthermore, Class I is divided into three Subclasses: long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons (comprising of long and short interspersed elements, shortly LINEs and SINEs, respectively) and DIRS-like elements (Kejnovsky *et al.*, 2012).

The proliferation of the LTR retrotransposons is orchestrated by the enzyme reverse transcriptase (RT), protease (PR), RNaseH (RH) and integrase (INT), which resemble the transposition mechanism observed in retroviruses (Kejnovsky *et al.*, 2012). The final step of the transcription and reverse transcription is the integration of the copy sequence elsewhere in the genome (Kejnovsky *et al.*, 2012).

On the other hand, the TEs from Class II could transpose through DNA intermediate. In Subclass I DNA transposons have "cut-and-paste" transpositional mechanism and subclass II includes rolling-circle transposons (Helitrons) and large virus-like elements such as Mavericks/Polintons, which employ distinct replication and integration strategies (Wicker *et al.*, 2007; Kejnovský *et al.*, 2012). Although canonical cut-and-paste transposition is non-replicative in principle, several DNA transposon families can nevertheless increase their copy number through family-specific mechanisms, including transposition coupled to DNA replication, interactions between autonomous and non-autonomous elements, or other replicative pathways (Feschotte and Pritham, 2007; Kejnovsky *et al.*, 2012). Consequently, the high copy number of

certain Class II transposons in eukaryotic genomes reflects lineage- and family-specific evolutionary dynamics rather than a single, universal mode of amplification.

A gene which expresses the enzyme TPase is required for their transposition. Therefore, those DNA transposons that encode the TPase enzyme are called autonomous transposons (Kejnovsky *et al.*, 2012). The TPase enzyme recognizes the terminal inverted repeats (TIR) of the transposon, excises it from its chromosomal location and inserts it into another location. Meanwhile, the DNA break created by the TPase activity could be repaired by homologous recombination with a template from the same transposon located in another homologous chromosome. Additionally, in case the mobility is performed during cellular replication in the S phase, a sister chromatid can serve the same role as the homologous chromosome mentioned above (Kejnovsky *et al.*, 2012).

Across angiosperms, including *Gossypium*, genome size variation is shaped by episodic bursts of transposable element (TE) activity. In *Gossypium*, these bursts scale with genome size, but in smaller-genome species, TE removal especially of *Gorge3* elements outpaces insertion, leading to genome contraction (Hawkins *et al.*, 2009). Similar TE amplifications have occurred in species like *Oryza australiensis*, where LTR elements doubled genome size in just three million years (Piegu *et al.*, 2006). These patterns suggest that environmental stress and hybridization, both common in *Gossypium*, may trigger TE activation by disrupting epigenetic silencing. A key question is why some lineages tolerate genome expansion more than others, possibly due to differences in TE suppression, DNA removal efficiency, or ecological constraints. Understanding these dynamics is crucial to explaining genome size evolution in plants (Cronn and Wendel, 2003).

The transposons from Class II, Subclass I (TIR elements), named the miniature inverted-repeat transposable elements (MITEs), are non-autonomous TEs. They are fully dependent on the TPase from another autonomous TE, which recognizes the same TIRs. Remarkably, their numerous invasions across the eukaryotic genome are associated with amplification through cross-mobilization (Feschotte *et al.*, 2003). Due to a co-evolution of the TIRs with the TPase, the transposition could be performed by distantly related autonomous TEs (Feschotte *et al.*, 2003). Insertions of MITEs within or near genes can affect host gene expression. Epigenetic silencing of inserted elements (e.g. via DNA methylation or repressive chromatin) can repress expression of adjacent or interrupted host genes, and silencing may also reduce the immediate fitness cost of insertions, allowing otherwise deleterious copies to persist and evade rapid purifying removal (Feschotte & Pritham, 2007). Feschotte *et al.* (2003) proposed that MITEs with lower sequence similarity to their autonomous progenitors might amplify while partly

escaping homology-dependent silencing mechanisms remains an attractive hypothesis, but it requires locus-specific and population-level evidence to establish generality.

3.2 Fruit trees genomes

Taxonomically, the genus *Prunus* is placed in the Kingdom *Plantae*, Division *Spermatophyta*, Subdivision *Angiospermae*, Class *Magnoliopsida*, Order *Rosales*, Family *Rosaceae*, Subfamily *Amigdaloidae*, and Genus *Prunus* (Rodríguez-Robles *et al.*, 2024). The genus *Prunus* is a commercially important group of fruit trees, including widely known diploid stone fruits such as peach (*Prunus persica* (L.) Batsch), apricot (*Prunus armeniaca* L.), sweet cherry (*Prunus avium* L.), Japanese plum (*Prunus salicina* Lindl.), Japanese apricot (*Prunus mume* (Siebold) Siebold & Zucc.), and almond (*Prunus dulcis* (Mill.) D. A. Webb), all of which have 16 chromosomes ($2n = 2x = 16$). Additionally, the genus includes tetraploid species like sour cherry (*Prunus cerasus* L., $2n = 4x = 32$) and hexaploid European plum (*Prunus domestica* L.) with $2n = 6x = 48$ (Zhang *et al.*, 2012).

Historical evidence has shed light on the global spread and distribution of *Prunus* species (Rodríguez-Robles *et al.*, 2024). The genus *Prunus* (*Rosaceae*) includes economically and culturally significant temperate fruits while almond is grown for its seeds, the others are prized for their flavourful fruit flesh. Originating in Central Asia, many *Prunus* species have secondary centres of diversity in Eastern Asia, Europe, and North America. Historical records suggest their spread to the Mediterranean basin was facilitated by Alexander the Great's conquests, which connected trade routes like the Silk Road (Janick, 2008; Rodríguez-Robles *et al.*, 2024). Roman names for *Prunus* fruits reflect their perceived origins, e.g., peach (*persica*, Persia), apricot (*armeniaca*, Armenia), and cherry (*cerasus*, Kerasun on the Black Sea shore), though these locations only served as trade waypoints (Rodríguez-Robles *et al.*, 2024).

The fruit (mesocarp) or nut (endocarp-enclosed seed) of such species are consumed fresh and processed (dried, canned or fermented), and are an important source of sugar, minerals, antioxidants and other health-promoting compounds (Meena *et al.*, 2021). Based on morphological features, five subgenera were described (Rehder, 1940), that were later decreased to three (*Prunus*, *Cerasus* and *Padus*) by molecular studies (Shi *et al.*, 2013). With the use of RAD-seq and genome skimming analyses, the molecular phylogeny and inflorescence evolution of *Prunus* have been studied, shedding light on the evolution of inflorescence structures (Su *et al.*, 2023). The presence of three *Prunus* subgenera is congruent with the inflorescence structure

as *Prunus*, *Cerasus* and *Padus* species have solitary flowers, corymbose and racemose inflorescences, respectively (Chin *et al.*, 2014).

The *Prunus* genus is thought to have originated ~61 Mya (million years ago) in eastern Asia and spread to its current distribution through different times and dissemination routes shaped by the driving forces of various oscillations in the climate and the geological influences (Chin *et al.*, 2014). *Cerasus* and *Prunus* lineages have diverged ~54 Mya. The sour cherry is an allotetraploid hybrid of *P. fruticosa* and *P. avium* (Olden and Nybom, 1968; Wöhner *et al.*, 2021). Within *Prunus*, apricot and plums including the European plum were estimated to diverge between 35-40 Mya (Chin *et al.*, 2014). Peach and almond originated from the same progenitor species in central Asia after exposing the two lineages to different levels of water availability (Tao *et al.*, 2007), and their divergence time is estimated approx. 8-5.8 Mya (Velasco *et al.*, 2016; Alioto *et al.* 2020).

Prunus domestica has been considered as a hexaploid hybrid between *P. cerasifera* (diploid) and *P. spinosa* (autotetraploid) by Crane and Lawrence (1934). Several researchers have tried to confirm the theory that *P. domestica* is a product of hybridization. Thereby, Rybin (1936) performed artificial hybridization with *P. spinosa* and *P. cerasifera*, as well as Endlich and Murawski (1962). Later on, Bajashvili (1991) did the same experiment, that again proved that the cross could generate fertile offspring. However, he also suggested that there is a possibility that the European plum is created by an autopolyploidy of *P. cerasifera* or of *P. spinosa*. No wild types have been found in the Caucasus area by Rybin (1936) nor by Bajashvili (1991), although, *P. domestica* originates and has inhabited the region for millenniums (Hartmann and Neumüller, 2009). To date, the studies on the genome of *P. domestica* have not yet reached a firm conclusion whether it is a hybrid by origin. Still the presence of *P. spinosa* and *P. cerasifera*, in the genesis of European plums have been confirmed (Hartmann and Neumüller, 2009; Zhebentyayeva *et al.*, 2019).

Plums exhibit remarkable diversity and have been domesticated on three continents. Key species include European plum (*P. domestica*), Japanese plum (*P. salicina*), and American plum (*P. americana*). While the European plum is a hexaploid derived from wild progenitors, the Japanese and American plums are diploids. Plums have been cultivated since at least 300 BCE and are used fresh, dried (as prunes), or processed into juices, brandies, and baked goods. Interspecific hybrids, such as plumcots (plum × apricot), highlight ongoing innovation in plum breeding (Janick 2005).

FAOSTAT data for 2024 indicate that global *Prunus* production reached 26.35 million metric tons for peaches and nectarines, 12.39 million metric tons for plums and sloes, approximately 4.36 million metric tons for cherries (sweet and sour combined), and 3.86 million metric tons for apricots, underscoring the economic importance of this genus in global fruit agriculture (FAO, 2024). China is the leading producer of peaches and plums, while Turkey dominates apricot production and ranks among the top producers of cherries. Global almond production output is largely concentrated in the United States, with additional major contributions from Spain and Australia (FAO, 2024; USDA FAS, 2024).

3.3 Genetic variability induced by transposition

Transpositional dynamism refers to activity level, mobilization frequency and population behaviour in the genomic composition of TEs over time (Barteri, 2019). TE mobilization can sometimes produce a dynamic balance between novel insertions and subsequent losses, but this putative equilibrium is not universal and instead depends on TE family characteristics and host-species biology. In reference to the impact on the structural and functional genomics changes made by TEs, the genetic variability could be seen in form of somatic, intra- and interspecific variability (Barteri, 2019).

One phylogenomic analysis demonstrated the relationship between almond and peach cultivars (Barteri *et al.*, 2020). The thorough analysis on LTR annotations in peach and almond, showed their chromosomal localization near the centromeric regions. Almond has displayed roughly seven-fold higher nucleotide diversity, as well as greater retention of ancestral LTR retroelements, whereas peach has shown more recent retrotranspositional activity and fewer polymorphic ancestral insertions due to its SC mating type and a strong reduction in population size (Barteri *et al.*, 2020).

Comparative examinations of fourteen *Rosaceae* genomes including various *Prunus* taxa have demonstrated that, even though protein-coding regions remain strongly preserved, long terminal repeat retrotransposons (LTR-RTs) fluctuate substantially in abundance, age, and chromosomal distribution (Yu *et al.*, 2024). By assessing TE copy number, insertion age, insertion-site preferences, conserved protein domains, phylogenetic affiliations, and transcriptional responsiveness, Yu *et al.* (2024) observed that numerous LTR-RT families become transcriptionally responsive under environmental cues, indicating that novel insertions can modulate neighbouring gene expression and promote clade-specific adaptation. GO enrichment of 3,903 genes adjacent to intact TEs revealed that ancient “cut-and-paste” DNA transposons are

linked to development, catalytic function, and fundamental metabolic pathways, while newer TE insertions are situated near genes implicated in stress responses and membrane transport. This pattern suggests that ancestral TEs remain as enhancer-like regulators or transcription factor binding sites within growth and metabolic networks, whereas emergent TEs can be harnessed for swift stress adaptation. Yu *et al.* (2024) also remark that current genome assemblies and TE annotation pipelines often fail to detect elements within highly repetitive regions. To minimize false omissions, their study concentrated on structurally complete retrotransposons (characterized by paired LTRs or conserved protein domains) and DNA transposons with unmistakable characteristics.

One study has provided evidence of occurrence of the retrotransposons *Ty1-copia* and *Ty3-gypsy* in the Japanese apricot *Prunus mume* cultivar ‘Koume’ (Wang *et al.*, 2010). The abundance of the retrotransposons from *Ty1-copia* group have been found to be represented with 18.4% throughout the genome of *P. mume*. Furthermore, *Ty3-gypsy* retrotransposon group is present with almost 33.3% of the *P. mume* genome. Although such abundance signals a substantial capacity to influence genome architecture and generate genetic variation, additional functional, ecological or populational genetic evidence would be necessary to confirm it (Wang *et al.*, 2010).

The regulation of the expression of nearby genes can give positive, negative or neutral implication on the host genome (Qiu and Köhler, 2020). Sometimes TEs could encompass binding motifs for TFs, so their movement might affect the gene regulatory network systems (Baud *et al.*, 2019; Qiu and Köhler, 2020). The emergence of C₄ photosynthesis highlights the evolutionary significance of TE activity in plant survival. The development of the C₄ species can be characterized by convergent evolution. Taken that more than 60 lineages have independently evolved into C₄ plants in a geologically short timeframe (in approximately 25-30 Mya) (Sage *et al.*, 2011), often involving a combination of gene duplication, neofunctionalization, and extensive regulatory rewiring. In several C₄ plants, pre-existing *cis*-regulatory elements have been sometimes derived from or associated with TEs and have been co-opted to confer cell-specific expression on photosynthetic genes, even when coding sequences have remained largely conserved relative to the C₃ relatives (Kajala *et al.*, 2012). The frequent association of these recruited regulatory elements with TEs supports a contributory role for transposable elements in the convergent evolution of C₄ photosynthesis (Cao *et al.*, 2016).

Accumulating evidence demonstrated the influence of the TEs on the gene regulatory networks related to the adaptation and flowering time in *Prunus persica* (Morata *et al.*, 2018). The presence of TF binding motifs (TFBMs) in MITEs, 11 out of 12 investigated genotypes,

contained TFBM in MITEs fragments (Morata *et al.*, 2018). The avalanche existence of a bZIP60 TF binding motif responsible for the regulation of bud dormancy (Fu *et al.*, 2014), in the *P. mume* and *P. persica* genomes suggests MITE mobility in a common ancestor (Morata *et al.*, 2018). Notwithstanding, the binding site of an additional TF, PIF3, was inserted with MITE in the upstream region of many genes involved in flower regulation in *P. mume* that might have influenced the flowering time in this early blooming species (Morata *et al.*, 2018).

3.4 Phenotypic traits associated with transposable elements

One observable consequence of transposable element (TE) activity is phenotypic variation, and such variation has been used as measurable evidence of TE activity in stone fruit trees, including traits of economic relevance (Tab. 1).

Vegetative propagation serves as a good practice in woody trees propagation, and bud sports are spontaneous mutations that introduce a new trait to the population, without losing the already existing traits. In one report, it has been found that about 170 cultivars of peach and nectarines are created as a result of bud sport mutations (Foster & Aranzana, 2018). This phenomenon of bud sports has captured the attention of Charles Darwin. Long before the advent of molecular investigations, Darwin described the metamorphic changes in size, colour, formation of double flowers in *Gentiana amarella* (Darwin, 1843; Foster & Aranzana, 2018). The movement of TEs in fruit trees could be caused by wounding, pruning, or viral infections (Foster & Aranzana, 2018). Furthermore, some of the obvious differences between peach and almond can be molecularly explained through the activity of TEs. Firstly, the lack of expression of the genes involved in the primary steps of the amygdalin biosynthesis in almond seeds is responsible for non-toxic and not bitter edible seeds (Alioto *et al.*, 2020; Thodberg *et al.*, 2018). Namely, two MITE insertions were identified in the upstream region of the gene *CYP71AN24*, one of which was designated *TIR2*. Their presence has caused high methylation, and influenced the expression of the corresponding gene (Alioto *et al.*, 2020). A second example involves the insertion of TEs approximately 1,000 bp upstream of six genes: *WUSCHEL* (*WUS*), *CLAVATA3* (*CLV3*), *SHOOT MERISTEMLESS* (*STM*), *APETALA1* (*AP1*), *FRUITFULL* (*FUL*), and *AINTEGUMENTA* (*ANT*), which collectively function in meristem organization, floral development, and fruit cell proliferation (Alioto *et al.*, 2020). As in the previous example, the TEs are silenced by methylation, which leads to altered expression of the six genes respectively (Alioto *et al.*, 2020).

The fruit hardness in peach (*Prunus persica*) is related to the concentration of ethylene present in the process of ripening (Tatsuki *et al.*, 2018). The firmness can be vastly influenced by the activity of the enzyme endo-polygalacturonase (PG) (Peace *et al.*, 2005). In recent studies, one gene candidate for fruit softening in peach *PpYUC11* has been investigated. The presence of a transposon at the 5' flanking region of the gene decreases the level of expression. The peach genotypes containing two pairs of this insertion as a homozygous genotype, possess a stony-hard fruit phenotype (Tatsuki *et al.*, 2018). Tatsuki *et al.* (2018) demonstrated that a transposon-like insertion in the 5'-flanking region of *PpYUC11* is associated with markedly reduced *PpYUC11* expression and co-segregates with the stony-hard phenotype. *PpYUC11* encodes a *YUCCA* enzyme involved in auxin (IAA) biosynthesis (Pan *et al.*, 2015). Reduced auxin at ripening is expected to limit induction of ethylene biosynthesis and downstream ethylene-responsive softening genes; ethylene in turn transcriptionally induces endo-polygalacturonase and other cell-wall-modifying enzymes that drive fruit softening (Hayama *et al.*, 2006). Thus, the transposon-associated down-regulation of *PpYUC11* provides a parsimonious mechanism linking reduced auxin to lowered ethylene production and consequent failure to activate cell-wall-remodeling required for normal softening, producing the stony-hard phenotype (Tatsuki *et al.*, 2018).

A molecular flower chimerism is a feature that could be detected through a comparative single base methylome and transcriptome sequencing (Ma *et al.*, 2018). While studying the spatial and temporal epigenetic variation between differently expressed genes (DEGs) in the ornamental fruit tree *Prunus mume* (Pm) 'DanbanTiaozhi', they noticed that those DEGs contain TE insertions (Ma *et al.*, 2018). Strikingly, by comparing the epigenetic variation between white petal tissues (WT) and red petal tissues (RT), they observed differently methylated regions (DMRs) (Ma *et al.*, 2018). They observed DMRs in red versus white petal tissues are associated with altered expression of genes functionally grouped in auxin biosynthesis, flavonoid/anthocyanin biosynthesis, and glycosyltransferases (Ma *et al.*, 2018).

Ma *et al.* (2018) reported that spatially distinct petal sectors in *Prunus mume* ('DanbanTiaozhi') display differential DNA methylation and that those differently methylated regions frequently colocalize with transposable element insertions. These epigenetic differences are associated with altered expression of genes from specific functional categories relevant to pigmentation and development: increased expression in red petal tissue was observed for genes involved in auxin biosynthesis (*YUCCA* family) and for certain UDP-glycosyltransferases implicated in anthocyanin modification, whereas several members of the flavonoid biosynthetic

pathway (notably *dihydroflavonol-4-reductases*, *DFRs*, and associated glycosyltransferases) showed reduced expression in red sectors.

The development of bioinformatics tools assists the annotation and functional characterization of TEs. Ma *et al.* (2018) have created the first *Rosaceae* plant TEs database (RPTeddb) based on 19,596 annotated TEs. Among 52 differentially expressed long non-coding RNAs (lncRNAs) in the transcriptomes of *P. mume* ‘FubanTiaozhi’, 16 have been located near TEs linked to transcription factor genes such as *MYB*, *bHLH*, and those encoding WD-repeat proteins and UDP-glycosyltransferases. This suggests that TEs, possibly in coordination with nearby lncRNAs, may influence gene expression involved in chimeric flower coloration (Ma *et al.*, 2018).

A cluster of *MADS-box* genes are involved in photoperiod perception and phytochrome function. Peach which has no perception of the short days, lower temperature, and is not forming terminal vegetative buds is considered as evergrowing (*evg*) (Bielenberg *et al.*, 2008). Naturally all *Prunus* fruit trees go into dormancy during the winter period, so the molecular background of this exception is related to mutations caused by TEs (Bielenberg *et al.*, 2008). A fraction of an autonomous transposable retroelement has been discovered in the exon of a polyprotein *EVG10* (Bielenberg *et al.*, 2008).

In *Prunus dulcis*, Castanera *et al.* (2024) showed that structural variation was largely driven by heterozygous TE insertions affecting gene regions. Notably, about 9% of genes display allele-specific expression (ASE) linked to such TE-based variations with the allele carrying a TE insertion generally expressed at lower levels than its counterparts. Crucially, many of these TE insertions remain polymorphic (heterozygous) in almond populations for extended periods, providing a persistent source of allelic diversity. Castanera *et al.* (2024) argued that this long-term maintenance of TE-induced variation allows natural selection to “test” new expression patterns: while deleterious insertions are often purged, those that persist can generate novel phenotypes. In effect, heterozygous TE insertions create lasting intra-population genomic variation that can be acted on by evolution enhancing the species’ adaptive potential to changing environments (e.g. seasonal shifts or long-term climate changes) (Castanera *et al.*, 2024).

3.5 Transposable elements causing the breakdown of self-incompatibility

Self-incompatibility (SI) is widespread across angiosperms, occurring in at least 19 orders, 71 families and 250 genera (Allen and Hiscock, 2008; Morimoto *et al.* 2015). The genetic diversity and outcrossing in fruit trees are preserved by gametophytic self-incompatibility (GSI) (Tsukamoto *et al.*, 2006) to help avoid inbreeding. The multiallelic S-locus is responsible for SI in *Prunus* by having a pistil-expressed *S-ribonuclease* (*S-RNase*) gene and a pollen-expressed *S-haplotype-specific F-box* (*SFB*) gene (Tab. 1) (Sonneveld *et al.*, 2005; Halasz *et al.*, 2019). Sexual compatibility is regulated by the haploid genome of the pollen and the diploid genome of the pistil (Hauck *et al.*, 2006b). In diploid species, the recognition of the self and non-self-pollen is done by the identification of the S-allele product in the pollen tube by the two S-allele products in the pistil (Huang *et al.*, 1994). In case of self-compatible (SC) plants, the pollen carrying the same S-allele as in the pistil can achieve fertilization. However, in that case, SI plants fail to reject the growth of the self-pollen tubes by the *S-RNase* enzymes produced in the pistil (Huang *et al.*, 1994).

One study on sour cherry (*Prunus cerasus*) revealed that a non-functional S-haplotype could be preserved and the sour cherry maintains SI, since it is tetraploid (Hauck *et al.*, 2006b). So, at least alterations in two S-haplotypes are needed for having a SC sour cherry. The changes in the S-alleles imply transposon-like element insertion in the S_{6m} -haplotype with a functional pollen and non-functional pistil gene (Yamane *et al.*, 2003). One plausible proposition could be that as a recent polyploid, sour cherry experiences bigger transpositional activity and chromosomal rearrangement, described as genome fluidity (Tsukamoto *et al.*, 2006; Voytas and Naylor, 1998). Likewise, the duplication of each S-haplotype gene speeds up the rate of evolution, the frequency of loss-of-function mutations, such as deletions, insertions and transpositions (Tsukamoto *et al.*, 2006). A Ds-like element of 615 bp size has been found in the coding sequence of the S_1 -haplotype pollen gene, termed S_1' (Tab. 1), which is attributed to a MITE, called *mMoshan* (Hauck *et al.*, 2006a; Wang *et al.*, 2016). When a transposon is cut, the target size duplication created upon transposon insertion can be kept, resulting in a footprint, a direct repeat of a small motif. A 5 bp insertion in the SFB_2 gene of peach presents a putative example of transposon footprints, that introduces a translation termination codon in the middle of the gene (Tao *et al.*, 2007).

The non-autonomous *Mutator* MITE element, named *Falling Stones* (*FaSt*), identified in apricot, *P. armeniaca*, gives rise to SC (Halász *et al.*, 2014). This 358 bp MITE element is inserted into the ORF of a *S-haplotype-specific F-box* gene, disturbs its function and results in

SC (Vilanova *et al.*, 2006; Halász *et al.*, 2014). Recent findings have confirmed *P. avium* cultivar ‘Cristobalina’ SC is due to a mutation outside the S-locus (Cachi *et al.*, 2014; Ono *et al.*, 2018). A transposon-like insertion was identified in the M locus-encoded *glutathione S-transferase* (*MGST*) gene, located on the edge of chromosome 3 (Ono *et al.*, 2018). In addition, a pollen-part mutation has been described in *P. armeniaca* ‘Canino’ and ‘Katy’, that is in a similar location as *MGST* (Vilanova *et al.*, 2006; Zuriaga *et al.*, 2013). Interestingly, this insertion is identical to the *FaSt* transposon described by Halász *et al.* (2014) in the apricot *SFB_C* allele (Tab. 1). It disrupts the function of the gene encoding for a disulfide bond A-like oxidoreductase protein, *ParMDO*, by an insertion in the ORF, therefore four exons are missing from the truncated protein (Muñoz-Sanz *et al.*, 2017).

Table 1. Examples of mutations caused by TEs in fruit species and their literature sources

Mutation	Crop species	Phenotypic effect	References
Insertion of a TE into the promoter of <i>VvTFL1A</i>	<i>Vitis vinifera</i> cultivar ‘Carignan’	Reiterated reproductive meristem RRM	Fernandez <i>et al.</i> (2010)
Insertion of a LTR in the coding sequence of <i>PpMYB25</i>	Nectarine, <i>Prunus persica</i>	Hairless phenotype	Vendramin <i>et al.</i> , (2014)
Insertion of a retrotransposon adjacent to a MYB transcriptional activator of anthocyanin production	Sicilian blood oranges (<i>Citrus sinensis</i>)	Blood orange varieties	Butelli <i>et al.</i> (2012)
A retrotransposon in the promoter of white berry allele of <i>VvMYBA2</i>	<i>Vitis vinifera</i>	White grapes	Walker <i>et al.</i> (2007)
An insertion of a transposon in the 5'-flanking region of <i>PpYUC11</i>	<i>Prunus persica</i>	Late-ripening stony hard peaches	Tatsuki <i>et al.</i> (2018)
Highly methylated TE insertions in the close proximity of <i>CYP71AN24</i>	<i>Prunus dulcis</i>	Sweet almond	Barteri <i>et al.</i> (2020)
LTR-type retrotransposon inserted in <i>MdPI</i>	<i>Malus domestica</i>	Seedless fruit	Yao <i>et al.</i> (2000)
615-bp <i>Ds</i> -like element inserts 225 bp downstream of the start codon of the <i>S₁-haplotype-specific F-box gene SFB₁</i>	<i>Prunus cerasus</i> L.	Breakdown of SI	Hauck <i>et al.</i> (2006a)

Mutation	Crop species	Phenotypic effect	References
<i>FaSt</i> MITE located in the coding region of the <i>SFB</i> gene	<i>Prunus armeniaca</i>	Breakdown of SI	Halász <i>et al.</i> (2014)
<i>FaSt</i> MITE located in <i>M</i> -locus <i>DsbA</i> -like oxidoreductase gene, <i>ParMDO</i>	<i>Prunus armeniaca</i>	Breakdown of SI	Muñoz-Sanz <i>et al.</i> (2017)
Transposon-like insertion in the <i>M</i> locus-encoded glutathione <i>S</i> -transferase (<i>MGST</i>)	<i>Prunus avium</i>	Breakdown of SI	Ono <i>et al.</i> (2018)
<i>PpCCD4</i> gene contains a LTR element in an intron	<i>Prunus persica</i>	Yellow flesh in peach	Falchi <i>et al.</i> (2013)

3.6 RNA-Seq Analysis and Read Alignment

RNA sequencing (RNA-seq) was first introduced in 2008 to profile transcriptomes with unprecedented resolution and sensitivity (Wang *et al.*, 2009). It rapidly replaced earlier transcriptomic methods such as expressed sequence tags (ESTs) and DNA microarrays due to its higher sensitivity, ability to detect novel transcripts, and independence from prior sequence knowledge. In plant research, RNA-seq has become a cornerstone technique for studying gene expression, developmental processes, stress responses, and evolutionary adaptations (Upton *et al.*, 2023). Its versatility has enabled transcriptome profiling across a vast diversity of species. Notably, the 100 Plant transcriptomes initiative alone generated RNA-seq data for over 1,173 plant species (Carpenter *et al.*, 2019), including *Prunus* species (Martínez-Gómez *et al.*, 2011).

RNA-Seq analysis begins with the alignment of sequencing reads to a reference genome. If no reference genome is available for a given species, alignment can be performed against a corresponding transcriptome, either an existing reference transcriptome or one assembled *de novo* from the analyzed data. Each read is assigned a match or no-match status based on whether it aligns to the reference (Jiung-Wen *et al.*, 2023).

However, certain challenges arise during alignment. The presence of repetitive sequences and paralogous gene copies can prevent reads from uniquely mapping to a single position in the genome or transcriptome (Chen *et al.*, 2023; Conesa *et al.*, 2016). In alignment-based RNA-seq analyses, the proportion of uniquely mapped reads is commonly used as a quality metric, reflecting both data quality and the effectiveness of the alignment algorithm (Oshlack *et al.*, 2010). The proportion of uniquely mapped reads serves as an important quality metric,

reflecting both data quality and the effectiveness of the alignment algorithm (Oshlack *et al.*, 2010). Before alignment, raw cDNA short fragments must undergo quality control. If the quality is sufficient, downstream analyses can proceed, typically involving splice-aware aligners such as STAR (Dobin *et al.*, 2013; Van den Berge *et al.*, 2019), HISAT (Kim *et al.*, 2015), Subread (Liao *et al.*, 2013), and GMAP (Wu *et al.*, 2005). When aligning reads to a reference genome, splice junctions must be considered since cDNA fragments may span intronic regions. In contrast, when aligning to a reference transcriptome, reads are already spliced and do not require junction mapping (Van den Berge *et al.*, 2019). After alignment, reads are stored in either Sequence Alignment/Map (SAM) or Binary Alignment/Map (BAM) file formats. The next step involves associating aligned reads with specific genes or exons. While *de novo* (reference-free) assembly is an option, it generally performs less effectively than reference-guided approaches (Hayer *et al.*, 2015; Chen *et al.*, 2023).

Following alignment, RNA-Seq reads can be quantified at either the gene level or transcript level, depending on the research objective (Van den Berge *et al.*, 2019). Tools such as RSEM implement a probabilistic model with expectation–maximization to assign fragments and estimate transcript abundances. These abundance estimates are commonly reported as normalized measures such as transcripts per million (TPM). (Li *et al.*, 2011; Van den Berge *et al.*, 2019).

Several other statistical models have been developed to handle multi-mapped reads and improve quantification accuracy. These include MMSEQ (Turro E. *et al.*, 2011), eXpress (Roberts and Pachter, 2013), Cufflinks (Trapnell *et al.*, 2010), and BitSeq (Glaus *et al.*, 2012), among others (Van den Berge *et al.*, 2019). For species without an available reference genome or transcriptome, pseudo-alignment tools such as Kallisto (Bray *et al.*, 2016) and Salmon (Patro *et al.*, 2017) which bypass full read alignment. At the gene level, there are two main approaches to read assignment. One method directly counts reads that overlap gene features. Alternatively, transcript-level quantifications can be aggregated to gene-level counts (Van den Berge *et al.*, 2019). When a reference genome includes an annotated gene features file, it can be used for gene-level counting. Common annotation databases include Ensembl, RefSeq, UCSC, and GENCODE (Chen *et al.*, 2023). Gene-level quantification considers all reads aligning to any exon within a gene. Tools such as featureCounts (Liao Y. *et al.*, 2014), HTSeq (Anders *et al.*, 2015), and STAR (Dobin *et al.*, 2013) provide splice-aware quantification by counting fragments mapped to genes (Fig. 8).

When the number of sequences associated with a gene or transcript is calculated, `featureCounts`, a program within the `Rsubread` package in R can be used (Liao Y. *et al.*, 2014). RNA-Seq analysis consists of two main steps. The first involves transforming count data for visual assessment of sample relationships. The second uses raw counts for statistical analysis, as statistical tests require unaltered data to accurately estimate measurement precision (Love *et al.*, 2016). Following the initial step, data filtering is necessary. Since count data often contain rows with zero or single counts, these should be omitted from further analysis (Love *et al.*, 2016). Next, variance stabilization was performed. We used the `rlog` transformation for exploratory visualization where sample size was moderate, however if the dataset is medium to large size, the `rlog` is computationally intensive and variance-stabilizing transformation (VST) is recommended (Love *et al.*, 2016). To perform differential expression analysis, the `DESeq` function is employed, which processes raw count data according to the predefined experimental design. The main steps involve calculating size factors to normalize differences in sequencing depth across samples, estimating dispersion values for each gene, and fitting a generalized linear model to the data. The resulting `DESeqDataSet` object retains all computed parameters, allowing for the extraction of relevant result tables for further analysis (Love *et al.*, 2014; 2016).

Transcriptomic research in *Prunus* species has largely focused on fruit quality traits due to their high commercial and nutritional value. Integrated transcriptomic and metabolomic studies have revealed that genes involved in the phenylpropanoid, flavonoid, carotenoid, and carbohydrate pathways are among the most dynamically regulated during ripening (García-Gómez *et al.*, 2020). For instance, carotenoid cleavage dioxygenase 4 (*CCD4*) has been linked to carotenoid accumulation and light-yellow or white flesh, while sucrose synthase (*SUSY*) expression parallels sugar content, highlighting their potential as molecular markers for breeding programs targeting flesh color and sweetness. Cultivar- and stage-specific expression of carotenoid biosynthesis genes, cell wall-modifying enzymes, *SUSY*, and *LOX2* correlates with differences in fruit firmness and ripening patterns (Ortuño-Hernández *et al.*, 2024). Hormonal treatments, particularly ethylene and its inhibitors, alter the expression of *pectinesterase*, *polygalacturonase*, *IAA*, *ferredoxin*, and *aquaporin* genes, emphasizing the role of hormone signaling in regulating texture and ripening (Salazar *et al.*, 2022).

Beyond candidate gene studies, network-based approaches have enhanced understanding of complex traits. A large-scale peach coexpression network (COO300) has helped identify regulators of fruit softening and anthocyanin accumulation, including *PpMYB10.1* (Pérez de los Cobos *et al.*, 2024). Structural variations, such as transposable element insertions, also modulate

gene expression. In sweet cherry, a retrotransposon in the promoter of *MGST* reduces transcript levels, affecting pollen-part self-compatibility (Ono *et al.*, 2018), whereas in Japanese plum, an insertion in *PsMYB10.2* enhances expression, promoting anthocyanin accumulation and red-flesh coloration (Fiol *et al.*, 2022). Similarly, TE-associated promoter methylation contributes to downregulation of genes such as *cytochrome P450* and *UDP-arabinose epimerase* in peach (Rothkegel *et al.*, 2021).

Transcriptomic studies in reproductive tissues remain limited but reveal key regulatory networks. In Chinese cherry and *P. sibirica*, flowering and dormancy were linked to stress, hormone, and trehalose pathways (Zhu *et al.*, 2015; Xu *et al.*, 2022). Floral pigmentation and morphology genes were reported in *P. humilis* and *P. mume* (Zhang *et al.*, 2024; Zhu *et al.*, 2022; Xiujun *et al.*, 2022). miRNAs (*miR156*, *miR172*, *miR167*) regulate floral bud break (Zhang *et al.*, 2022), while pistil development and self-incompatibility involve thousands of DEGs (Shi *et al.*, 2020; Gómez *et al.*, 2019). These studies demonstrate that transcriptional and post-transcriptional regulation coordinates reproductive development and stress responses in *Prunus*.

Additional studies provide detailed insights into specific traits. In red-fleshed *P. salicina*, Fang *et al.* (2016) identified 13 DEGs involved in anthocyanin biosynthesis, confirming roles for *MYB* and *bHLH* transcription factors, with similar results in *P. persica* and *P. avium* (Liu *et al.*, 2019; Wei *et al.*, 2015). Genes affecting almond kernel size (Jafari *et al.*, 2022) and regulators of *P. mume* aroma compounds (Xiujun *et al.*, 2022) have also been described. Collectively, these studies demonstrate that integrating transcriptomics with metabolomics, coexpression networks, and epigenetic analyses provides a robust framework for understanding the genetic control of both fruit quality and reproductive traits in *Prunus* species.

3.7 Molecular markers developed for genetic analysis of *Prunus* species

The biology of *Prunus* species imposes inherent constraints on breeding and genetic improvement. As perennial tree crops, these species exhibit extended juvenile periods of three to five years before flowering and fruiting, and many cultivated gene pools have limited genetic variation. Complex reproductive systems, including gametophytic self-incompatibility, polyploidy in some cherries and plums, and scion–rootstock incompatibilities, further reduce the efficiency of selection and increase the time and cost required to achieve genetic gain (Carrasco *et al.*, 2013; Pinosio *et al.*, 2020). Conventional breeding, which relies primarily on phenotypic

evaluation, is slow and often ineffective for traits with complex inheritance or strong environmental influence. These limitations are further exacerbated by biotic stresses, such as *Monilinia* spp. and Plum pox virus, and abiotic stresses driven by climate change. Together, these factors underscore the urgency of rapidly identifying, introgressing, and managing adaptive variation from wild and underutilized germplasm (Mariette *et al.*, 2016; De Mori & Cipriani, 2023; Hegedűs *et al.*, 2023). Gene-targeted molecular markers have therefore received growing attention because they sample genic regions and are more likely to be linked to functional variation.

The integration of molecular markers into *Prunus* breeding programs has substantially increased efficiency and precision. Marker-assisted selection (MAS) enables early identification of desirable genotypes based on markers linked to key traits such as fruit shape, acidity, ripening time, and self-compatibility, reducing the need for large progeny populations and lengthy field trials. At IRTA, MAS has been applied to thousands of peach and almond seedlings, accelerating the development of cultivars with targeted characteristics (Pérez de los Cobos *et al.*, 2024). For example, producing 100 peach hybrids with nectarine type, yellow flesh, acidity, and flat shape through conventional breeding would require approximately 12,800 seedlings; with reliable markers tracking all target traits, this number drops to about 124 seedlings, representing a roughly 100-fold reduction in land and labor requirements. Beyond selection, MAS allows controlled introgression of alleles from related species, such as disease resistance and novel quality traits, and genome-wide marker-based strategies can reconstruct elite genotypes with improved traits without relying on genetic engineering. These applications illustrate how molecular markers can accelerate breeding, reduce costs, and enable outcomes that would otherwise be difficult or impossible to achieve through conventional phenotypic selection (De Mori & Cipriani, 2023).

Several marker systems have been applied to study genetic diversity and support breeding in *Prunus*. Simple sequence repeats (SSRs) are codominant, reproducible, and generally highly polymorphic, making them suitable for cultivar identification, parentage analysis, and population structure studies (Decroocq *et al.*, 2004; Dirlewanger *et al.*, 2002; Halász *et al.*, 2019). However, most SSR loci are located in non-coding regions, so they frequently lack direct functional association with traits. Variants such as expressed sequence tag-simple sequence repeats (EST-SSRs) and SSRs in regulatory or exon regions are derived from expressed or regulatory sequences, which increases the likelihood of functional association with phenotype. For instance, EST-SSR markers developed from flavonoid pathway genes in *Prunus salicina*

have successfully clustered cultivars according to fruit color, illustrating their biological relevance (González *et al.*, 2016). While these markers are more labor-intensive to develop and may exhibit reduced allelic diversity due to selective constraints, they provide greater interpretability for trait-focused studies and marker-assisted selection (García-Gómez *et al.*, 2018).

Dominant PCR-based markers such as RAPD, ISSR, and AFLP offer inexpensive, rapid surveys of genome-wide diversity. These markers are useful for preliminary assessments but have significant limitations, including the inability to distinguish heterozygotes, anonymity of loci, and limited reproducibility and transferability across laboratories and taxa, which restricts their use in breeding-focused applications (Martínez-Gómez *et al.*, 2003; Poczai *et al.*, 2013). Start codon targeted (SCoT) markers preferentially amplify coding-region flanks using a single primer, producing highly polymorphic fragments suitable for diversity surveys and population structure analyses. Studies in *Prunus* have demonstrated that SCoT markers can reveal strong within-population diversity and high discrimination among cultivars (Buer *et al.*, 2022; Thakur *et al.*, 2022; Atapour *et al.*, 2022). Despite their utility, SCoT markers are dominant and produce anonymous fragments, which limits their direct application in MAS and fine mapping without further sequencing.

Transposable element (TE)-based markers, including retrotransposon- and DNA transposon-anchored assays, iPBS, S-SAP, and MITE-TD, exploit the presence and recent activity of TEs to capture insertional polymorphisms, structural variation, and lineage-specific differences (Roy *et al.*, 2015; Baránek *et al.*, 2012). Unlike SSRs or SNPs, TE-based markers detect variants that reflect recent genome dynamics and structural changes. TE insertions can disrupt coding sequences, modify regulatory regions, and influence gene expression, which may produce phenotypic effects subject to selection. This makes TE-based markers particularly informative for detecting variation in traits such as self-compatibility, fruit development, and stress responses (Halász *et al.*, 2014; Muñoz-Sanz *et al.*, 2017). These markers also reveal structural rearrangements, duplications, deletions, and recombination events, capturing a layer of genome evolution that SSRs and SNPs alone cannot detect. While TE-based markers are typically dominant and may be affected by homoplasy in highly active families, they provide high resolution for distinguishing closely related cultivars and populations, which is critical for germplasm characterization, fingerprinting, and pre-breeding.

By characterizing recent structural variation across germplasm, identifying candidate TE insertions influencing reproductive and fruit-quality traits, and developing practical marker

panels for early selection, TE-based and gene-targeted markers are particularly appropriate. TE-anchored assays detect functionally relevant variants and lineage-specific insertions, providing high-resolution discrimination among closely related cultivars and wild populations. When integrated with EST-SSRs and SCoT markers, they complement codominant SSR and SNP datasets, maximizing coverage of both allelic and structural variation. PCR-based TE and SCoT assays also offer cost-effective workflows for screening large seedling populations and underutilized germplasm collections where high-throughput sequencing may not be feasible.

In conclusion, molecular markers differ in the genomic layers they sample and the evolutionary processes they reflect. SSRs and SNPs provide codominant genotyping for allelic variation, EST-SSRs and regulatory SSRs capture functionally relevant variation, SCoT markers enrich for gene-associated polymorphisms, and TE-based markers reveal structural, regulatory, and recent lineage-specific variation. Combining TE-based markers with EST-SSRs and SCoT markers maximizes the detection of biologically meaningful variation, supports germplasm characterization, and enables efficient early selection, bridging the gap between genotype and phenotype while enhancing the applied utility of molecular tools in *Prunus* breeding.

4. MATERIALS AND METHODS

4.1 Plant material

4.1.1 Plant material for transcriptome analyses

Three apricot cultivars ('Pannónia', 'Ceglédi óriás' and 'Gönci magyar kajszi'), one sweet cherry cultivar ('Carmen'), one sour cherry open-pollinated seedling, and one hexaploid plum cultivar ('Hanita') were used for the PCR amplification with *FaSt* specific primers. These accessions are maintained at the experimental orchard at Soroksár, Budapest, Hungary. The polyploid species were included intentionally to evaluate primer specificity and amplification behavior across different ploidy levels within *Prunus*. For each cultivar, floral and vegetative tissues were collected during spring from field-grown trees: flower buds at the balloon stage (late pre-anthesis), anthers and pistils (dissected from the same flowers so that male and female reproductive tissues are paired), petals and styles (from the same flowers), and fully expanded mature leaves (vegetative control). For each tissue and cultivar we collected three independent biological replicates and each biological replicate consisted of tissue pooled from three individual flowers (or three leaves) taken from a single tree. Sampling was performed between 09:00 and 11:00 local time to minimize diurnal variation. Samples were immediately frozen in liquid nitrogen and stored at -70°C until RNA extraction.

4.1.2 Plant material for molecular marker development

Two sets of experiments were carried out, the first on two cultivars of six *Prunus* species, *P. armeniaca* L. (apricot), *P. dulcis* (Mill.) D.A. Webb (almond), *P. avium* L. (sweet cherry), *P. persica* (L.) Batsch (peach), *P. cerasus* L. (sour cherry) and *P. domestica* L. (European plum). In the second experiment, 28 cultivars of European plum were analyzed; all cultivars are listed in Tab. 2. The trees were sampled in the *Prunus* germplasm collection of the Department of Plant Biotechnology, Hungarian University of Agriculture and Life Sciences (MATE, Budapest, Hungary).

Table 2. Origin of cultivars used in the study

Cultivar name	Species	Pedigree¹	Country of origin¹
Besztercei Bt2	European plum	Clone of Besztercei (1974)	Hungary
Bühler Frühzwetschge	European plum	German landrace (1840)	Germany
Canada giant (Sumgita)	sweet cherry	Van × Sam (1992)	Canada
Ceglédi óriás	Apricot	Local selection, Izsák (1953)	Hungary
Collins	Peach	Jerseyland × NJ-188 (1955)	USA
Condor	Peach	Unknown	USA
Czar	European plum	Prince Engelbert × Early Prolific (1874)	United Kingdom
Duránci	European plum	Hungarian landrace (Borsod)	Hungary
Elena	European plum	Fellenberg × Stanley (1993)	Germany
Empress	European plum	Unknown (2000)	Italy
Fehérszilva	European plum	Hungarian landrace (Duna- Tisza-köz, Borsod, Bácska) (2014)	Hungary
Feketicsi meggy	sour cherry	Landrace, Bácsfeketehegy	Serbia
Giant	European plum	Ageni × Pond's Seedling (1893)	USA
Gönci magyar kajszi	Apricot	Selected clone of Hungarian Best (1960)	Hungary
Grand prize	European plum	Unknown seedling of Burbank (1937)	USA
Haganta	European plum	Cacanska najbolja × Valor (2003)	Germany
Hanita	European plum	President × Auerbacher (1980)	Germany
Haroma ²	European plum	(Ortenauer × Stanley 34) × Hanita (1993)	Germany
Jojo	European plum	Ortenauer × Stanley (1991)	Germany
Kántorjánosi 3	sour cherry	Selected clone of a landrace, Kántorjánosi (1994)	Hungary
Katalin	sweet cherry	Germersdorfer × Podyebrad (1989)	Hungary
Nemtudom P3	European plum	Selected clone of a landrace (2012)	Hungary

Cultivar name	Species	Pedigree ¹	Country of origin ¹
Opal	European plum	Early Favourite × Oullins gage (1925)	Sweden
Páczelt szilvája	European plum	Unknown, Páczelt János, Nagykároly (end of the 19th century)	Hungary
Presenta	European plum	President × Ortenauer (1996)	Germany
Sanctus Hubertus	European plum	Mater Dolorosa × Early Rues (1963)	Belgium
Soviet 2	European plum	Soviet seedling	Russia
Soviet 3	European plum	Soviet seedling	Russia
Tétényi keményhájú	almond	Open pollination from Burbank magonca (1980)	Hungary
Topend plus	European plum	Cacanska najbolja × Valor (1994)	Germany
Topfive	European plum	Cacanska najbolja × Bühler Frühzwetsche (1999)	Germany
Tophit	European plum	Cacanska najbolja × President (1988)	Germany
Toptaste ²	European plum	Valor × Hauszwetschke (1994)	Germany
Tuleu timpuriu	European plum	Tuleu gras × Sermina (1967)	Romania
Tuono	almond	Unknown (~1830)	Italy
Vengerka virila	European plum	Unknown	Russia
Vörös nagygyümölcsű	European plum	Landrace	Hungary
Wagenheim	European plum	Unknown (1837)	Germany

¹The pedigree and country of origin of accessions are given according to Surányi (2013) and Makovics-Zsohár *et al.* (2017)

²European plum cultivars included in both sets of experiments.

4.1.3 Plant material for RNA-Seq analyses

Samples for RNA-Seq were collected in a single year (2021) during blossoming at the balloon stage of flower development. For each accession and tissue (anther or style), three independent biological replicates were analyzed. Each biological replicate consisted of tissue pooled from approx. 40 flowers collected from a single tree, where available, replicates were obtained from different trees of the same accession. Samples were immediately frozen in liquid nitrogen and stored at −70 °C until RNA extraction.

Total RNA was isolated from each biological replicate independently using a standard phenol–chloroform extraction protocol. RNA quality and concentration were assessed prior to library preparation. For RNA-Seq library construction, 50 ng total RNA per biological replicate was used. Libraries were prepared and sequenced separately for each replicate.

RNA was extracted from the anthers of 3 SC cultivars ('Gönci magyar kajszi' AMK, 'Pannónia' APA and 'Rózsakajsz C. 1406' ARK) and 3 SI accessions ('Ceglédi óriás' ACO, M604 AM6 and 'Harmat' AHA) apricot cultivars and from the styles of the same 3 SC ('Gönci magyar kajszi' SMK, 'Pannónia' SPA and 'Rózsakajsz C. 1406' SRK) and 3 SI accessions ('Ceglédi óriás' ACO, M604 AM6 and 'Harmat' AHA).

4.2 RNA and DNA extraction

4.2.1 RNA extraction and cDNA synthesis for transcriptome analyses

Total RNA was extracted from each biological replicate independently. For qRT-PCR analysis, samples were extracted using the CTAB method. RNeasy Mini Kit (Qiagen, Hilden, Germany) including on-column RNase-free DNase I treatment (Qiagen) following the manufacturer's instructions was used for RNA used reverse transcribed to cDNA and used for the transcriptome PCR assays. For mRNAseq analysis, a standard phenol–chloroform method extraction protocol was used.

Extracted RNA concentrations were measured by fluorometric assay (Qubit) and RNA integrity was assessed by Agilent Bioanalyzer (RNA 6000 Nano/Pico) or, when a Bioanalyzer was not available, by 1% agarose gel electrophoresis. For RNA-Seq, samples were required to meet a pre-specified integrity threshold ($RIN \geq 7$) to proceed to library construction. For qRT-PCR, the RNA concentration would have required an impractically large volume for 1 µg input (e.g., >50 µL), the sample was re-extracted or excluded. All retained samples were treated with RNase-free DNase I prior to reverse transcription.

Reverse transcription was performed on DNase-treated RNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) with oligo(dT)18 primers according to the manufacturer's instructions. For qRT-PCR assays, 1 µg total RNA (per replicate) was used as input where feasible. When starting material was limiting, the maximum feasible input was used and cDNA reaction volumes were equalized across samples. Synthesized cDNA was stored at -20 °C until use.

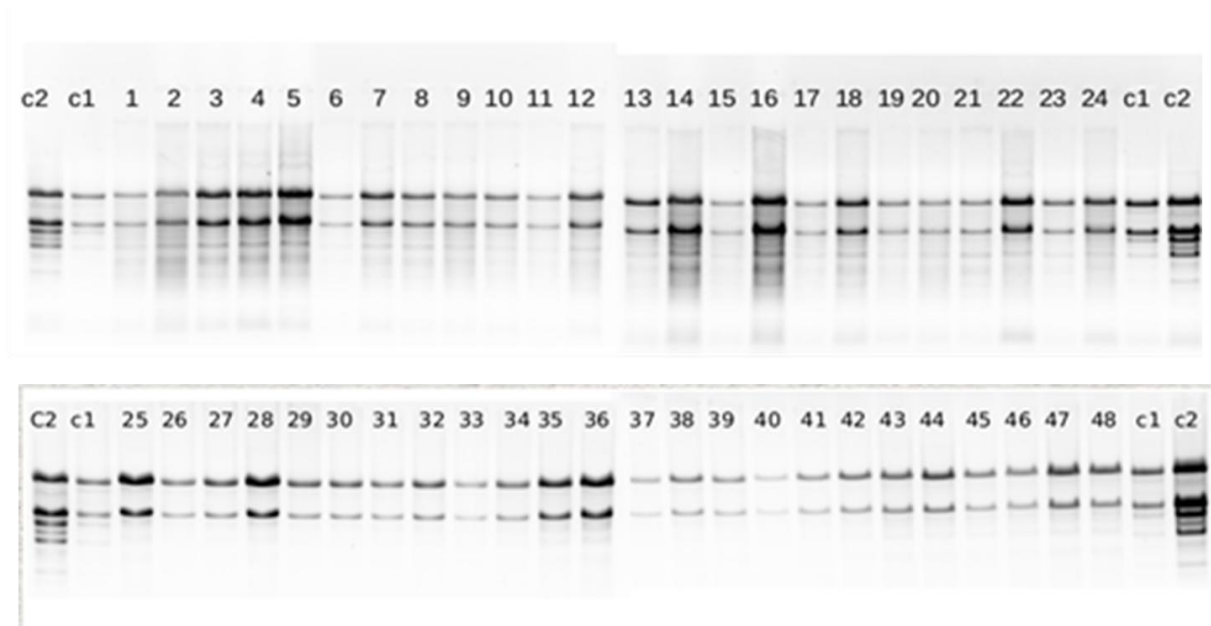


Figure 1. Agarose gels of total RNA extracted from style and anther tissues from apricot samples. C1 and C2 are controls from *Arabidopsis thaliana* Columbia; Style tissue from: 1-6, ‘Pannónia’ SPA, 7-12 and 24 ‘Rózsakajsi C. 1406’ SRK, 13-18 (‘Ceglédi óriás’ SCO, 19-24 ‘Harmat’ SHA; Anther tissue from 25-30, ‘Pannónia’ APA, 31-36 and 24 ‘Rózsakajsi C. 1406’ ARK, 37-42 ‘Ceglédi óriás’ ACO, 43-48 ‘Harmat’ AHA

4.2.2 Genomic DNA extraction

Total genomic DNA was isolated from buds using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. DNA concentration and purity were measured for all samples with a NanoDrop™ ND-1000 spectrophotometer (Bio-Science, Budapest, Hungary). DNA integrity was assessed by electrophoresis on 1% (w/v) agarose gels stained with ethidium bromide. Intact high-molecular-weight genomic DNA presented as a single predominant band with minimal smearing. Functional integrity was additionally confirmed by PCR: samples that produced the expected amplification products with the study primers were retained for downstream analyses. For genomic PCR assays, 40–70 ng genomic DNA was used per reaction.

4.3 Sequencing and library preparation for RNA-Seq analyses

cDNA libraries were prepared using poly-T oligo-attached magnetic beads to isolate messenger RNA from total RNA by Novogene (Beijing, China). After fragmentation, first-strand cDNA synthesis was conducted using random hexamer primers, followed by second-strand synthesis with dUTP incorporation for directional libraries. cDNA libraries were prepared and

QC-checked by Novogene (Beijing, China). Twelve libraries were prepared and sequenced on an Illumina NovaSeq 6000 platform (150 bp paired-end). Libraries were multiplexed using unique Illumina index sequences and demultiplexed by Novogene. The project targeted ~12 Gb (~80 million 150 bp paired-end reads) per library. After provider quality control, per-library raw read counts ranged approximately 80.3–87.1 million reads (12.12–13.07 Gbp raw bases) with 78–85 million clean reads per library (clean bases 11.7–12.8 Gbp) (Novogene report, QC summary). Full QC and mapping metrics are provided in the Novogene report (Contract ID H204SC22012203).

4.4 PCR analyses

4.4.1 PCR parameters and detection of amplified fragments for transcriptome analyses

A pair of primers was used to amplify the *FaSt* sequence from cDNA: FaSt-F (5'-GAGATATTTAGTGATATACCCATT-3') and FaSt-Rev5 (5'-CCCAATATGAAGACCCAAATTA-3'). The designation “Rev5” denotes reverse primer number 5 from an initial set screened during optimization. Namely, FaSt-Rev5 produced a single, correctly sized product on cDNA templates and was therefore selected for the transcriptome PCR assays. Primer properties (melting temperature, hairpin, self- and hetero-dimers) were evaluated using the OligoAnalyzer tool (IDT; <https://eu.idtdna.com/calc/analyzer>). For genomic PCR targeting the Falling Stone MITE, a separate genomic primer (FaSt-R; 5'-TCTTAGAAATTACAAAACCTACC-3') was designed and used as described in the genomic PCR section.

Actin was employed as an internal control for conventional PCR to verify cDNA amplifiability. The use of actin follows its widespread application in *Prunus* and related *Rosaceae* studies and was supported by our empirical verification. Actin primers consistently amplified the expected fragment in all genotypes retained for *FaSt* analyses (the single genotype that failed to amplify actin was excluded). We acknowledge that reference-gene stability is context dependent and that other genes may be preferable for qRT-PCR normalization in specific experimental designs (Niu *et al.*, 2014; Wang *et al.*, 2014; Tong *et al.*, 2009). In this work actin served strictly as a control of successful reverse transcription and PCR amplifiability rather than as the primary normalization gene for quantitative expression analyses. Reaction conditions were as follows: initial denaturation at 98°C for 30 s and 35 cycles of 98°C for 5 sec, 50°C for 10s, 72°C for 40s and 72°C for 1 min. PCR was carried out in a Swift MaxPro (Esco Healthcare Pte)

thermocycler. 1-2 µl cDNA was used for PCR amplification in a 20 µl reaction volume containing 4 µl 5X Phire Reaction Buffer, 0.2 mM dNTPs, 0.5 µM of the adequate primers, 3 % DMSO, and 0.4 U of Phire Hot Start II DNA polymerase (Thermo Fisher Scientific). The PCR products were separated on 1% TBE agarose gels at 80 V for 30 min and DNA bands were stained with ethidium bromide. Fragment sizes were estimated by comparison with the 1 kb and 100bp + DNA ladder (Thermo Scientific).

4.4.2 Real-Time PCR

Real-time PCR was performed on a BIO-RAD CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories) using PCRBIO SyGreen Mix (PCR Biosystems Ltd., London, UK). Mastermix was prepared in a final reaction volume of 8 µL comprising: 4 µL PCRBIO SyGreen Mix, 1 µL cDNA template, forward and reverse primers at a final concentration of 200 nM each, and molecular-grade water to 8 µL. All reactions were run in three technical replicates and included a no-template control on each plate. Cycling conditions were: 2 min initial denaturation at 98 °C, followed by 40 cycles of 98 °C for 10 s (denaturation), 55 °C for 20 s (annealing) and 60 °C for 30 s (extension). Melt-curve analysis was performed from 65 °C to 95 °C with increments of 0.5 °C to verify product specificity. A four-point dilution series of pooled cDNA (serial dilutions prepared from a cDNA pool) was used to generate standard curves for each assay; these curves were used to evaluate linear dynamic range and to calculate amplification efficiencies. Relative fold change (FC) was calculated with the $\Delta\Delta C_q$ method (Bookout & Mangelsdorf, 2003), after confirming approximately similar amplification efficiencies between target and reference assays (Fig. 2).

Table 3. Primers used for qRT-PCR

Primer name	Sequence (5' → 3')	Target	Amplicon size (bp)	Assay
FaSt-F	GAGATATTTAGTGATATACCCATT	FaSt element	331	Conventional PCR
FaSt-Rev5	CCCAATATGAAGACCCAAATTA	FaSt element	331	Conventional PCR
FaSt-R	TCTTAGAAATTACAAACTACC	Falling Stone MITE	~variable (see text)	Genomic PCR
FaStqRT-F1	GAGATATTTAGTGATATACCCATT	FaSt element	116	qRT-PCR

Primer name	Sequence (5' → 3')	Target	Amplicon size (bp)	Assay
qRT-Rev1	TCAATAAGGCCAGTTGTCAT	FaSt element	116	qRT-PCR
qRT-FW2	TACTTTAATGCCCTATTGAGTG	FaSt element	96	qRT-PCR
qRT-Rev2	TCTTAGAAATTACAAAACCTACC	FaSt element	96	qRT-PCR
UBQ10-F	AAGGCTAAGATCCAAGACAAAGAG	Ubiquitin 10	146	qRT-PCR (reference gene)
UBQ10-R	CACGAAGACGAAGCACTAAG	Ubiquitin 10	146	qRT-PCR (reference gene)
ACT-F	CTCCGTGTTGCTCCAGAAG	Actin	~130	Conventional PCR (control)
ACT-R	CAGGCACGTTGAAGGTTTC	Actin	~130	Conventional PCR (control)

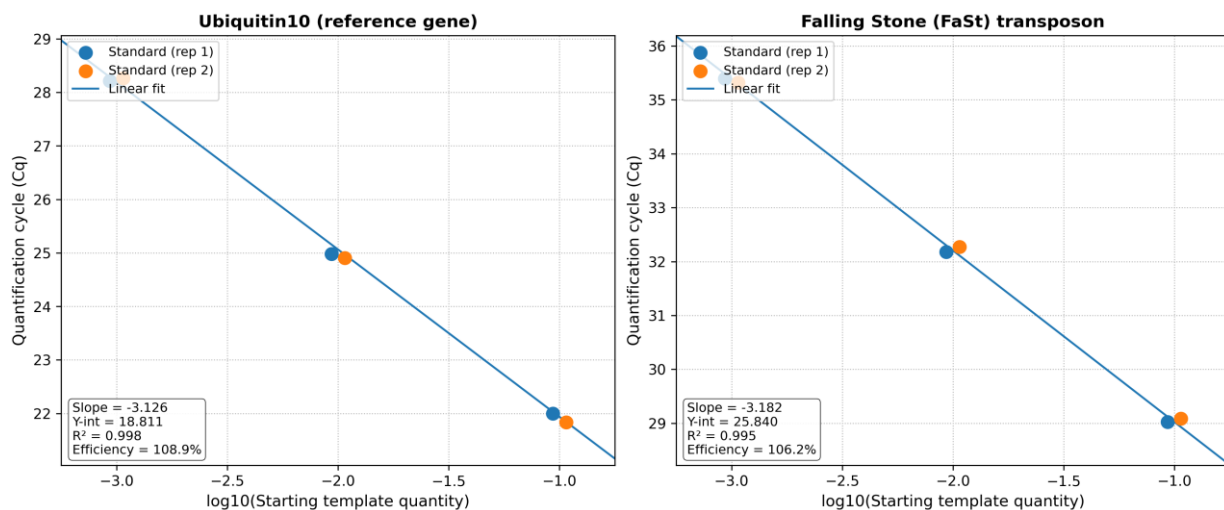


Figure 2. Calibration curve from almond anther tissue with *ubiquitin10* gene as a reference gene (on the left) and almond anther tissue with *Falling Stone (FaSt)* transposon sequence (on the right).

4.4.3 Genomic PCR with SCoT and *FaSt*-specific primers

For genomic PCR, 19 (1, 2, 4-7, 11-13, 16, 18, 19, 21-24, 27, 32 and 34) and 17 (1-5, 7-11, 15, 16, 19, 21-23, and 34) SCoT primers were used for two cultivars each of six *Prunus* species and 28 *P. domestica* cultivars, respectively. A new primer, FaSt-R (5'-TCTTAGAAATTACAAAACCTACC-3') was designed to anneal to a unique sequence of the *Falling Stone* MITE. The FaSt-R primer was used in two complementary ways: (i) paired with

FaSt-F (conventional *FaSt* amplicon detection) to confirm presence of the *FaSt* sequence itself, and (ii) combined with individual SCoT primers (one SCoT primer plus FaSt-R in the same reaction) to test for *FaSt* insertions adjacent to SCoT-amplified loci. Using FaSt-R in combination with SCoT primers allows detection of insertion-specific bands (presence/absence polymorphisms) and provides gene-context information for *FaSt* insertions. SCoT primers used alone provide independent, gene-targeted polymorphism profiles across the genome (Collard and Mackill, 2009). . Approximately 40–70 ng of genomic DNA was used for PCR amplification in a 12 μ L reaction volume, containing 10 $\times$ DreamTaq™ Green buffer (Thermo Fisher Scientific, Waltham, MA, United States) with final concentrations of 1.5 mM MgCl₂, 0.2 mM of dNTPs, 0.4 μ M of the adequate primers, and 0.625 U of DreamTaq™ DNA polymerase (Thermo Fisher Scientific). The PCR amplification was carried out in a Swift MaxPro thermocycler (ESCO Healthcare) . Amplification was run at 94 °C for 5 min, followed by 40 cycles of denaturation at 94 °C for 1 min, primer annealing at 50 °C for 1 min and elongation at 72 °C for 2 min, final extension was 10 min at 72 °C. Amplicons were detected by 1.2 % TBE agarose gels stained with ethidium bromide (EtBr) at 80 V for 3 h. Fragment sizes were estimated by comparison with the GeneRuler™ 1-kb DNA Ladder (Thermo Fisher Scientific).

4.5 Cloning and DNA sequencing

Direct cloning of PCR products was carried out using the InsTAclone PCR Cloning Kit (Thermo Fisher Scientific) and JM109 *Escherichia coli* competent cells (Zymo Research Corp., Irvine, CA, United States). TNucleotide sequences of 3 samples were determined for each fragment in both directions by using M13 sequencing primers. The plasmid DNA fragments were purified using the EZ-10 Spin Column Plasmid DNA kit (Bio Basic Inc., Markham, Canada) and then sequenced in an automated sequencer ABI 3500 XL Genetic Analyzer (Applied Biosystems, Foster City, CA, United States).

4.6 Data evaluation and bioinformatics analyses

4.6.1 Data evaluation and bioinformatics analyses for molecular marker development

Banding pattern was scored as absent (0) or present (1) with only the strong and fully separated bands considered. Weak or doubtful bands were neglected. Total number of scorable bands (TNB), number of differently sized bands (NDB), size range of amplicons (SRA), percentage of amplicons with size $\leq$ 1000 bp (PA $\leq$ 1kb) and percentage of the polymorphic bands

(PPB) were calculated in MS Excel, while polymorphism information content (PIC), Nei's gene diversity (h), Shannon's Information index (I) and Resolving power (R_p) were determined using the iMEC (<https://irscope.shinyapps.io/iMEC/>) (Amiryousefi *et al.*, 2018) and Popgene 1.32 (Yeh and Boyle, 1997) programs. PCA and a neighbor-joining (NJ) tree based on Jaccard's index were constructed using the software PAST 4.03 (Hammer *et al.* 2001). To examine overall genetic relationships among accessions, principal component analysis (PCA) was performed on the binary marker matrix using PAST v4.03 (Hammer *et al.*, 2001). PCA was used to visualize multivariate patterns of marker variation and to assess clustering tendencies among *Prunus* species, as reported in the Results section.

Table 4 *Prunus* genomes used and the BLASTN search parameters used for homology calls.

Species	Cultivar / Accession	Genome version	Databas e
<i>Prunus cerasus</i>	'Montmorency'	v1.0.a2	GDR
<i>Prunus avium</i>	'Tieton'	v2.0	GDR
<i>Prunus fruticosa</i>	27e12(2)	v1.0	GDR
<i>Prunus dulcis</i>	'Lauranne'	v1.0.a1	GDR
<i>Prunus dulcis</i>	'Nonpareil'	v1.0	GDR
<i>Prunus dulcis</i>	'Texas'	v3.0	GDR
<i>Prunus armeniaca</i>	'Stella'	v1.0	GDR
<i>Prunus persica</i>	—	Genome V2.0.a1	GDR
<i>Prunus yedoensis</i>	—	Genome v1.0	GDR
<i>Prunus domestica</i>	—	Draft Genome v1.0	GDR
<i>Prunus campanulata</i>	—	v1.0	GDR
<i>Prunus kansuensis</i>	—	genome v2.0	GDR
<i>Prunus mume</i>	'Nanko'	v1.0	GDR
<i>Prunus salicina</i>	'Sanyueli'	genome v2.0	GDR
<i>Prunus speciosa</i>	IZO01	v1.0	GDR
<i>Prunus davidiana</i>	—	Genome v2.0	GDR
<i>Prunus mandshurica</i>	—	genome v1.0	GDR
<i>Prunus mongolica</i>	—	v1.0	GDR
<i>Prunus mira</i>	—	Genome v2.0	GDR
<i>Prunus sibirica</i>	CH320_5	genome v1.0	GDR
<i>Prunus zhengheensis</i>	—	v1.0	GDR

In parallel, sample distance analysis was conducted using Jaccard's similarity coefficient, and a neighbor-joining (NJ) dendrogram was constructed from the resulting distance matrix using PAST v4.03. Node robustness was evaluated by 1000 bootstrap replicates, and bootstrap values were displayed on the corresponding branches. These distance-based analyses were used to assess genetic affinities and groupings consistent with PCA patterns. The BlastN analysis at the National Center for Biotechnology Information (NCBI) and the available genome sequences of the *Prunus* species at the GDR (<https://www.rosaceae.org/blast>) databases were used for homology searches (Altschul *et al.*, 1990). The BLASTN 2.12.0+ analysis at GDR (Zhang *et al.*, 2000) was carried out using *Prunus* species genomes (Tab. 4). Sequences were aligned and presented with MEGA7 (Kumar *et al.* 2016) and BioEdit v. 7.2.0. (Hall 1999), respectively. The phylogenetic chronogram of 12 *Prunus* species and one *Malus* species (Figure 20) was reconstructed based on divergence time estimates obtained from the TimeTree 5 database (<https://timetree.org>; Kumar *et al.*, 2022), using calibration points reported by Chin *et al.* (2014). Divergence times were expressed in millions of years ago (Mya), and the geologic timescale was displayed to contextualize lineage diversification. The presence or absence of *FaSt* transposon-associated insertions and deletions was mapped onto the chronogram to infer lineage-specific sequence alterations, as described in the Results section. Finally, the relative frequency of *FaSt* transposon insertions across *Prunus* genomes was visualized using a heatmap generated with Heatmapper (Babicki *et al.*, 2016).

4.6.2 Bioinformatic tools for *in silico* analyses of the presence of *FaSt* in the promoter region of the *CBF* gene

DNA sequences were compared using the NCBI BLASTN 2.11.0 software (Altschul *et al.*, 1990). The aligned sequences were manually edited and presented using the BioEdit program v.7.2.5 (Hall, 1999). *PdCBF1* (JQ317157) and *PdCBF2* (JQ317158) sequences have been downloaded from NCBI GenBank database (<http://www.ncbi.nlm.nih.gov>). The transcription factor binding sites were predicted using the PlantRegMap tool (Tian *et al.*, 2020). and visualized by the Plant Promoter Analysis Navigator (PlantPAN; <http://PlantPAN2.its.ncku.edu.tw>). We used the PlantRegMap tool with the species-specific binding motifs of *P. mume* to predict transcription factor binding sites in the promoter region of *PdCBF1* and *PdCBF2*. On the grounds that an additional 374 bp were present in the *PdCBF1* promoter sequence, the *FaSt* element had brought the chance for binding of other TFs, thus, suggesting possible transcriptional changes in comparison with its homologue, *PdCBF2*.

4.6.3 Bioinformatic tools for *in silico* analysis of the occurrence FaSt in *Prunus* genome and transcriptome

We analysed available *P. armeniaca*, *P. persica* and *P. dulcis* transcriptome datasets and eleven *Prunus* genome assemblies (covering seven species) to assess presence and frequency of FaSt copies. Screening of genomes and transcriptomes was performed via the Genome Database for Rosaceae (GDR; <https://www.rosaceae.org>). Similarity searches were conducted with BLASTn using the parameters specified above (E-value threshold 0.001; word_size 11; gap open/extension penalties 5 and 2; mismatch penalty -2; reward 1; no culling limit). Hits were filtered by E-value, percent identity and coverage and overlapping hits collapsed as described; transcript-level frequency was calculated as the percentage of transcripts in a dataset containing at least one qualifying hit, and genomic copy number was reported both as total copies and as copies per megabase (see frequency-estimation calculation).

4.6.4 Data analyses from the qRT-PCR

The qRT-PCR was performed using the CFX96 Touch Real-Time PCR Detection System, and data were analysed with CFX Maestro Software (Bio-Rad Laboratories, n.d.).

4.7 Analyses of RNA-Seq data

4.7.1 Pre-processing of sequencing data

Raw RNA-Seq reads were first assessed for quality using FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequences and low-quality bases were trimmed using Trimmomatic v1.0 (Bolger *et al.*, 2014). After filtering, high-quality reads were aligned to the haploid apricot reference genome *Apricot Genome CUR v1*, cultivar ‘Rojo Pasi3n’ (Campoy *et al.*, 2020), using STAR v2.6 (Dobin *et al.*, 2015) with default parameters (Appendix A2 and A3). Alignment statistics and mapping efficiency for all libraries are summarized in Figure 8.

- FastQC documentation and download: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> . Accessed November 2023.
- Trimmomatic documentation: <http://www.usadellab.org/cms/?page=trimmomatic>. Accessed November 2023.
- STAR documentation: <https://github.com/alexdobin/STAR> . Accessed November 2023.

Full command lines and parameter settings used for trimming and alignment are recorded in Appendix A2–A3 and in the analysis script deposited with the manuscript.

4.7.2 PCA and sample-distance analysis (quality control and exploratory data analysis)

Exploratory ordination and distance metrics were computed to evaluate global sample relationships and to detect outliers prior to differential testing. Implementation details are reproduced here to satisfy reproducibility requirements; the exact R commands executed are included in Appendix A3 and in the deposited R script.

Raw counts in the DESeqDataSet were variance-stabilized with the DESeq2 regularized log transformation: `rld <- rlog(dds, blind = FALSE)`

`blind = FALSE` was used so that the transformation would preserve variance associated with the experimental design (Love *et al.*, 2014). For larger sample sizes, the variance-stabilizing transformation (`vst`) is an acceptable, faster alternative.

Principal components analysis (PCA) was performed on the `rlog` values. We used DESeq2's `plotPCA()` function (default `ntop = 500` most variable genes) and also extracted PCA coordinates (`returnData = TRUE`) for custom plotting with `ggplot2`. PCA plots report percentage variance explained by PC1 and PC2 and are annotated for compatibility status and cultivar. (`ggplot2`: <https://ggplot2.tidyverse.org/> . Accessed November 2023).

Sample-to-sample distances. Two complementary dissimilarity measures were computed:

1. Euclidean distances on `rlog` values:

```
sampleDists <- dist(t(assay(rld)))
sampleDistMatrix <- as.matrix(sampleDists)
```

The Euclidean sample distance matrix was visualized as a hierarchical clustering heatmap using `pheatmap`; when precomputed distances were supplied to `pheatmap`, the same distance matrix was used for both row and column clustering to preserve the dendrogram ordering.

2. Poisson distance on raw counts:

```
poisd <- PoissonDistance(t(counts(dds)))
samplePoisDistMatrix <- as.matrix(poisd$dd)
```

To account for the discrete count structure and mean–variance relationship, the Poisson distance (PoissonDistance <https://cran.r-project.org/web/packages/PoiClaClu/index.html>, accessed November 2023), was computed directly on raw counts (samples as rows) and visualized as an alternative heatmap (pheatmap: <https://cran.r-project.org/web/packages/pheatmap/index.html>, accessed November 2023) and MDS.

Hierarchical clustering used complete linkage unless otherwise noted. Heatmap color palettes were generated with `colorRampPalette(rev(brewer.pal(9,"Blues")))` (RColorBrewer: <https://cran.r-project.org/web/packages/RColorBrewer/index.html>, accessed November 2023) for consistent visual scaling across figures. PCA scatterplots and MDS (cmdscale) visualizations were produced with `ggplot2` and labeled using `ggrepel` for legible sample names.

Biological replicates were expected to show the smallest pairwise distances and to cluster together. Concordant patterns between PCA and both distance heatmaps (Euclidean on rlog and Poisson) were required to support any claims about lack of segregation by compatibility status (SC vs SI) or about cultivar-specific effects.

4.7.3 Quantification of gene expression levels

Gene-level read counts were obtained using *featureCounts* v2.0.8 (Liao *et al.*, 2014), producing raw integer counts of uniquely mapped reads per annotated gene. For descriptive reporting and visualization, raw counts were converted to Fragments Per Kilobase of transcript per Million mapped reads (FPKM) by normalizing for gene length and library size; these FPKM values were used only for expression plots and summary tables, and were not used as input to differential expression testing.

4.7.4 Differential expression analysis

Differential expression analysis was performed in R using DESeq2 (Love *et al.*, 2014). Raw *featureCounts* integer counts (not FPKM) were used as DESeq2 input. Where applicable, cultivar identity was included as a blocking factor in the design (design = ~ cultivar + compatibility); otherwise design = ~ compatibility. Prior to model fitting, genes with row sums ≤ 1 across all samples were removed (`dds <- dds[rowSums(counts(dds)) > 1, ]`). DESeq2's default pipeline (size-factor estimation, gene-wise dispersion estimation, and negative-binomial GLM fitting) was applied unless otherwise indicated in Appendix A3. P-values were adjusted using the Benjamini–Hochberg false discovery rate; genes with $\text{padj} \leq 0.05$ were considered

significant. Complete reproducible commands are provided in Appendix A3. (DESeq2 documentation: <https://bioconductor.org/packages/DESeq2> . Accessed November 2023).

4.7.5 Functional annotation of differentially expressed genes

DEG loci were annotated by retrieving sequence records and functional descriptions from UniProt and NCBI, followed by manual curation of UniProt functional notes to assign concise gene descriptions and putative molecular functions for representative DEGs. All gene identifiers and exported DEG tables are provided as Supplementary Files and in the deposited analysis outputs (UniProt: <https://www.uniprot.org/>; NCBI: <https://www.ncbi.nlm.nih.gov/>; both accessed November 2023).

We did not perform formal over-representation or pathway enrichment testing on the published DEG lists because the absolute number of statistically significant DEGs was limited and therefore insufficient for robust enrichment statistics. Enrichment methods applied to very small or heterogeneous gene sets are prone to low power, instability, and bias (for example gene-length and expression-level effects), which can lead to misleading conclusions if presented without strong supporting evidence. Accordingly, manual curation of UniProt/NCBI annotations was used to conservatively characterize notable DEGs.

5. RESULTS AND DISCUSSION

5.1 Detection of *FaSt*-related transcripts in *Prunus* by RT-PCR

Falling Stones MITE elements were detected in many copies throughout all the 8 chromosomes of *Prunus* species (Halász *et al.*, 2014) and they were also detected in two expressed genes (Halász *et al.*, 2014; Munoz-Sanz *et al.*, 2017). We wanted to see if *FaSt* copies can be found in other transcripts. The cDNA samples were used for PCR analysis with actin primers (used as internal control) that amplified a fragment of 130 bp in *Prunus* samples. The actin amplification was successful in case of *P. domestica* ‘Hanita’, a *P. cerasus* chance seedling, *P. armeniaca* ‘Ceglédi óriás’, ‘Magyar kajszi’ and ‘Pannónia’. *P. avium* ‘Carmen’ did not show amplification of actin fragment and hence was excluded from further analyses. The *FaSt*-specific primers annealing to the TIR regions of the transposon were expected to amplify a fragment of 331 bp. The amplification was successful in all the five studied tissues (leaf, calyx, petals, anthers and styles) of *P. domestica* (Fig. 3). These RT-PCR results provide targeted evidence for *FaSt* transcription in the samples tested.

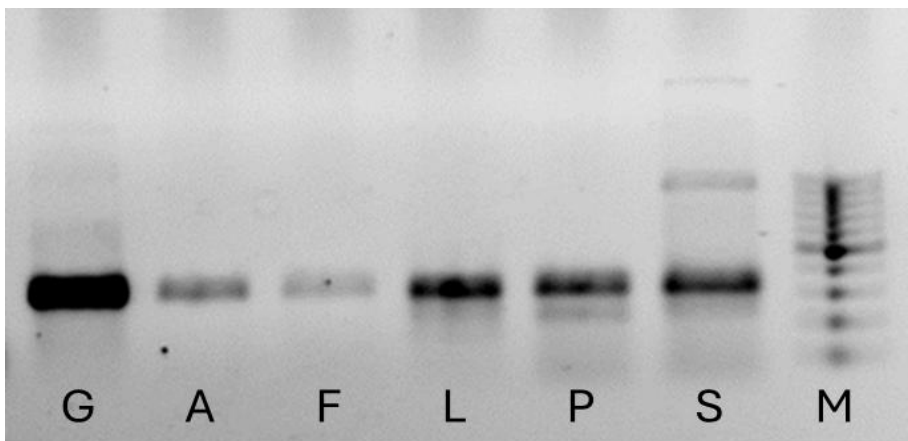


Figure 3. PCR analysis of *Prunus domestica* ‘Hanita’ cultivar with the *Falling Stones* (*FaSt*)-specific primer pair *FaSt*-F and *FaSt*-Rev5, amplifying 331 bp of the *Prunus*-specific non-autonomous transposon. The samples are as follows: and. genomic DNA as control (G), anthers (A), flower buds (F), leaf (L), petal (P), style (S), and 100bp +DNA molecular weight marker (M).

5.2 Real-time PCR detection of *FaSt*-derived transcripts

Using qRT-PCR, *FaSt*-derived transcripts were detected in all tissues examined (leaf, flower buds, calyx, petals, anthers and styles) from the sampled *Prunus* accessions. Detectable expression was observed in the tested sweet cherry and apricot cultivars. The three apricot cultivars included in this study showed qualitatively similar expression patterns. *FaSt* transcripts were also detected in the sampled tetraploid sour cherry and hexaploid European plum tissues. These data indicate that *FaSt* transcription was detectable in every tissue and accession included in our experimental set; they do not imply that *FaSt* is universally transcribed across the entire *Prunus* genus. Details on qRT-PCR calibration, threshold selection and Ct determination (Fig. 2). Quantitative analysis of *FaSt* signal across four tissues of *P. armeniaca* ‘Ceglédi óriás’ showed the highest relative signal in anther tissue (Fig. 4). This observation is descriptive and was not subjected to formal statistical testing. Likewise, a comparison between *P. domestica* ‘Hanita’ (hexaploid) and *P. armeniaca* ‘Pannónia’ (diploid) showed a stronger signal in the anther of the sampled diploid cultivar (Fig. 5), but the effect of ploidy was not tested.

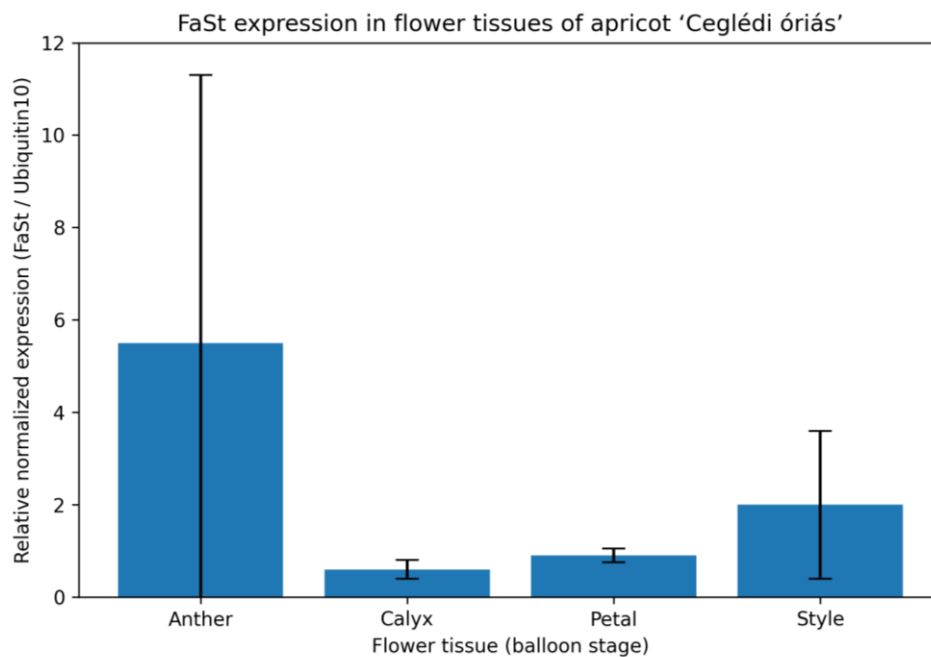


Figure 4. Relative expression of the *Falling Stone* (*FaSt*) transposon in four floral tissues of apricot (*Prunus armeniaca* L.) cultivar ‘Ceglédi óriás’ at the balloon stage. Expression levels were quantified in anthers, calyx, petals, and styles and normalized to the Ubiquitin10 reference gene. Bars represent relative normalized expression values, and error bars indicate variability among replicates. *FaSt* expression was highest in anthers, while lower expression levels were observed in calyx, petals, and styles.

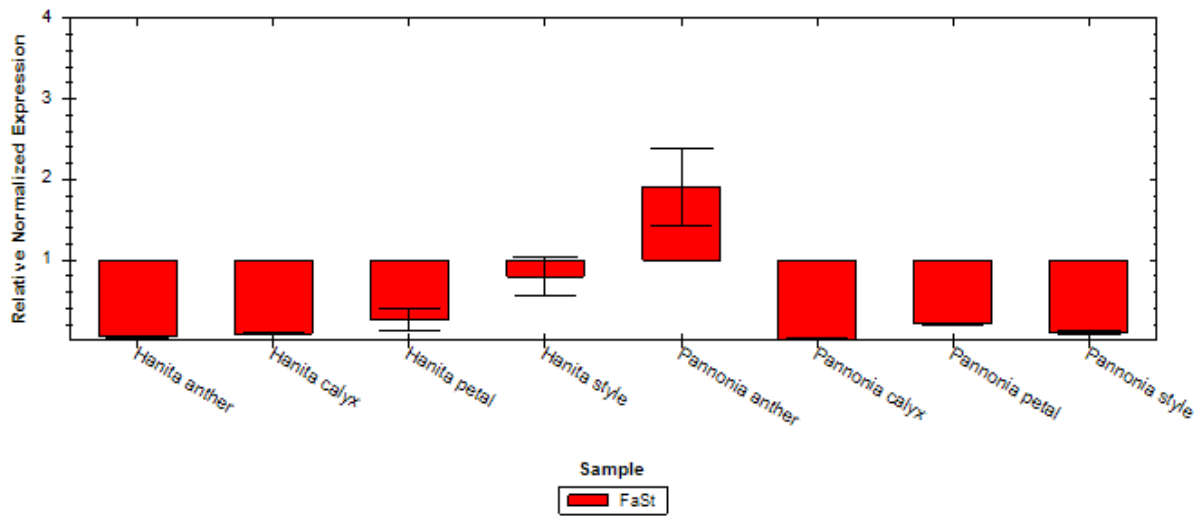


Figure 5. Relative expression of the *Falling Stone* (*FaSt*) transposon across four floral tissues in two *Prunus* species: apricot (*Prunus armeniaca* L.) cultivar ‘Pannónia’ and European plum (*Prunus domestica* L.) cultivar ‘Hanita’. Expression levels were measured in anthers, calyx, petals, and styles and normalized to a reference gene. Bars represent relative normalized expression values, and error bars indicate variability among replicates.

5.3 *In silico* analysis of the occurrence of *Falling Stones* (*FaSt*) transposable elements in *Prunus* genome and transcript datasets

We searched available *Prunus* genome assemblies and associated RNA-Seq datasets in the Genome Database for Rosaceae (GDR) to survey the genomic and transcriptomic occurrences of the *FaSt* MITE family. Genome assemblies and associated RNA-Seq libraries interrogated in this study included *P. armeniaca* (‘Sungold’, ‘Longwangmao’, ‘Marouch’, ‘Stella’ v1.0), *P. mume* v1.0, *P. mandshurica* v1.0, *P. sibirica* CH320_5 and F106 v1.0, peach v2.0.a1, *P. dulcis* v2.0, and *P. salicina* ‘Sanyueli’ v2.0. Because raw counts of element copies are affected by assembly size and completeness, cross-species comparisons below are based on copy numbers normalized to assembly length.

Transcriptomic datasets linked through GDR (all publicly available tissue datasets for the species examined) were searched to detect *FaSt* sequences in RNA-Seq data. The available RNA-Seq resources included, for example, fruit, leaf, and phloem for *P. dulcis* v2.0. Using strict sequence matching criteria, we detected complete *FaSt* sequences in a small number of mRNA assembled transcript sequences: 9 in *P. armeniaca*, 11 in *P. dulcis*, and 11 in *P. persica*. Of these 31 complete transcript-embedded copies, 28 were located in annotated 3’ UTR regions and 2 in 5’ UTR regions; a single case involved the 5’ end of the *FaSt* sequence contributing a short 9-

amino-acid fragment that matched the C-terminal motif of a cytochrome P450 sequence, with the remaining sequence situated in the 3' UTR. We report this unusual structural arrangement as an observation that may warrant further functional and evolutionary study, but the current data do not provide experimental evidence for functional consequences.

When we inspected genomic localization patterns, chromosomes 1 and 6 tended to carry higher numbers of *FaSt* copies in the majority of the assemblies examined, whereas chromosome 5 was often among the chromosomes with fewer copies (Fig. 6). This statement summarizes a descriptive tendency observed across the examined assemblies. The precise ranking of chromosomes by copy number varied among species. The distribution of the 575 partial *FaSt* fragments across the eight chromosomes was examined separately. These fragments show a broadly similar trend but are reported independently from complete element counts because they may represent degraded or truncated sequences.

Overall, integration of genome and transcriptome searches provides targeted evidence that *FaSt* elements are present as complete and partial copies in multiple *Prunus* assemblies and that *FaSt* sequences can be detected within limited sets of mRNAs in *P. armeniaca*, *P. dulcis*, and *P. persica*. However, the data presented here are primarily descriptive; functional impacts and hypotheses about roles in trait evolution require further experimental validation and appropriately powered statistical tests.

The pattern of *FaSt* accumulation in *Prunus* genes indicates the differences in AT frequencies and functional constraints at different parts of protein-coding genes and provides an opportunity to explore the genetic background of phenotypic traits.

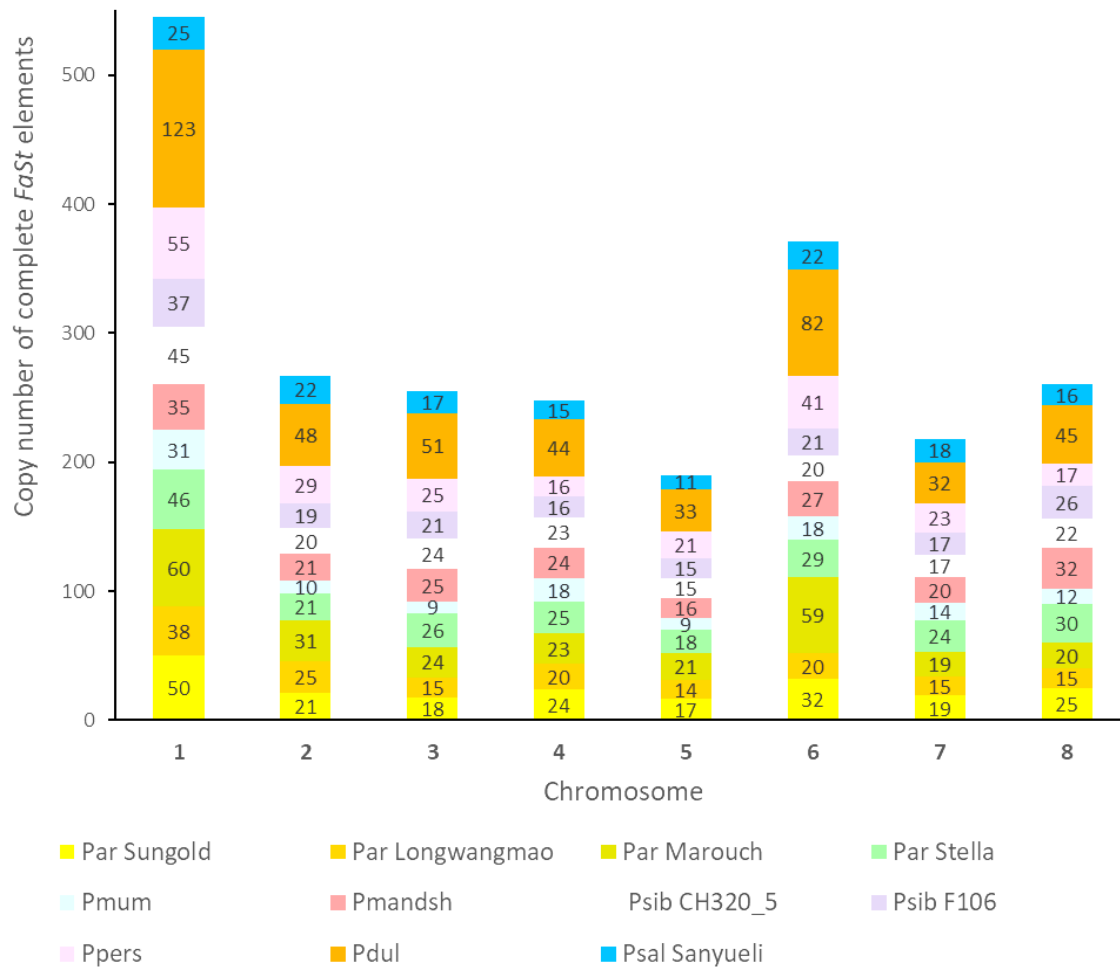


Figure 6. Copy number of the *FaSt* elements distributed on the 8 chromosomes (axis x) in 11 genome sequences of 7 *Prunus* species. Par, *P. armeniaca*; Pmum, *P. mume*, Pmands, *P. mandshuriaca*; Psib, *P. sibirica*; Ppers, *P. persica*; Pdul, *P. dulcis*; and Psal, *P. salicina*.

5.4 Insertion of a *FaSt* in the promoter region of an almond *C-repeat-binding factor (CBF)* gene

During screening of the NCBI nucleotide database, BLASTN analysis identified a putative *FaSt* transposable element in the upstream regulatory region of the *Prunus dulcis* *PdCBF1* gene (GenBank accession JQ317157). No *FaSt* insertion was detected in the corresponding promoter region of *PdCBF2* (JQ317158) (Fig. 7). We used bioinformatics approaches to test if the presence of *FaSt* might be associated with the previously detected alterations in gene expression patterns of *CBF* genes (Ivanovska *et al.*, 2022). Comparison of the upstream sequences revealed that the *PdCBF1* promoter is approximately 374 bp longer than that of *PdCBF2*. This length difference is attributable to the insertion of a 374-bp *FaSt* element, resulting in a *PdCBF1* promoter region of approximately 1.1 kb, compared with approximately

0.75 kb for *PdCBF2*. The *FaSt* insertion in *PdCBF1* is located upstream of the transcription start site and is oriented in the reverse direction relative to the reference *FaSt* element originally described in the *Prunus armeniaca* SFB gene (Halász *et al.*, 2014). BLASTN alignment of the 349-bp insertion identified in the *PdCBF1* promoter against the reference *FaSt* element described in the apricot SFB gene (Halász *et al.*, 2014), has revealed 84.2% nucleotide identity across the full length of the element, with sequence variation consisting of single nucleotide polymorphisms and small insertions/deletions. The characteristics of the 82-bp terminal inverted repeat arms and 9-bp target site duplications (TATTTT) confirmed this insertion being a *FaSt* transposon. Interestingly, such an insertion was not observed in the promoter region of the *PdCBF2* gene. Formerly, TEs inserted in the promoter region were shown to efficiently modify the expression patterns of the corresponding genes (Wang *et al.*, 2016).

Motifs for the SQUAMOSA subfamily of MADS-box (SBP) TFs (Fig. 7) have been observed. The SBPs have functional role in the development of flowers (Guo *et al.*, 2008; Klein *et al.*, 1996). Moreover, the *FaSt* elements contained binding motifs for the TFs involved in plant growth, such as the SHI-RELATED SEQUENCE (SRS), as well as WRKY transcription factors, taking part in abiotic and biotic plant stress protecting pathways (Fridborg *et al.*, 1999; Rushton *et al.*, 2010; Zhao *et al.*, 2020). These TF families are known to participate in developmental processes and stress-related regulatory pathways, however, their presence here is based solely on *in silico* prediction. The presence of predicted motifs within the *FaSt* sequence indicates a potential, but unvalidated, cis-regulatory contribution of the insertion. No *FaSt*-related sequence and no corresponding predicted TF binding sites were identified in the *PdCBF2* promoter region.

The *FaSt* insertion in the *PdCBF1* promoter introduces predicted cis-regulatory motifs that may contribute to differences in cold-responsive regulation between *PdCBF1* and *PdCBF2*. However, these are *in silico* predictions only and do not demonstrate causality. Notably, although *PdCBF1* and *Prunus persica* *PpCBF1* are not closely related (Barros and Gonçalves, 2012), phylogenetic distance by itself does not indicate functional divergence, and hence functional validation is required.

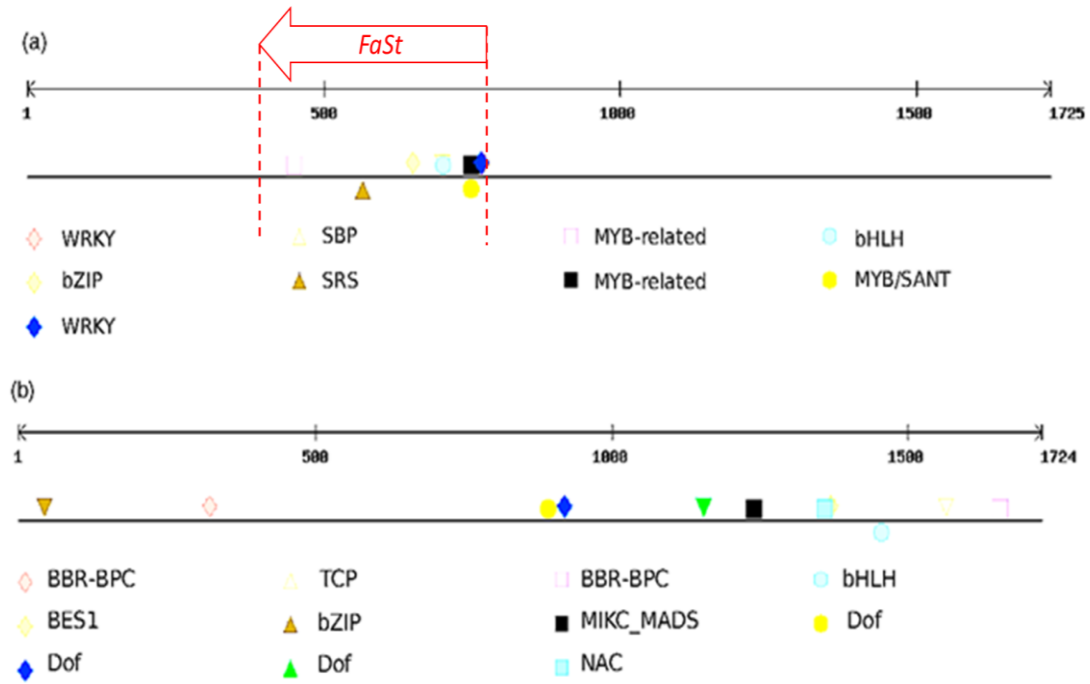


Figure 7. A prediction of some transcription factor binding sites within the promoter region of (a) *PdCBF1* containing a 349-bp insertion of the *FaSt* transposable element (its position and direction is indicated by a red arrow); and (b) *PdCBF2* without a *FaSt* insertion. The figure is created with visualization of the TFs on The Plant Promoter Analysis Navigator (PlantPAN).

Thereby, implying that the aforementioned insertion influenced the diversification of the *CBF1* gene by the introduction of novel characteristics in the genome of *Prunus dulcis*. As *CBF1* is characterized as being essential in the cold acclimation (Barros and Gonçalves, 2012), further functional advancement might be anticipated, considering for instance, that WRKY TFs are involved in cold stress management in plants (Barros and Gonçalves, 2012). The current work is limited to genomic sequence analysis and *in silico* prediction of transcription factor binding sites. No conclusions regarding altered gene expression, regulatory activity, or functional divergence can be drawn from these data alone. Closer studies are needed for deeper understanding of the consequences of the detected TE insertion.

5.5 Results for RNA-Seq analysis and read alignment

To examine the impact of a specific TE (*FaSt*) insertion on gene expression in reproductive tissues in apricots, we generated RNA-Seq profiles for twelve samples representing four experimental groups: “insertion_pollen” (anther tissues of SC apricot, where an *SFB* gene

with a *FaSt* insertion is highly expressed); “insertion_style” (style tissues of SC apricot); “noInsertion_pollen” (anther tissues of SI apricot, where the highly expressed *SFB* gene does not contain a *FaSt* insertion) and “noInsertion_style” (style tissues of SI apricot). Each group was represented in biological triplicate. After gene-level quantification, raw read counts were loaded into DESeq2, with a single “condition” factor encoding all four group combinations (Appendix A3).

After normalization and dispersion estimation, we focused our analysis on the pollen tissue, motivated by the fact, that the non-autonomous *Mutator* MITE element *Falling Stones* (*FaSt*) has been originally discovered in apricot (*P. armeniaca*) and shown to disrupt an *S-haplotype-specific F-box* gene ORF, leading to SC (Vilanova *et al.*, 2006; Halász *et al.*, 2014). To investigate the transcriptional consequences of *FaSt* insertions in their native reproductive context, we focused on pollen and used styles as within-genotype reference samples. RNA was extracted from anthers of three SC cultivars and three SI cultivars, as well as from styles of the corresponding SC and SI genotypes (Tab. 5). By comparing gene expression in tissues expressing *FaSt*-containing versus non-containing pollen *SFB* alleles by using their paired style samples to account for baseline tissue-specific expression, we tried to identify genes whose transcriptional response is uniquely associated with the presence of the 358 bp MITE insertion in pollen.

Table 5. Apricot samples used for RNA-Seq

Genotype common name	Short code (pollen)	Short code (style)	Compatibility status
Gönci magyar kajszi	AMK	SMK	SC
Pannónia	APA	SPA	SC
Rózsakajszi C.1406	ARK	SRK	SC
Ceglédi óriás	ACO	SCO	SI
M604	AM6	SM6	SI
Harmat	AHA	SHA	SI

5.5.1 Read alignment

For read alignment, we used the haploid apricot reference genome *pruArmRojPasHapCUR* (GenBank GCA_903112645.1; *Prunus armeniaca* cv. ‘Rojo Pasion’; Campoy *et al.*, 2020) and mapped reads with STAR v2.6 against this 40× contig-level assembly. Figure 8 shows the number and proportion of uniquely aligned reads per library: mapping rates are consistently high across all twelve samples, and both anther and style libraries display similar proportions of uniquely aligned reads. The absence of systematic differences in alignment

efficiency between SC and SI cultivars indicates that downstream expression differences are unlikely to be driven by alignment artifacts and supports the suitability of this reference for comparative analysis. This uniform mapping performance suggests that observed count differences reflect biological variation and provides confidence for subsequent normalization and differential-expression procedures.

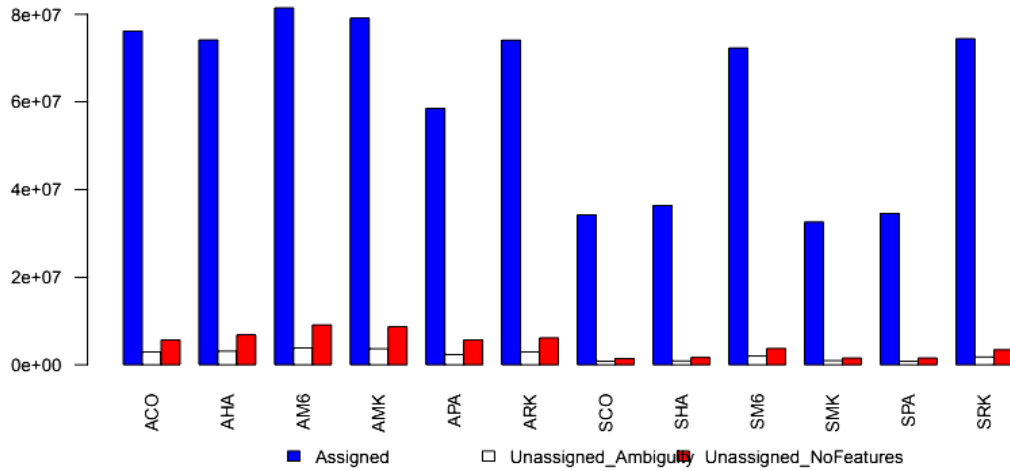


Figure 8. Barplot of the aligned reads on the genome by STAR aligner of the 12 samples apricot cultivars against the genome of ‘Rojo Passion’ apricot. 3 SC anthers (‘Gönci magyar kajszai’ AMK, ‘Pannónia’ APA and ‘Rózsakajszai C. 1406’ ARK) and 3 SI (‘Ceglédi óriás’ ACO, M604 AM6 and ‘Harmat’ AHA) and 3 SC styles (‘Gönci magyar kajszai’ SMK, ‘Pannónia’ SPA, ‘Rózsakajszai C. 1406’ SRK) and 3 SI (‘Ceglédi óriás’ SCO, M604 SM6, ‘Harmat’ SHA). Anthers are shown on the left, and styles on the right.

5.5.2 Scatterplots of transformed counts

Because RNA-Seq data are heteroskedastic (variance tends to increase with mean expression), we compared $\log_2$ -transformations ($\log_2(\text{count} + \text{pseudocount})$) to DESeq2’s regularized logarithm (rlog) transformation and examined the resulting sample relationships. Figure 9 shows scatterplots for the six anther samples where the left panel displays $\log_2$ -scaled counts and the right panel displays rlog-transformed counts; Figure 10 shows the analogous plots for the six style samples. Methodologically, the scatterplots demonstrate that the rlog transformation reduces the exaggerated variance at low counts seen with simple $\log_2$ scaling and yields a more homoskedastic dataset appropriate for PCA and clustering. Biologically, the rlog-transformed scatterplots indicate broadly comparable expression distributions across SC and SI pollen samples (Fig. 9), and largely overlapping profiles across styles (Fig. 10). These results do not suggest a pronounced global transcriptomic shift in either tissue associated with the *FaSt* insertion. Meanwhile, they do not exclude more subtle, locus- or pathway-specific transcriptional effects that would require targeted differential-expression tests to detect.

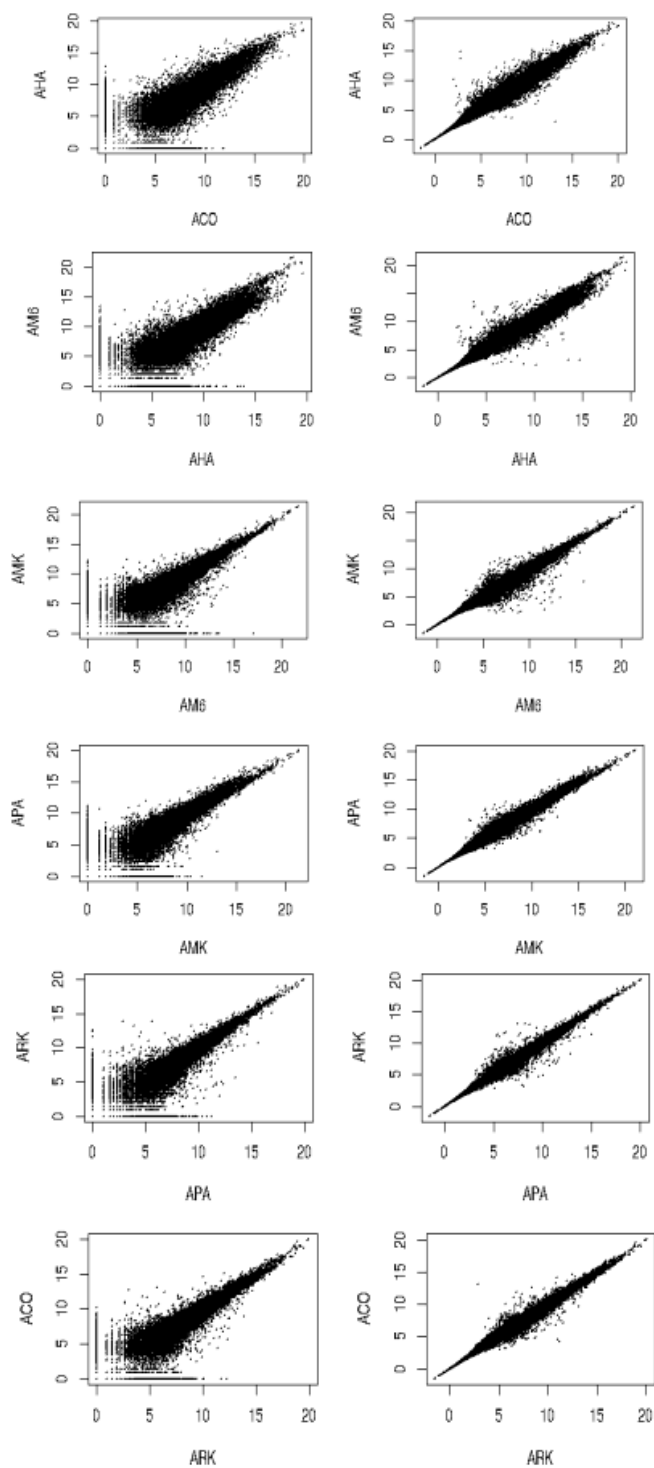


Figure 9. Scatterplots of transformed counts. Scatterplots displaying transformed counts for 6 apricot anther samples. The left panel shows $\log_2$ -transformed scaled counts, while the right panel presents counts processed with the regularized logarithm (rlog) transformation. 3 SC anthers (‘Gönci magyar kajszi’ AMK, ‘Pannónia’ APA and ‘Rózsakajszi C. 1406’ ARK) and 3 SI (‘Ceglédi óriás’ ACO, M604 AM6 and ‘Harmat’ AHA).

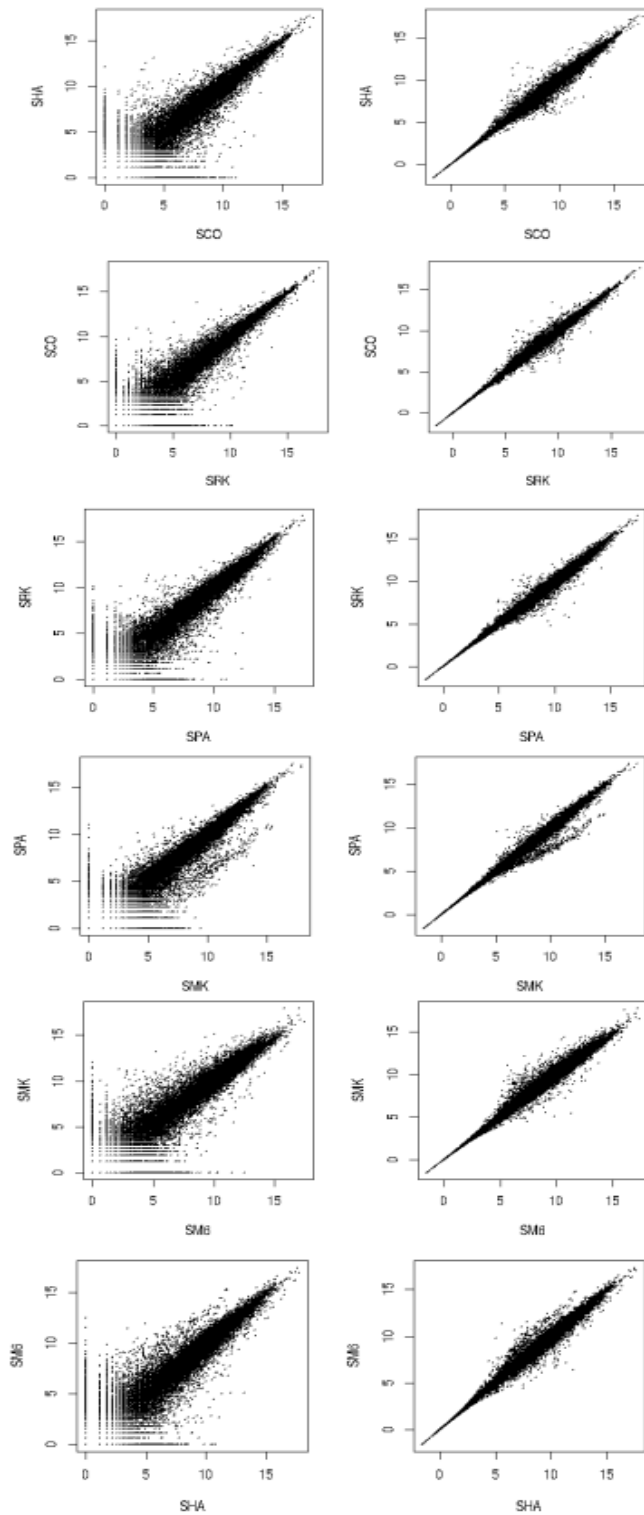


Figure 10. Scatterplots of transformed counts. Scatterplots displaying transformed counts for 6 apricot style samples. The left panel shows $\log_2$ -transformed scaled counts, while the right panel presents counts processed with the regularized logarithm (rlog) transformation. The style samples include three SC cultivars ('Gönci magyar kajszi' SMK, 'Pannónia' SPA, 'Rózsakajszai C. 1406' SRK) and three SI cultivars ('Ceglédi óriás' SCO, M604 SM6, 'Harmat' SHA).

5.5.3 Principal Component Analysis

Principal Component Analysis (PCA) of rlog-transformed counts was used as an exploratory approach to summarize the major axes of transcriptional variance among samples (Fig. 11). The first two principal components (PC1 and PC2) represent the largest sources of variance in the dataset but are not attributed to specific biological processes. Instead, they likely reflect a combination of factors including tissue identity, cultivar background, and overall expression differences. Within each tissue, SC and SI samples largely overlap in PCA space, indicating that phenotype status does not strongly structure the global transcriptome. A small number of cultivar-specific samples show modest deviations along PC1 or PC2. However, these deviations are relatively minor compared to the overarching tissue identity effect, supporting the notion that cultivar background rather than phenotype-specific reprogramming appears to contribute more strongly to the primary source of variation.

5.5.4 Sample Distance Analysis

In order to further examine the relationships between samples, we computed pairwise Poisson distances and visualized these distances using hierarchical clustering heatmaps for both anther (panel A) and style (panel B) libraries (Fig. 12). The heatmaps show that biological replicates (shown along the darkest diagonal cells) have the smallest Poisson distances, confirming high reproducibility across experimental replicates.

In both the anther and style heatmaps, SC and SI samples are interspersed within tissue branches, with no clear separation by phenotype. This observation aligns with the PCA results, where SC and SI samples also did not segregate into distinct clusters, indicating that self-compatibility status does not significantly influence the overall transcriptomic distance between samples. However, certain cultivars, such as ‘Ceglédi óriás’ (ACO) and ‘Harmat’ (AHA) in anthers, show relatively larger pairwise distances, suggesting that cultivar-specific effects may contribute to some divergence in gene expression. These patterns of cultivar-specific variation further support the notion that the presence of the *FaSt* insertion does not appear to induce a widespread, SI-driven transcriptomic reprogramming. Instead, the modest cultivar-specific differences observed are more consistent effects rather than a common program of self-incompatibility-driven changes in gene expression.

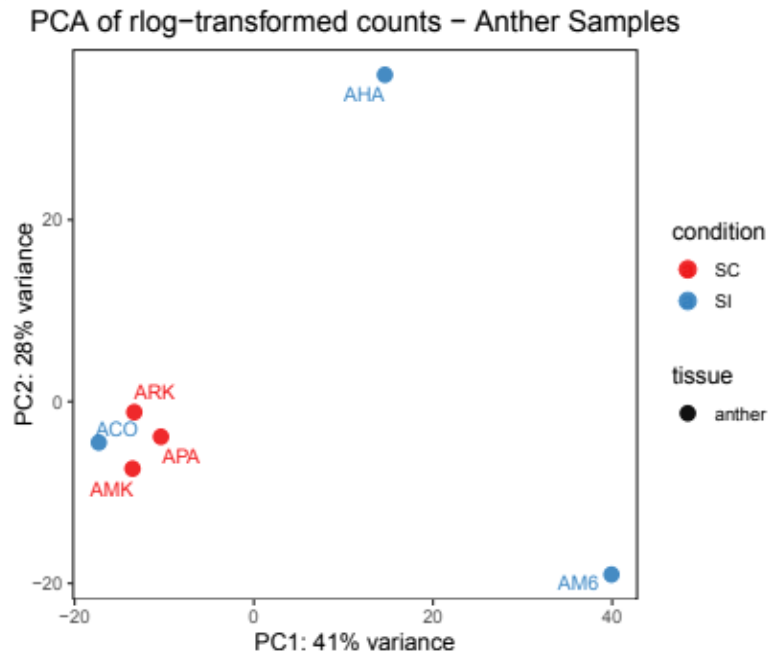
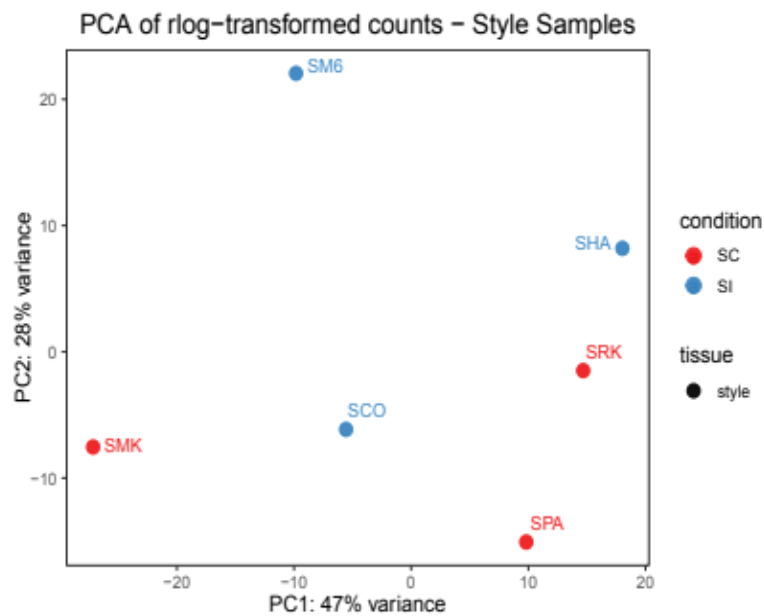
A**B**

Figure 11. Principal component analysis (PCA) of rlog-transformed count values showing clustering of SC and SI cultivars. 3 SC anthers shown in red ('Gönci Magyar Kajszi' AMK, 'Pannónia' APA and 'Rózsakajsz C. 1406' ARK) and 3 SI shown in blue ('Ceglédi óriás' ACO, M604 AM6 and 'Harmat' AHA) and 3 SC styles ('Gönci Magyar Kajszi' SMK, 'Pannónia' SPA, 'Rózsakajsz C. 1406' SRK) and 3 SI ('Ceglédi óriás' SCO, M604 SM6, 'Harmat' SHA). SC cultivars are shown in orange, and SI cultivars in blue. (A) Anther samples; (B) Style samples.

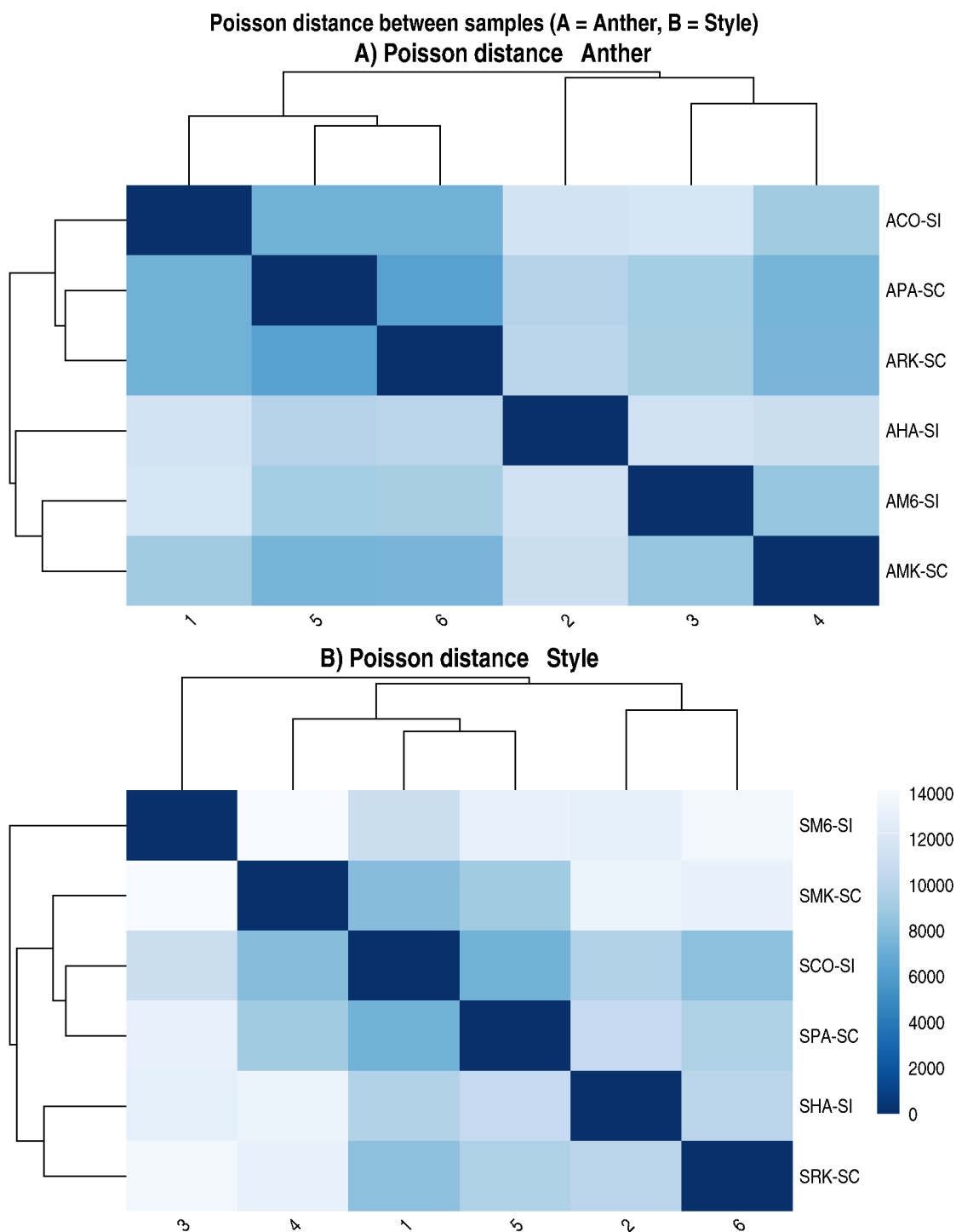


Figure 12. Heatmap of distance between samples with Poisson Distance. A) Anthers of ‘Gönci Magyar Kajszi’ AMK, ‘Pannónia’ APA, ‘Rózsakajsi C. 1406’ ARK, ‘Ceglédi óriás’ ACO, M604 AM6 and ‘Harmat’ AHA; B) Styles of ‘Gönci Magyar Kajszi’ SMK, ‘Pannónia’ SPA, ‘Rózsakajsi C. 1406’ SRK, ‘Ceglédi óriás’ SCO, M604 SM6 and ‘Harmat’ SHA. Anthers are shown on the left, styles on the right. The colors represent color coded expression matrix, where the rows and columns are clustered using hierarchical clustering. Dark blue represents the distance being the smallest and white the furthest.

5.5.5 Independent filtering

After confirming that global patterns are dominated by tissue identity and that *FaSt*-related effects are likely localized, we applied independent filtering prior to differential-expression testing to increase detection power. Lowly expressed genes were binned by average expression and, within each bin, the proportion of genes with p-values below 0.05 was assessed. The independent-filter diagnostic showed the enrichment of low p-values in particular expression bins. Excluding genes with very low counts before multiple-testing correction reduces their disproportionate influence on the false discovery rate and improves the ability to detect truly differentially expressed genes. The independent filtering approach implemented in DESeq2 uses the *genefilter* package and follows the statistical rationale described by Bourgon *et al.* (2010).

5.5.6 Differential expression analysis in anther tissues

Differential expression (DE) analysis was conducted on RNA-Seq data from anther and style tissues to investigate transcriptional differences between self-incompatible (SI) and self-compatible (SC) cultivars. Tissue was not included in the model because analyses were performed separately for each tissue type.

In anther tissues, 28,106 genes exhibited nonzero total read counts across samples. At the stringent significance threshold of $\text{padj} < 0.05$, DESeq2 identified 90 significantly differentially expressed genes (DEGs) between SI and SC cultivars, corresponding to approximately 0.32% of the expressed transcriptome. Among these, 67 genes were upregulated and 23 genes were downregulated in SI relative to SC cultivars.

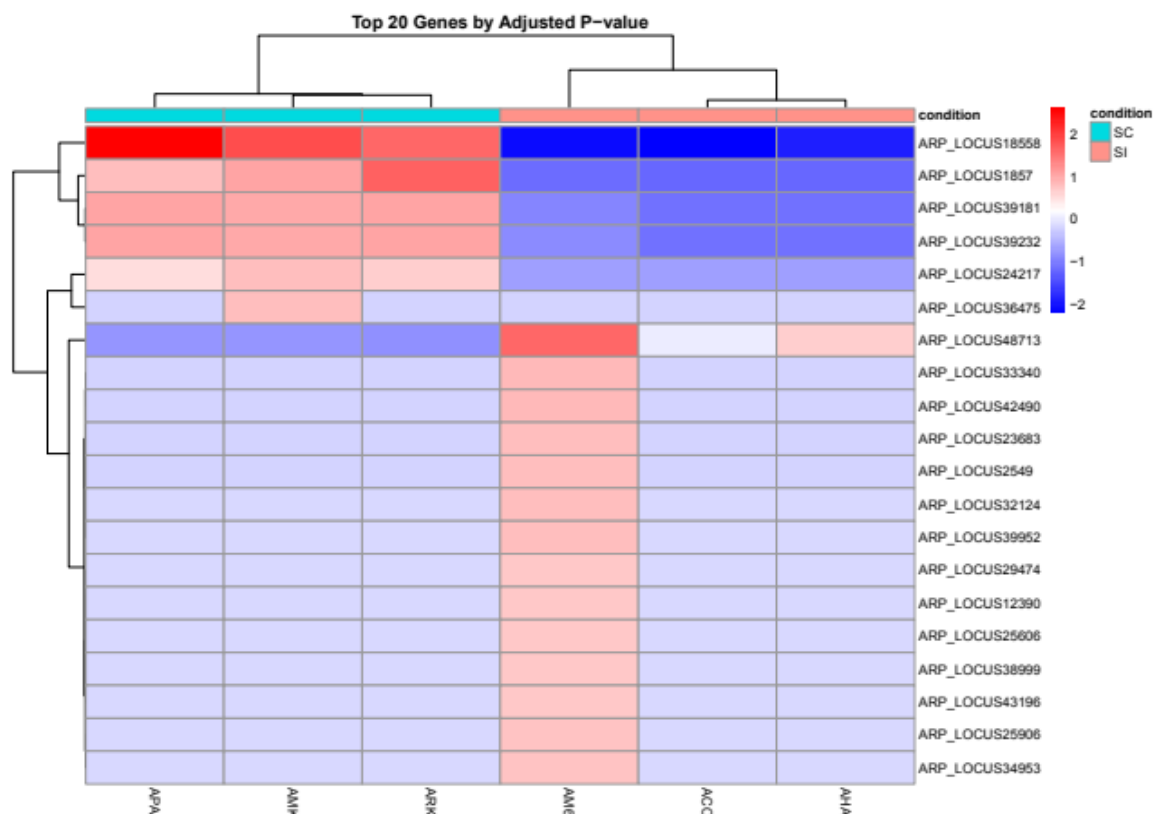
Using a more permissive threshold ($\text{padj} < 0.1$), 137 DEGs were detected, including 95 upregulated and 42 downregulated genes in SI compared to SC. Independent filtering excluded 4,178 low-count genes (mean normalized count < 4), thereby increasing statistical power for multiple testing correction. Additionally, 663 genes were flagged as outliers based on Cook's distance and were not considered significant.

Regularized log (rlog) transformation was applied for visualization and sample-level quality assessment. Distance-based clustering of rlog-transformed expression values revealed that biological replicates clustered primarily according to compatibility phenotype rather than cultivar identity, indicating a consistent transcriptional signal associated with SI versus SC status in anther tissue (Fig. 13).

Among the most strongly downregulated genes were an uncharacterized protein ($\log_2FC = -21.50$, $p_{adj} = 3.62 \times 10^{-5}$), *cell division control protein 24 OB domain-containing protein* ($\log_2FC = -10.77$, $p_{adj} = 2.93 \times 10^{-9}$), and an *amino acid transporter transmembrane domain-containing protein small RNA target site* ($\log_2FC = -9.03$, $p_{adj} = 3.49 \times 10^{-15}$).

The most upregulated genes included Uncharacterized protein, *PGG domain-containing protein*, and Uncharacterized protein LOC110752720, all with a $\log_2FC$ of +21.06 and $p_{adj} = 1.80 \times 10^{-5}$, as well as *non-specific serine/threonine protein kinase* ($\log_2FC = +20.89$, $p_{adj} = 3.62 \times 10^{-5}$) (Fig. 13). To assess sample similarity and reproducibility, $rlog$ -transformed counts were used to compute pairwise sample distances and hierarchical clustering. In anther tissue, biological replicates consistently clustered together and separated according to reproductive condition, confirming high reproducibility and supporting the robustness of the differential expression analysis.

Overall, the anther transcriptome exhibited a modest number of differentially expressed genes between SI and SC cultivars. While the number of DEGs was limited, several genes showed large expression differences, suggesting potential roles in reproductive phenotype differentiation. Further functional studies will be required to elucidate the biological significance of these genes, particularly those with unknown molecular functions.



	Locus	Gene name
ARP_LOCUS18558	CAL8132126.1	Amino acid transporter transmembrane domain-containing protein
ARP_LOCUS1857	CAL8079863.1	Cell division control protein 24 OB domain-containing protein

Figure 13. Heatmap of the top 20 genes selected by adjusted p-value (padj), displaying relative rlog-transformed expression values across anther tissue samples. Anthers of ‘Gönci Magyar Kajszi’ AMK, ‘Pannónia’ APA, ‘Rózsakajszi C. 1406’ ARK, ‘Ceglédi óriás’ ACO, M604 AM6 and ‘Harmat’ AHA.

Similarly, the differential expression analysis was performed on RNA-seq data obtained from style tissues, comparing SI and SC conditions. Raw counts were filtered to remove low-count genes, retaining 29,914 genes with nonzero read counts for analysis. DESeq2 (Love *et al.*, 2014) was applied using the model formula $\sim$ condition, where “condition” denotes SI or SC. Tissue was not included in the model because analyses were performed separately for style and anther samples (Fig. 14). Size-factor normalization and dispersion estimation were carried out using default DESeq2 settings, and statistical significance was assessed using Wald tests with Benjamini–Hochberg correction for multiple testing.

At an adjusted p-value threshold of 0.05, 75 genes were identified as significantly differentially expressed between SI and SC styles, corresponding to approximately 0.25% of the expressed transcriptome. When applying a more permissive, exploratory threshold of $\text{padj} < 0.1$, a total of 119 differentially expressed genes (DEGs) were detected. Of these, 31 genes were upregulated and 88 genes were downregulated in SI relative to SC cultivars. Independent filtering excluded 6,960 low-count genes (mean normalized count < 4), thereby improving statistical power for multiple testing correction. In addition, 203 genes (0.68%) were flagged as outliers based on Cook's distance and were not considered significant. Although 1,477 genes showed nominal significance ($\text{p-value} < 0.05$), only the subsets reported above remained significant after multiple testing correction.

The most strongly upregulated gene was *GDSL esterase/lipase* ($\log_2\text{FC} = +10.13$, $\text{padj} = 7.65 \times 10^{-4}$), while the most strongly downregulated gene was *CCHC-type domain-containing protein* ($\log_2\text{FC} = -10.57$, $\text{padj} = 1.58 \times 10^{-12}$). rlog-transformed counts were used to compute pairwise sample distances and hierarchical clustering. Several additional genes exhibited large effect sizes but did not pass the most stringent significance threshold due to variability or limited expression levels.(Fig. 14).

Despite the presence of differentially expressed genes, no consistent expression pattern could be directly attributed to the presence or absence of the *FaSt*-transposable element insertion. Thus, while compatibility status is associated with measurable transcriptional variation in style tissue, the data do not support a direct causal effect of the *FaSt* insertion on global gene expression in this tissue. Further studies are needed to determine the biochemical functions of proteins encoded by uncharacterized DEGs.

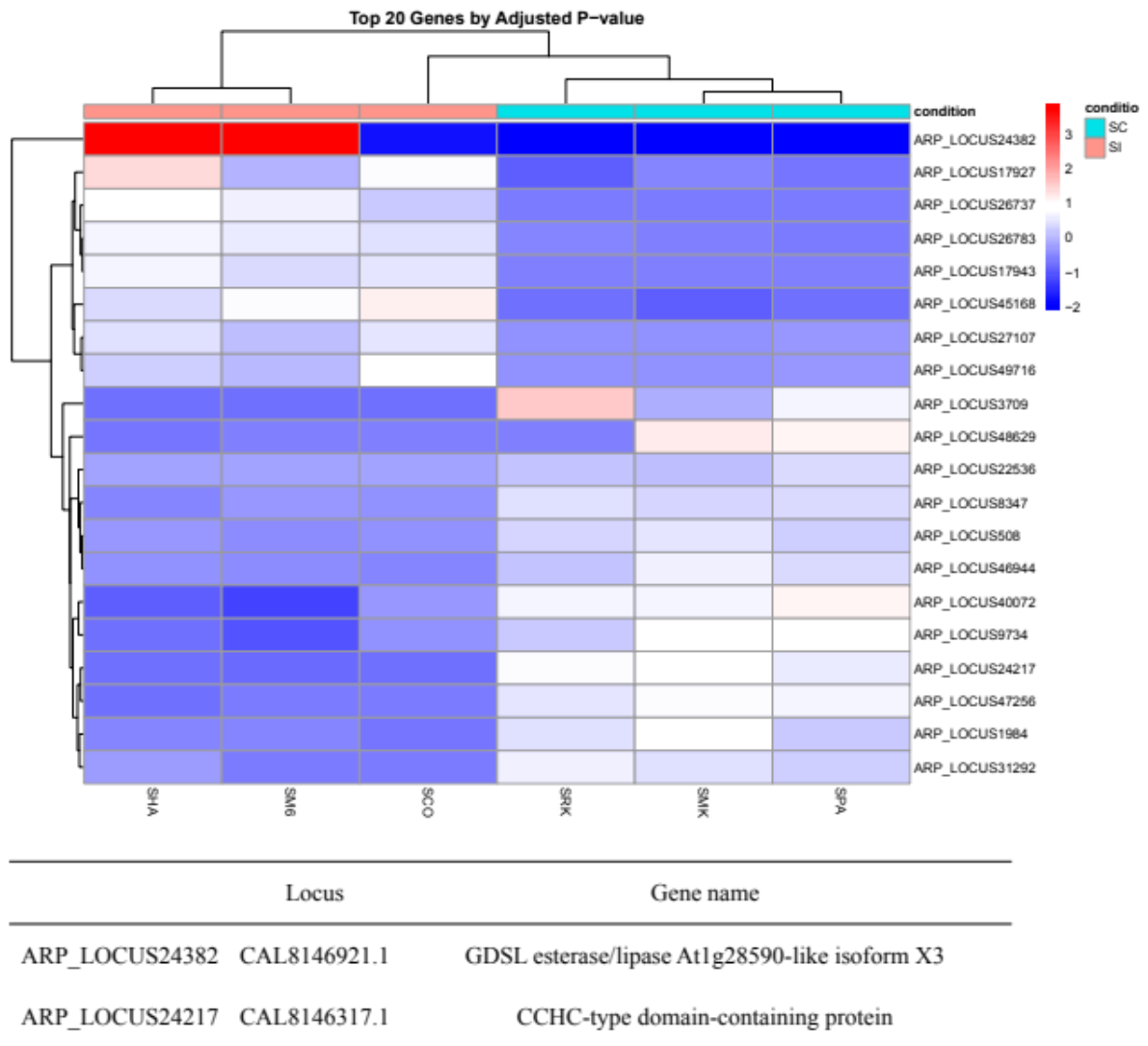


Figure 14. Heatmap of the top 20 genes selected by adjusted p-value (p_{adj}), displaying relative log-transformed expression values across style tissue samples. Styles of ‘Gönci magyar kajszai’ SMK, ‘Pannónia’ SPA, ‘Rózsakajszai C. 1406’ SRK, ‘Ceglédi óriás’ SCO, M604 SM6 and ‘Harmat’ SHA.

5.5.7 Discussion on the RNA-Seq analysis of apricot samples

RNA sequencing of apricot (*Prunus armeniaca*) floral tissues revealed detectable transcriptional differences between SC and SI genotypes and might provide a focused molecular baseline for studies of self-incompatibility. Although RNA-Seq has been applied widely in *Prunus* for flower color traits (Zhang *et al.*, 2024), stress responses (Karimi *et al.*, 2016;) and flowering (Vimont *et al.*, 2019; Villar *et al.*, 2020; Canton *et al.*, 2021), relatively few studies target anthers, pistils, pistil abortion (Shi *et al.*, 2020) or pollen and pistil interactions. Work in

almond and sweet cherry suggests that pollen and pistil responses associated with SC can involve extensive transcriptional programs. Gómez *et al.* (2019) identified more than 1,000 DEGs during pollen and pistil interactions in almond, implicating hydrolases, LRR receptor kinases and transporters, and Feng *et al.* (2025) reported over eight thousand DEGs in pollination time course experiments in *Prunus avium*, with MAPK, hormone and plant pathogen interaction pathways and a core set of transcription factors implicated in compatibility signalling. Studies of pistil failure and floral development also often report large DEG sets: Shi *et al.* (2020) found 3,882 DEGs associated with pistil abortion in *Prunus mume* enriched for hormone, cell wall and carbohydrate metabolism. Zhang and Wei (2022) identified 163 DEGs linked to pistil abortion in apricot long branch buds, notably *XTH* and *beta glucosidase* genes, while dormancy and floral identity studies in cherry and other *Prunus* species reported hundreds to thousands of DEGs dominated by MADS box and hormone pathways (Zhu *et al.*, 2015; Canton *et al.*, 2021; Vimont *et al.*, 2019; Canton *et al.*, 2022). High resolution spatial and developmental studies (Shang *et al.*, 2022; Villar *et al.*, 2020) further resolve stage and cell type specific regulators, including WRKY, SBP and multiple MADS and other transcription factor families, and identify anther development genes such as AMS, MS188 and MS1 that might coordinate organogenesis and dormancy transitions. Male sterility work in *Prunus sibirica* (Chen *et al.*, 2021) implicates carbohydrate metabolism, proteasome activity, calcium signalling and programmed cell death linked to tapetum dysfunction, and these processes appear to overlap with networks reported in other reproductive transcriptomes.

Against this background, our SI and SC contrasts produced a modest DEGs yield, 90 in anthers (Fig. 13) and 75 in styles (Fig. 14). This outcome might reflect the narrow biological contrast, namely SI versus SC at a single floral stage, and the possibility that the *FaSt* transposon insertion in the *S-locus F-box SFB* gene causes local, allele specific or post transcriptional effects rather than broad steady-state mRNA changes. Functionally, the DEGs we observed belong to gene families repeatedly implicated in reproductive biology. Anther DEGs include a non-specific serine/threonine kinase and a PGG domain protein that are upregulated, and a nucleolin-like gene, an amino acid transporter and a cell division control protein that are downregulated. Style DEGs include a strongly upregulated GDSL esterase lipase and a strongly downregulated CCHC type zinc finger. Kinases, transporters, cell cycle regulators, GDSL family members and CCHC CCHC zinc finger proteins have prior links to pollen development, pollen and pistil recognition, cell wall and lipid metabolism and post transcriptional control in *Prunus* and other angiosperms (Gómez *et al.*, 2019; Shi *et al.*, 2020; Chen *et al.*, 2021; Zhang *et al.*, 2022).

The limited DEG number therefore needs not imply lack of biological importance. Instead it might indicate subtler regulatory modes, such as allele-specific expression, chromatin or epigenetic effects, or small RNA mediated control. Integrated mRNA and miRNA profiling in *Prunus mume* and small RNA studies in almond highlight the contribution of miRNA and phasiRNA networks to floral transitions and stress responses (Zhang *et al.*, 2022; Karimi *et al.*, 2016). In addition, recent *Armeniaca* pangenome work suggests that structural variation including TEs may be widespread and that graph-aware or allele inclusive mapping could recover reads that are underrepresented by a single reference sequence (Blanchard *et al.*, 2025). To strengthen mechanistic inference, follow up work might therefore include allele aware mapping or inclusion of the *FaSt* allele in the reference, targeted genomic validation such as long read sequencing or locus specific PCR to resolve insertion breakpoints, and small RNA sequencing to identify potential transposon or compatibility related regulatory RNAs. Such integrated approaches could help determine whether the *FaSt* insertion contributes to local transcriptional or epigenetic modulation of SI in *Prunus armeniaca*.

In conclusion, these results suggest that SI in apricot involves targeted modulation of signaling, transport and regulatory genes rather than broad transcriptional reprogramming, and they provide a concise set of candidate loci that might be prioritized for mechanistic follow up and breeding applications.

5.6 Development of a new marker system based on *Falling Stones* and SCoT primers

Single SCoT primers can only achieve amplification if two genes are located within an amplifiable distance and in opposite orientation (Fig. 15 A). This is not identical to the mechanism of the FaSt-R primer designed in our study. The 349 bp *FaSt* element is composed of two TIRs of 82 bp each, which had to be avoided in primer design, otherwise primers would amplify only a single band from the many *FaSt* copies in the *Prunus* genome. Therefore, the region between nucleotide positions 212 to 233 was used to design the FaSt-R primer, which is a part of the internal sequence of *FaSt* (Fig. 16). Due to the high copy number in the *Prunus* genome and different orientations of the abundant *FaSt* elements, the 22 nt FaSt-R primer can also be used for a single primer PCR assay (Fig. 15 B) (Ivanovska *et al.*, 2025).

5.6.1 Amplification by the SCoT and FaSt-R primers and their combinations

To evaluate the efficiency of SCoT and FaSt-R primers and their combinations, we used two cultivars of six *Prunus* species (peach, almond, apricot, sweet and sour cherries and European plum) and ran the marker analysis with 19 SCoT primers using a single primer for the amplification. We also used the FaSt-R primer separately and then we combined FaSt-R with each of the 19 SCoT primers to check the amplification efficiency. For many primer combinations, the quality of patterns was improved due to the inclusion of the FaSt-R primer (Fig. 17). In some cases, most of the fragments were monomorphic (e.g., SCoT 23, Figure 17A), and the inclusion of FaSt-R resulted in higher PPB (Fig. 17C). Some SCoT primers amplified many weak and diffuse bands (e.g., SCoT 27, Fig. 17B), while the combination of SCoT-FaSt produced fewer and more intense, sharp bands easy to evaluate (Fig. 17D).

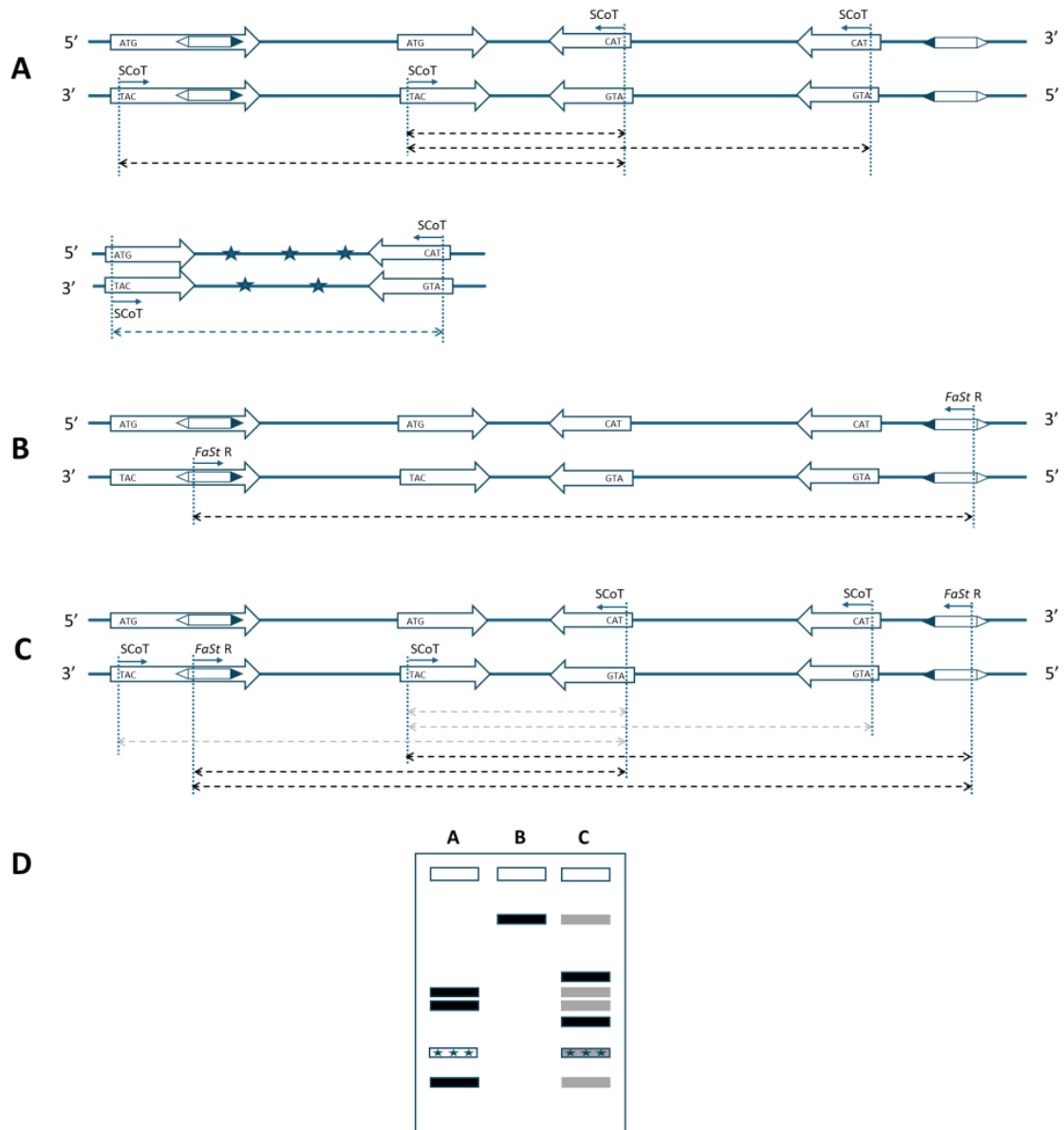


Figure 15. Schematic representation of a marker system based on start codon targeted (SCoT) and *Falling Stones* (FaSt)-specific primers for *Prunus* species. (A) Amplification in the original SCoT assay requires two genes in opposite orientation within an amplifiable distance. (B) A single FaSt-R primer can also yield amplicons under similar conditions, due to the high copy number and varied orientation of FaSt elements. (C) The combination of SCoT and FaSt-R primers can amplify fragments if a FaSt element is located in the proper orientation near a gene. As FaSt-R is sequence-specific, SCoT–FaSt combinations are expected to outperform the consensus SCoT-only assays and also amplify more reliably than primers annealing to the distantly located FaSt elements (weak amplification indicated by grey arrows). Thick arrows: genes in 5' to 3' orientation; hexagons: FaSt elements with black triangles marking the 5' end; thin arrows: primer annealing sites; dashed arrows: expected amplicons; asterisks: non-target DNA. (D) Representative banding patterns from the described PCR strategies. Grey bands indicate weak or undetectable amplification (Ivanovska *et al.*, 2025).

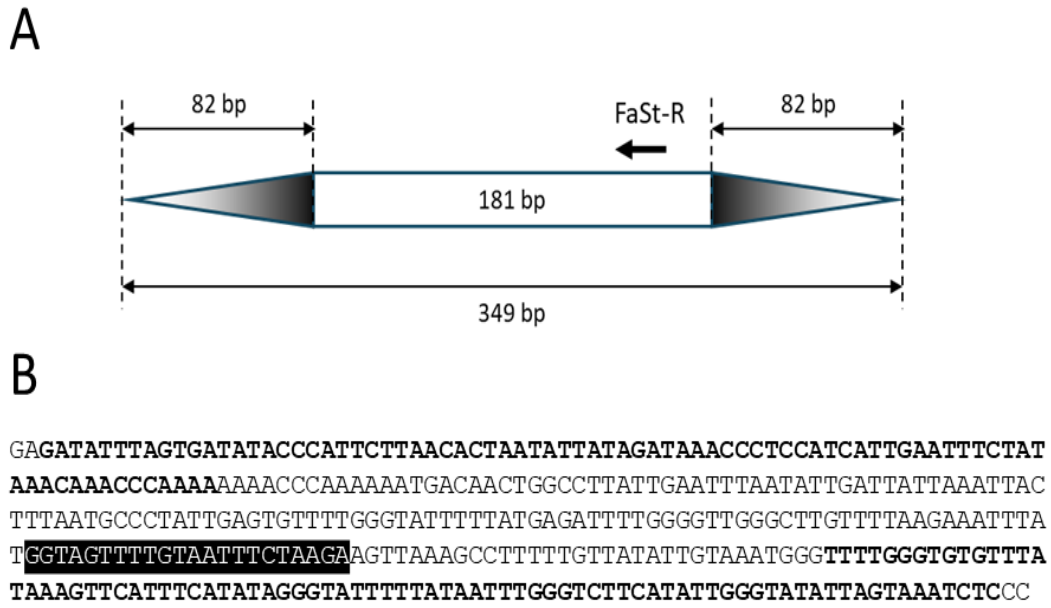


Figure 16. The structure and sequence of *Prunus armeniaca* Falling Stone (FaSt) transposon (KF956794). (A) The structural motifs of the non-autonomous transposon are shown to scale and their sizes are indicated in bp. The terminal inverted repeat (TIR) arms are indicated by gradient-colored triangles and the arrow represent the annealing site and 5' to 3' orientation of the FaSt-R primer. (B) The base sequence of the FaSt element with TIRs in bold and the FaSt-R primer annealing site is shown against black background (Ivanovska *et al.*, 2025).

Using only the FaSt-R primer resulted in the smallest TNB, NDB, PPB, *R_p*, *h* and *I* values (Tab. 6). The smallest fragment size was 700 bp, which is much higher than 250 bp and 150 bp for the SCoT only and SCoT-FaSt combinations, respectively. The averages of TNB, percentage of amplicons with size $\leq 1,000$ bp and *R_p* were the highest in case of SCoT-only analyses. Compared to SCoT-only analyses, the average TNB was lower, and the PPB was considerably higher for the SCoT-FaSt combinations, while NBD, PIC, *h* and *I* were similar for SCoT and SCoT-FaSt analyses.

The TNB was higher for 14 of the 19 SCoT primers than for the corresponding SCoT-FaSt combinations, which can improve the evaluation of gels. In case of fewer bands, the intensity and sharpness of bands were improved as in case of SCoT 27-FaSt-R (Fig. 17 D). Although the average value of the percentage of amplicons with size $\leq 1,000$ bp was higher for SCoT markers than for SCoT-FaSt combinations, two (SCoT 7 and SCoT 11) of the 19 SCoT primers produced small amplicons only in combination with the FaSt-R primer, and for six other primers (SCoT 2, 13, 19, 21, 22 and 34) the presence of FaSt-R significantly increased the number of small amplicons. When combined with FaSt-R, the PPB was unchanged at 100% for

two SCoT primers and it was increased for all but two SCoT primers (SCoT 13 and 24). The inclusion of FaSt-R primer decreased the PIC for 12 SCoT primers, although the magnitude of change was mild ranging from 3 to 25%. The rest of the SCoT primers showed a similar rate of increase in PIC values. The *Rp*, *h* and *I* were favorably changed in 7 of the 19 SCoT primers when combined with FaSt-R. In some cases, the inclusion of FaSt-R into the PCR resulted in a more pronounced improvement, e.g., *Rp* was increased for SCoT 11 by 339%, and *h* was increased by 50% for SCoT 2.

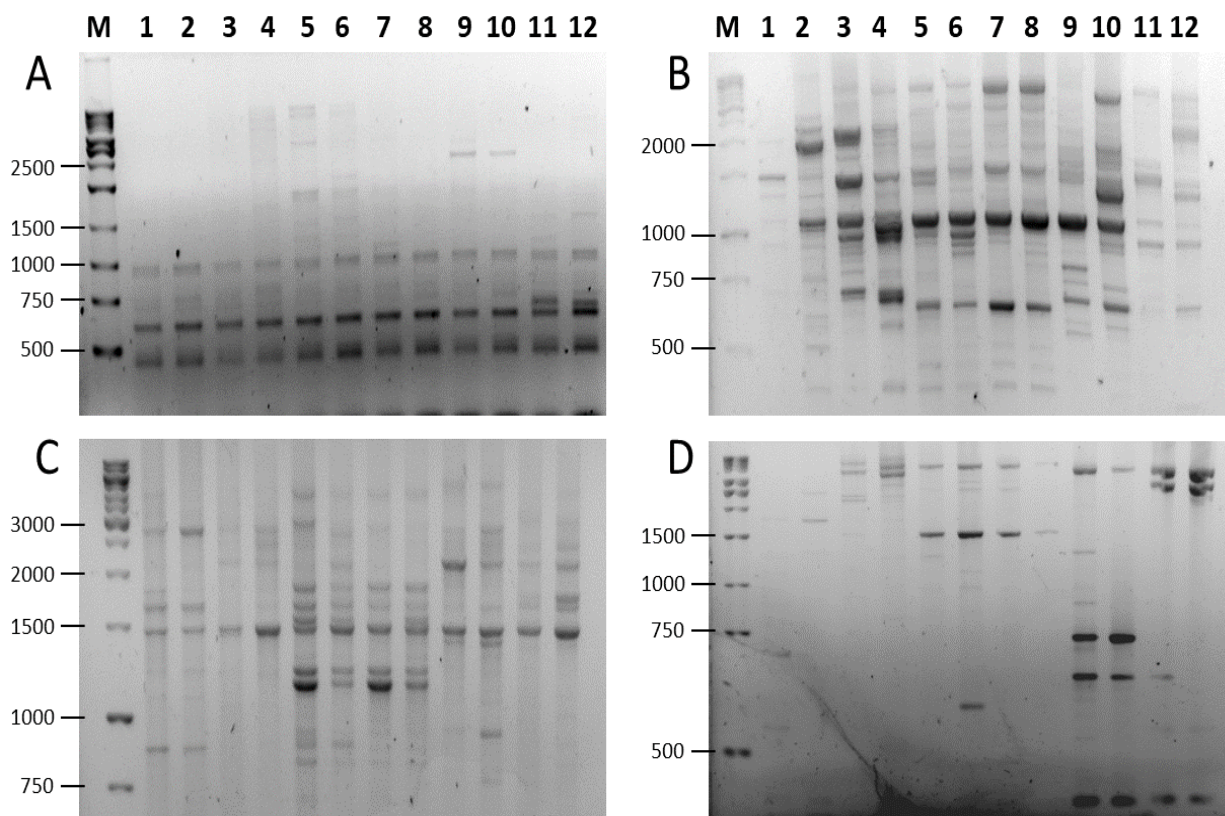


Figure 17. Patterns of amplified fragments in PCR of six *Prunus* species using a SCoT and a combination of SCoT (forward) and FaSt-R (reverse) primers. (A) SCoT23. (B) SCoT23 and FaSt-R. (C) Scot27. (D) SCoT27 and FaSt-R. Labels are the followings: (M) GeneRuler 1 kb DNA ladder, apricot (1) ‘Ceglédi óriás’ and (2) ‘Gönci magyar kajszi’; almond (3) ‘Tuono’ and (4) ‘Tétényi keményhéjú’; sour cherry (5) ‘Kántorjánosi 3’ and (6) ‘Feketici meggy’; sweet cherry (7) ‘Canada giant’ and (8) ‘Katalin’; peach (9) ‘Collins’ and (10) ‘Condor’; European plum (11) ‘Toptaste’ and (12) ‘Haroma’ (Ivanovska *et al.*, 2025).

Table 6. Statistical values of the 19 SCoT primers, the FaSt-R and their combinations in 12 cultivars across six species.

	TNB*	NDB*	SRA*	PA≤1kb*	PPB*	PIC*	Rp*	h*	I*
SCoT 1	73	11	350-2,200	43.8	83.6	0.37	6.17	0.35	0.51
SCoT 2	88	14	650-3,000	6.8	59.1	0.28	6.00	0.24	0.37
SCoT 4	133	23	250-4,000	59.4	72.9	0.35	12.50	0.28	0.43
SCoT 5	86	14	500-3,000	47.7	86.0	0.37	8.33	0.33	0.50
SCoT 6	81	12	350-3,000	55.6	85.2	0.33	5.83	0.30	0.45
SCoT 7	35	8	1,100-3,500	0.0	100.0	0.37	4.50	0.29	0.46
SCoT 11	51	7	2,000-4,500	0.0	52.9	0.33	3.83	0.24	0.37
SCoT 12	70	11	350-3,000	60.0	100.0	0.41	8.00	0.35	0.53
SCoT 13	98	16	300-4,500	36.7	100.0	0.38	11.00	0.33	0.50
SCoT 16	76	10	750-2,250	47.6	68.4	0.34	5.33	0.32	0.47
SCoT 18	119	23	300-2,500	67.2	69.7	0.37	12.83	0.28	0.44
SCoT 19	82	14	500-3,000	23.2	85.4	0.36	7.33	0.32	0.48
SCoT 21	78	12	500-2,500	21.8	84.6	0.37	6.67	0.34	0.51
SCoT 22	79	12	800-3,000	27.8	54.4	0.33	6.50	0.25	0.39
SCoT 23	54	10	400-4,000	27.5	33.3	0.23	3.00	0.26	0.32
SCoT 24	72	11	500-2,000	54.2	100.0	0.38	6.67	0.38	0.56
SCoT 27	90	14	300-6,000	64.4	73.3	0.37	7.17	0.29	0.44
SCoT 32	70	12	350-3,000	40.0	82.9	0.34	6.00	0.31	0.46
SCoT 34	63	12	650-4,000	17.5	81.0	0.39	7.83	0.30	0.46
Mean	78.8	12.9		36.9	77.5	0.35	7.13	0.30	0.46
FaSt-R	46	10	700-2,500	34.8	73.9	0.36	5.67	0.25	0.40
SCoT 1 + FaSt-R	73	10	600-4,000	32.9	100.0	0.35	6.17	0.35	0.52
SCoT 2 + FaSt-R	74	10	300-4,000	27.0	83.8	0.35	5.67	0.36	0.53
SCoT 4 + FaSt-R	49	11	450-3,000	36.7	100.0	0.34	5.50	0.29	0.45
SCoT 5 + FaSt-R	77	11	300-8,000	6.5	84.4	0.28	4.50	0.30	0.46

	TNB*	NDB*	SRA*	PA≤1kb*	PPB*	PIC*	R_p*	h*	I*
SCoT 6 + FaSt-R	65	14	150-3,500	47.7	81.5	0.32	6.50	0.26	0.41
SCoT 7 + FaSt-R	30	9	300-4,000	30.0	100.0	0.35	4.67	0.24	0.39
SCoT 11 + FaSt-R	112	21	250-4,000	33.9	78.6	0.38	13.00	0.29	0.45
SCoT 12 + FaSt-R	56	10	450-8,000	46.4	100.0	0.42	6.50	0.34	0.52
SCoT 13 + FaSt-R	84	13	700-5,000	41.7	85.7	0.35	7.00	0.34	0.50
SCoT 16 + FaSt-R	67	16	480-8,000	32.8	100.0	0.38	8.83	0.29	0.45
SCoT 18 + FaSt-R	86	14	500-5,500	11.6	84.8	0.32	6.33	0.30	0.46
SCoT 19 + FaSt-R	91	19	290-6,000	36.3	86.8	0.37	10.50	0.29	0.44
SCoT 21 + FaSt-R	75	13	150-6,000	42.7	100.0	0.35	6.83	0.34	0.51
SCoT 22 + FaSt-R	94	16	750-6,000	34.0	87.2	0.36	8.67	0.32	0.48
SCoT 23 + FaSt-R	70	13	750-4,500	20.0	91.4	0.38	6.33	0.26	0.42
SCoT 24 + FaSt-R	54	10	250-4,000	37.0	55.6	0.31	5.00	0.22	0.35
SCoT 27 + FaSt-R	31	12	180-3,500	45.2	100.0	0.34	4.83	0.22	0.37
SCoT 32 + FaSt-R	50	10	400-6,000	20.0	100.0	0.38	5.33	0.32	0.49
SCoT 34 + FaSt-R	45	16	280-3,000	44.4	100.0	0.30	5.83	0.20	0.34
Mean	67.5	13.1		33.0	90.5	0.35	6.74	0.29	0.45

*: TNB, total number of bands; NDB, number of differently sized bands; SRA, size range of amplicons; PA≤1kb, percentage of amplicons with size ≤ 1,000 bp; PPB, percentage of the polymorphic bands; PIC, polymorphism information content; *R_p*, resolving power; *h*, Nei's (1973) gene diversity; *I*, Shannon's information index.

Data were published in Ivanovska *et al.* (2025).

5.6.2 Phylogenetic analysis of *Prunus* using SCoT-FaSt markers

A phylogenetic tree was constructed based on the data from 19 SCoT-FaSt marker combinations to assess how reliably the genetic relationships among 12 cultivars belonging to 6 *Prunus* species can be shown. Five major clusters were obtained with 95-100% bootstrap support, each corresponding to peach, almond, European plum, apricot, and cherry. Within the cherry cluster, sweet cherries formed a subgroup with strong bootstrap support and ‘Feketicsi meggy’ positioned separately. The peach and almond clusters formed a common group with a high bootstrap value and hence they were closer to each other than to European plums, apricots and cherries. Higher ploidy levels (sour cherries and European plums) did not result in an increased number of detected alleles compared to diploid species.

Since the information in the SCoT-FaSt amplification patterns appeared sufficient to reliably reconstruct evolutionary inferences among *Prunus* species, we also conducted the analysis on a data pool containing SCoT-FaSt amplicons of 28 European plum cultivars (Fig. 18B). The phylogenetic tree consists of five major groups. Fig. 18BI consists of the modern German cultivars and some of those that have been used in their breeding program, Fig. 18BII contained two offspring of the cultivar ‘Stanley’, Fig. 18BIII and Fig. 18BIV had a common origin and included all Hungarian landraces, while Fig. 18BV collected old German, Russian, Swedish and USA cultivars. Another cluster of UK, Russian and Romanian cultivars was not supported statistically.

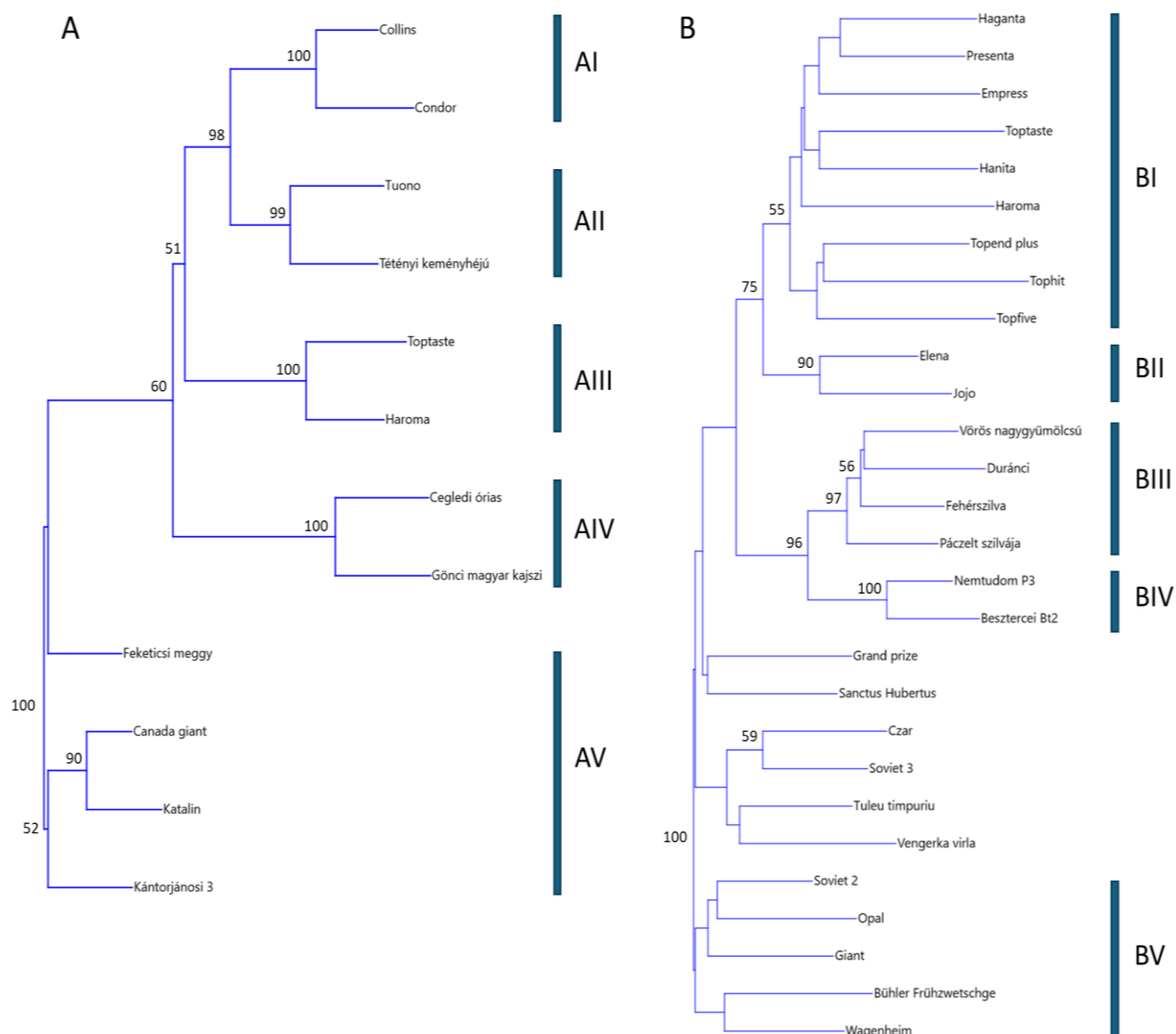


Figure 18. Phylogenetic analysis of *Prunus* using data from 19 SCoT-FaSt primer combinations and the neighbor joining method based on the Jaccard's similarity coefficients. (A) Twelve cultivars of six *Prunus* species (almond, apricot, European plum, peach and sour and sweet cherries). (B) Twenty-eight European plum cultivars. The numbers at specific nodes indicate percentage of 1,000 bootstrap replicates in which a given group was present (< 50% not shown). Numbers AI-AV and BI-BV indicate the statistically supported groups of cultivars (Ivanovska *et al.*, 2025).

The relationships among the assessed European plum cultivars were also analysed by a principal component (PCA) analysis (Fig. 19). The first two principal axes accounted for 23.8 and 14.8 % of the total variation, respectively. The PCA plot shows four distinct groups of

cultivars with ‘Jojo’ and ‘Elena’ located closely to the modern German cultivars. The two Hungarian landraces, ‘Besztercei Bt2’ and ‘Nemtudom P3’ were close to each other and both were strikingly separated along Component 2 axis from the rest of Hungarian landraces.

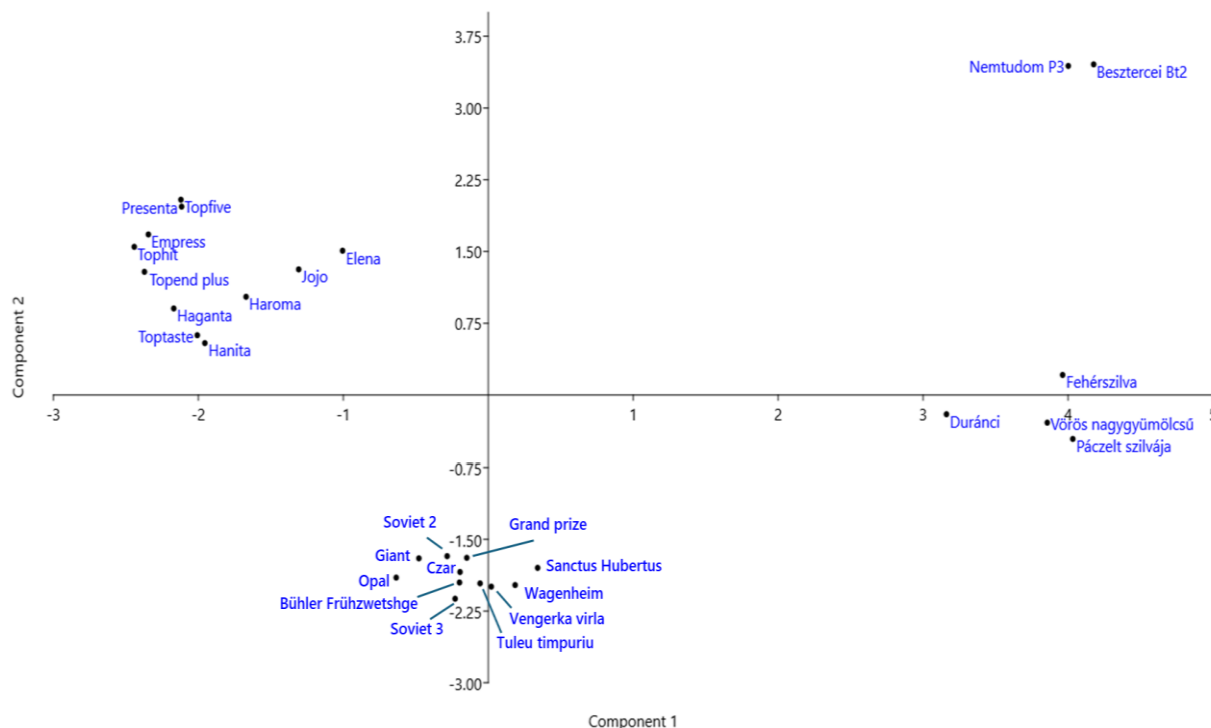


Figure 19. Distribution of *Prunus domestica* cultivars of different origins on the two first principal component analysis axes determined from the SCoT-FaSt marker data (Ivanovska *et al.*, 2025).

5.6.3 Sequence analysis of the amplified fragments

To confirm if the amplified fragments indeed contained recognizable parts of *FaSt* and genic regions, we cloned and sequenced a range of amplicons. A total of 32 sequences were obtained and submitted to the NCBI GenBank database under accession numbers of PQ869779-PQ869809. The obtained sequences were amplified in PCR using *FaSt*-R and the following SCoT primers SCoT1 (2 sequences), SCoT2 (4), SCoT3 (2), SCoT4 (8), SCoT16 (1), SCoT19 (1), SCoT21 (1), SCoT22 (7), SCoT32 (3) and SCoT33 (3). Fragments were cloned and sequenced from the assayed species including *P. domestica* (11 fragments), *P. cerasus* (9), *P. dulcis* (8), *P. persica* (2), *P. avium* (1) and *P. armeniaca* (1).

Twenty-six of the sequences contained an identifiable *FaSt* element with sizes ranging from 58 to 218 bp, and six contained regions where the *FaSt*-R primer annealed to a degenerate

complementary sequence, although it seems not to be a part of a *FaSt* element. In these cases, the annealing sites were located in the 3' UTR in four sequences (Pcer_KJ_2_1227, Pcer_FM_2_1227, Pcer_KJ_4_965 and Pdom_EL_4_766) and two were included in the intron (Pcer_KJ_4_662 and Pdom_TT_21_269). The affected genes include *beta-D-glucosyl crocetin beta-1,6-glucosyltransferase*, *peroxidase N1*, *methyltransferase PMT10*, *deoxyloganetin glucosyltransferase* and *WUSCHEL-related homeobox 9*. A total of 11 of the 26 *FaSt* elements identified were located in the 5' UTR (Pdom_1-TF-907, Pdom_BB_1_908 and Pdul_TU_3_544), 3' UTR (Pdom_NT_22_873 and Pdom_HR_32_1224) or intron (Pdul_TK_4_761, Pdul_TK_22_901, Pcer_KJ_4_1000, Pcer_FM_4_496, Pdom_NT_22_334 and Pdom_HR_33_1460) of specific genes. Those genes showed significant homology (*E* values ranging from 0.0 to 7e-14) to the following protein coding genes: *polygalacturonase*, *inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase*, *BTB/POZ domain-containing protein*, *pectinesterase*, *DEAD-box ATP-dependent RNA helicase 52C*, *ethylene-responsive transcription factor*, *glycosyltransferase*, *beta-D-glucosyl crocetin beta-1,6-glucosyltransferase*, *leaf rust 10 disease-resistance locus receptor-like protein kinase* and a gene coding for an uncharacterized protein (Tab. 7). A *FaSt* element was located in the intergenic region close to the *U1 spliceosomal RNA* coding gene (Pcer_FM_4_1056) and the rest of the *FaSt* elements) were identified in intergenic regions (Tab. 7) (Halász *et al.*, 2019).

Two sequences (Pdom_1-TF-907 and Pdom_BB_1_908) were homologous to the *Prunus persica polygalacturonase ADPG2* gene, while the *FaSt* element in the sequences aligned separately with several *P. persica FaSt* sequences. The Basic Local Alignment Search Tool (BlastN) analysis on the *P. domestica* genome showed 98% identity and two gaps on a 936 bp alignment. The *FaSt* element was located in the 5' UTR, upstream of the start codon of the open reading frame. The same sequence without major alterations was detected in both cultivars, 'Topfive' and 'Besztercei Bt.2' and was also present in the *P. salicina* genome sequence but its *FaSt* element was missing from all other *Prunus* species (Fig. 20).

The complete sequence of Pav_CG_3_910 is present in both the sweet and sour cherry genomes but those of the *P. fruticosa*, apricot, peach and all other *Prunus* species lack its *FaSt* element (Fig. 21). Similarly, the Pcer_KJ_2_343 sequence had a 48% coverage with its closest homolog after BlastN searches; however, blasting it against the sour and sweet cherry genome at the Genome Database of Rosaceae (GDR) platform confirmed the presence of the complete sequence (*E* value 0.0 with 100% identity), while it was absent from *P. fruticosa* and all the assayed *Prunus* species.

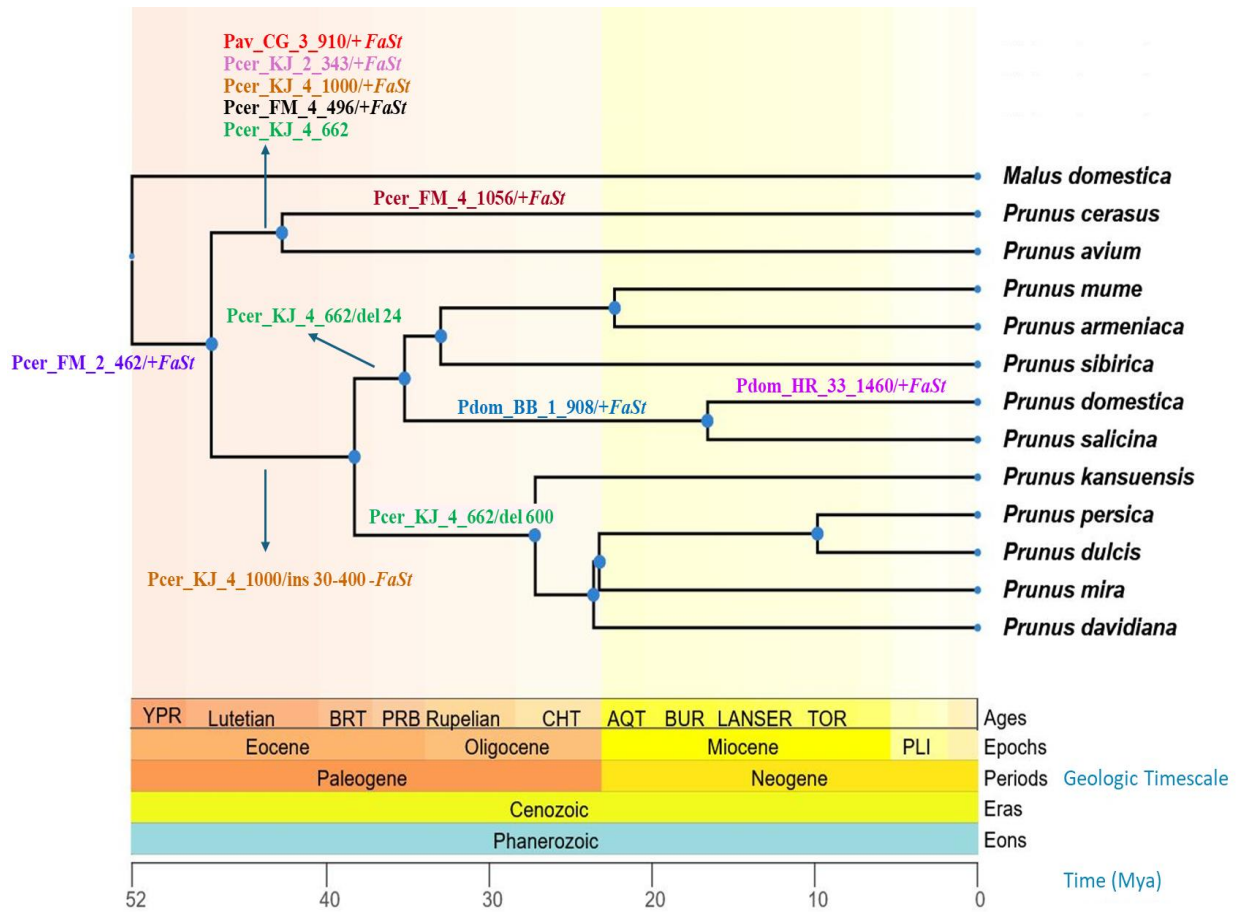


Figure 20. The chronogram of 12 *Prunus* and one *Malus* species derived from divergence time estimation based on the TimeTree 5 database (Kumar *et al.* 2022). The ages of geologic timescale with a scale as millions of years ago (Mya) are shown by background colors and below the cladogram. Arrows point to *Falling Stones* (FaSt) transposon induced sequence alterations (+ presence and – absence of FaSt, while “ins” and “del” refer to other sequence insertions and deletions, respectively). Sequence identifiers are explained in Tab. 7. Data were published in Ivanovska *et al.* (2025).

The Pcer_KJ_4_1000 sequence involves the intron of the gene coding for a *BTB/POZ domain-containing protein* and a *FaSt* copy within the intron region. The sequence was found in *P. cerasus* and *P. avium* genomes but in *P. fruticosa* and *P. yedonensis*, there is a small insertion (~60 bp) in the middle of the intron and outside the *FaSt* element, while in *P. campanula* the size of this insertion is somewhat larger (~200 bp). The genome of *P. salicina*, *P. mongolica*, *P. zhengheensis*, *P. dulcis*, *P. persica*, *P. armeniaca*, and *P. × kanzakura* lacked the *FaSt* element and contained an insertion of 30-400 bp. The Pcer_FM_4_496 sequence covers the intron of a *pectinesterase* gene and a *FaSt* element was present only in the *P. speciosa*, *P. avium* and *P. cerasus* genes, while other more distantly related species (e.g., *P. dulcis* and *P. salicina*) did not

have this *FaSt* copy. The *FaSt* in Pcer_FM_4_1056 was missing from the homologous sequences of *P. avium*, *P. persica* and *P. armeniaca*. Interestingly, the ‘Montmorency’ *snRNA* gene sequence had a 45 bp deletion. The Pcer_FM_2_462 intergenic sequence contained a *FaSt* element, which was present in all assayed *Prunus* species.

Pdul_TK_22_1866 has a *FaSt* in the intergenic region, which is also present in the ‘Nonpareil’ genome (98.89% identity) but is missing from the ‘Lauranne’ and ‘Texas’ homologs. All other *Prunus* species also lack this *FaSt* copy in their homologous regions. The *FaSt* in Pdul_TK_4_761 was also present in the homologous gene of ‘Texas’ but missing from ‘Lauranne’ and all other *Prunus* species. In addition, the Pdul_TK_22_901 has a *FaSt* in the intron 5 of the *DEAD-box ATP-dependent RNA helicase 52C-like* gene, which is missing from each of the genome sequences of ‘Nonpareil’, ‘Lauranne’ and ‘Texas’ cultivars. The *FaSt* in the 5’ UTR of *inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase* gene of ‘Tuono’ (Pdul_TU_3_544) was also carried by the ‘Lauranne’ gene but was not found in the genome sequence of ‘Nonpareil’ and ‘Texas’. The Pdul_TU_33_342 is a unique intergenic sequence with a *FaSt* copy, which could not be detected in any other almond cultivars or *Prunus* species.

The Pdom_HR_33_1460 sequence with its *FaSt* copy is present in the European plum draft genome while the *FaSt* was missing from *P. salicina* and other *Prunus* species. The *FaSt* in Pdom_NT_22_334 intron and the Pdom_EL_4_433 intergenic sequence was not found in the *P. domestica* and other *Prunus* genomes.

Variations were also detected in sequences without *FaSt* elements. Compared to the ‘Montmorency’ genome sequence, the Pcer_KJ_4_965 had a deletion of ~30 bp in the 3’ UTR, while the UTRs of *P. armeniaca* and *P. persica* were more divergent. The Pcer_KJ_4_662 contains an intron in the *peroxidase N1-like* gene and the *FaSt*-R primer annealed to a degenerate region, which is not part of a *FaSt* element. The complete sequence was found in sour and sweet cherries, *P. fruticosa*, *P. yedonensis*, *P. campanulata*, *P. speciosa*, while the homologous sequences in plums, *P. mume*, *P. armeniaca*, *P. humilis*, *P. sibirica* and *P. zhengheensis* had a short (24 bp) deletion, and *P. persica*, *P. dulcis*, *P. kansuensis*, *P. davidiana*, *P. mandshurica*, *P. mongolica* and *P. mira* contained a larger deletion of approx. 600 bp.

To compare the formerly detected dynamics in *FaSt* accumulation in specific genomic regions with a more general genomic landscape of *FaSt* elements, the total number of *FaSt* copies was determined. We calculated the number of the complete (≥ 340 bp) *FaSt* elements, the fragments exceeding 50 % of the original length (339-175 bp), those between 174 and 82 bp, and under 82 bp. Fig. 21 shows an uneven distribution of different sizes over the nine assayed

genomes. *P. persica* and *P. dulcis* had the largest number of full-length *FaSt* copies and the 339-175 bp category was predominant in plum species. The species in the *Cerasus* subgenus had the highest levels of truncated *FaSt* copies (lengths between 174 and 82 bp). *P. fruticosa* (1154) and *P. cerasus* (1594) had more full-length copies than *P. avium* (503) and *P. speciosa* (514). Similarly, *P. domestica* (1400) had considerably more full-length *FaSt*s than *P. salicina* (427). Interestingly, *P. armeniaca* was more similar to *Cerasus* species than those in the *Prunus* subgenus by having more fragmented *FaSt* copies.

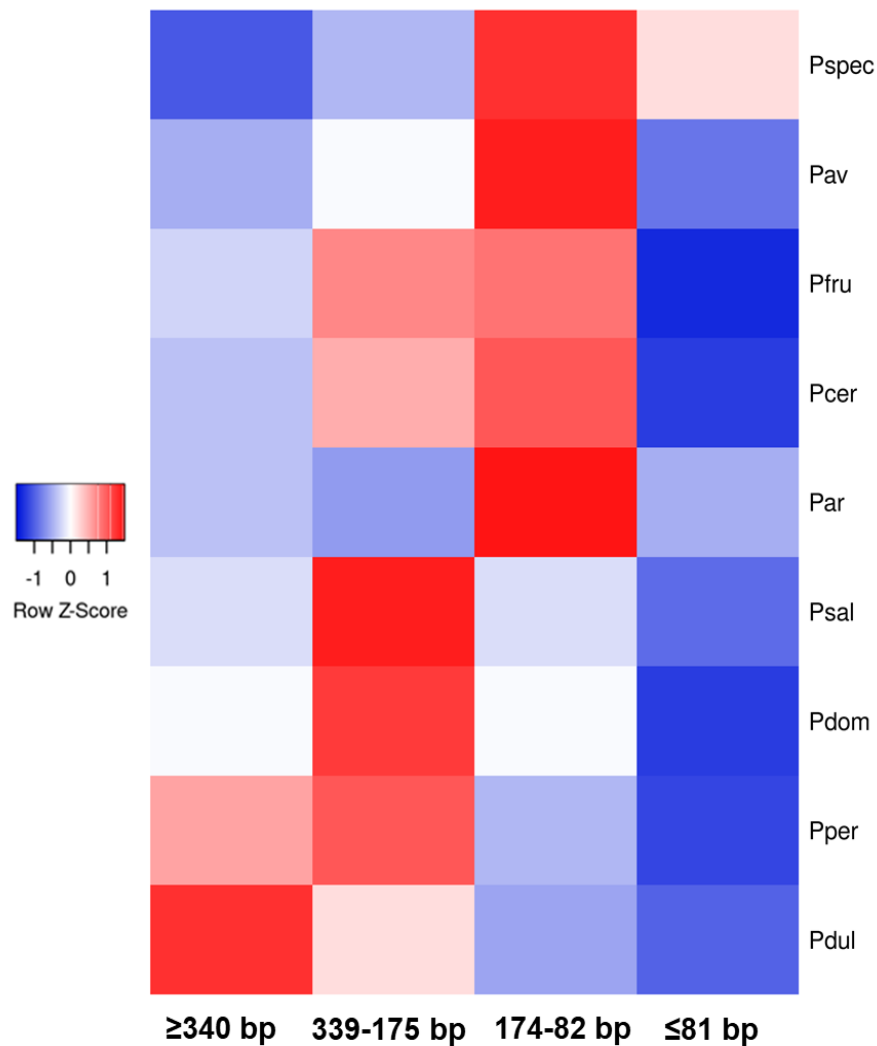


Figure 21. The heat map of *in silico* identified *Falling Stones* (*FaSt*) transposons in nine *Prunus* species. The four size categories are given as follows: almost complete, ≥ 340 bp; 339-175 bp; 174 and 82 bp and ≥ 82 bp, the colors red (high) and blue (low) indicate different genomic abundancies of *FaSt*, the columns represent the four size categories and the rows represent the assayed genomes (Pspec, *P. speciosa*; Pav, *P. avium*; Pfru, *P. fruticosa*; Pcer, *P. cerasus*; Par, *P. armeniaca*; Psal, *P. salicina*; Pdom, *P. domestica*; Pper, *P. persica*; and Pdul, *P. dulcis*). Data were published in Ivanovska *et al.* (2025).

5.6.4 Discussion of the SCoT-FaSt marker system

SCoT markers were designed to anneal to the conserved region flanking the start codon in plant genes (Collard and Mackill 2009) and were used in several plant species including *Prunus* fruit trees (Rai *et al.* 2023, Atapour *et al.* 2022, Buer *et al.* 2022, Guo *et al.* 2018). The NDB and the average for NDB/marker were similar for SCoT only and SCoT-FaSt combinations, and were comparable to values reported for the former SCoT marker analyses of *P. armeniaca* (Guo *et al.* 2018) and *P. sibirica* (Buer *et al.* 2022). The parameters characterizing marker efficiency (the average of PIC and the collective value of R_p) and genetic diversity (h and I) were not reduced by the addition of FaSt-R. In fact, the level of polymorphism was enhanced by SCoT-FaSt primer combinations. SCoT markers produced somewhat a lower percentage of polymorphic bands than in other *Prunus* species in previous assays (Buer *et al.* 2022, Guo *et al.* 2018) but it was considerably increased by the addition of the FaSt-R primer. The PIC values of a dominant marker range between 0 and 0.5, and all the assayed markers and their combinations revealed higher PICs than those reported for SCoT analysis in Chinese apricots (Guo *et al.* 2018).

TNB was lower and NDB was slightly higher for the SCoT-FaSt combinations than SCoT-only analyses. The fewer bands in more size categories covered a wider range of amplicon sizes, between 150 and 8,000 bp. There were smaller and bigger fragments than the exclusively used SCoT primers amplified in many other species, citrus, pomegranates or Jojoba (Mahjbi *et al.* 2015, Heikrujam *et al.* 2015, Karapetsi *et al.* 2021). The inclusion of FaSt-R made gel patterns less dense and allowed more reliable identification of differentially expressed fragments. FaSt-R was designed to anneal within the unique loop sequence of the non-autonomous transposon (Halász *et al.* 2014), allowing the recognition of complete or partial *FaSt* copies within the *Prunus* genome. This can significantly reduce the number of amplicons, since the peach genome contains almost 28,000 protein-coding genes (Verde *et al.*, 2013) and about 1,000 *FaSt* copies. The decrease in TNB and the alterations in band size suggest that FaSt-R primers indeed bring specificity to the PCR and, as predicted, amplification with only SCoT primers is less intense when FaSt-R is also available (Fig. 15D). A SCoT primer may also amplify gene fragments from pathogenic organisms infecting the target plant (Giri *et al.* 2024, Palacioğlu *et al.* 2023). However, by including the FaSt-R that selectively anneals within a *Prunus*-specific transposon, we can get rid of non-target amplicons disturbing the genetic analysis of a crop plant.

The SCoT-FaSt amplified fragments were used for phylogenetic analyses. Five subclusters were constructed on the NJ tree with 99 to 100% bootstrap support. Those included

the peach, almond, plum, apricot and cherry species, with peach and almond as well as sweet and sour cherries being relatively closer to each other. ‘Feketicsi meggy’ was not in a common cluster with ‘Kántorjánosi 3’, which provides support for its putative interspecific origin (Surányi 2016). Such phylogenetic relationships are consistent with the relative differences in divergence time between the corresponding species (Chin *et al.*, 2014; Velasco *et al.*, 2016). It confirms that the SCoT-FaSt marker combinations provide eligible levels of phylogenetic signals to reliably reconstruct evolutionary relationships among species. A more detailed analysis of 28 European plum cultivars also detected five statistically supported clusters. The grouping of cultivars matched their origin and pedigree with German cultivars in Fig. 18BI originated from two different breeding programs, Hohenheim and Geisenheim, with a similar set of parental genotypes, including ‘Cacanska najbolja’, ‘Auerbach’ and ‘President’ (Hartmann and Neumüller 2006). The cluster Fig. 18BII received 90% bootstrap support and grouped the half-siblings of ‘Stanley’ (Hartmann and Neumüller 2006). Another clade, Fig. 18BIII consisted of landraces of Hungarian origin and close to the statistically supported separate cluster of ‘Nemtudom’ and ‘Besztercei’ plums. The origin of those Hungarian plums is unknown but they have evolved in specific geographic regions and their separation from other landraces was also illustrated by SSR analyses and chloroplast DNA sequence variations (Halász *et al.* 2023; Makovics-Zsohár *et al.* 2017).

The cluster Fig. 18BV collects old cultivars from diverse countries and without any reported parentage. Another statistically unsupported cluster existed among Russian and Romanian cultivars and ‘Czar’. It is known that Romanian plum breeding programs back in 1960-1970 frequently used ‘Czar’ as a genitor of early-ripening phenotype (Butac *et al.* 2010). ‘Czar’ is the first European plum cultivar created in England (1843) that was named after the Russian emperor visiting England (Faust and Surányi 1999) and hence it is plausible that the accessions of this genotype made their way into the Russian breeding centers. The PCA analysis provided a more illustrative representation of the four major groups with one containing the two Hungarian landraces and considerably separated from other Hungarian plums, while cultivars of Fig. 18BI-BII were much closer together. Our results indicated that the SCoT-FaSt marker analysis provided enough phylogenetic signal to efficiently reconstruct genetic relationships among and within species.

Approximately one third of the amplified fragments had a size smaller than 1 kb. Small fragment sizes were also detected by the SCoT analysis of *Citrus* (Mahjbi *et al.* 2015), *Punica* (Karapetsi *et al.* 2021), *Simmondsia* (Heikrujam *et al.* 2015) and many other species. Based on the original concept of SCoT marker strategy, it would only be possible if two protein-coding

genes are located in opposite directions, but very close to each other. To verify if many genes are indeed that close to each other in the *Prunus* genome, and to make sure that the resulting fragments are indeed the result of a joint amplification of the SCoT and FaSt primer pairs, we cloned a number of fragments and determined their DNA sequence.

5.6.5 Discussion on functional and evolutionary aspects of the genomic landscape of *FaSt*

Each cloned amplicon was flanked by the 18 and 22 bp sequences of the SCoT and FaSt-R primers at their respective ends, indicating all the randomly chosen fragments were indeed amplified by a combination of those oligonucleotides. This confirms our assumption that fragments amplified by the specific FaSt-R primer predominate among the amplicons. However, six of those sequences contained only the FaSt-R primer sequence and none of the expected ~230 bp part of the *FaSt* element. It indicates the FaSt-R primer indeed anneals specifically but its target site is even more frequent in the genome than the complete *FaSt* elements, which increases the size of amplified fragments in all combinations. Such annealing sites may also have arisen through the accumulation of base substitutions outside *FaSt* or internal deletions within the *FaSt* since mutations in a MITE are most likely not to have a harmful effect on the fitness (Lisch *et al.* 1999; Bromham *et al.* 2015).

The vast majority (Halász *et al.*, 2010) of the sequences contained an identifiable *FaSt* element, 11 of those were inserted in the non-coding parts of protein-coding genes, the UTRs and introns. It fits well to the genome-wide distribution of MITEs in many species which is biased toward UTRs and introns having many more copies than exons (Oki *et al.*, 2008, Halász *et al.*, 2014, Zhou *et al.*, 2016). The other *FaSt* elements (Decroocq *et al.*, 2004) were detected in the intergenic regions that correspond to the insertion preference of *FaSt* in AT-rich sequences (Halász *et al.* 2014). SCoT markers were designed to anneal to the conserved region around the ATG start codon of plant genes (Collard and Mackill 2008). The size of the sequences ranged from 251 to 1,866 bp. In other studies, SCoT analysis produced very similar size ranges of amplicons (Mahjbi *et al.* 2015, Heikrujam *et al.* 2015, Karapetsi *et al.* 2021) what makes it hardly acceptable that SCoT primers amplified regions between protein coding genes in opposite orientations. Very few studies reported sequence data of SCoT amplified fragments to verify SCoT primers annealing to the start codon of genes. A short survey of sequences used to design SCAR markers (listed by Rai 2023) confirmed that most of them did not include the sequences surrounding the start codon. Our partial sequences of a total of 17 protein encoding genes verified that SCoT primers annealed to exonic regions outside the start codon. Plant introns are

AU-rich, in contrast to the GC-rich exons (Goodall and Filipowicz 1991, Xie and Wu 2002) and the higher GC content of exons is important for increased stability of RNA structures, including both mRNA and rRNA (Biro 2008, Šmarda and Bureš 2012). It explains why many SCoT-FaSt combinations amplified parts of the protein and RNA-coding genes. The non-genic segments of the genome show variable levels of GC contents (Šmarda and Bureš 2012), making it possible that some of the SCoT primers annealed to such regions. These results indicate that under the PCR conditions suggested for SCoT analysis (Collard and Mackill 2008), these primers will not selectively bind to the start codon regions but they show a definite tendency to anneal within the GC-rich segments of the genome, which nonetheless creates a good chance of finding genes.

FaSt copies were identified in the intron and 5' and 3' UTRs of several genes, some of which have a well-defined role in important physiological processes. A *FaSt* was identified in the intron of five genes, including the sour cherry *BTB/POZ domain-containing protein At4g08455-like* and *pectinesterase* genes. The BTB/POZ proteins were associated with growth and development (e.g., ethylene biosynthesis, disease resistance, hormone perception) through selective ubiquitination and kinase activation (Gingerich *et al.* 2005). Pectinesterases are involved in pectin degradation and hence they enhance the rate of fruit softening (Ma *et al.* 2023). The introns in two *P. domestica* genes had also *FaSt* insertions. The glycosyltransferase BC10-like proteins have roles in metabolic processes and influence fruit ripening and response to abiotic stresses (Wu *et al.* 2017), while the leaf rust 10 disease-resistance locus receptor-like protein kinase-like proteins are members of LRR-NBS resistance proteins and can enhance resistance to *Xanthomonas* (Foix *et al.* 2021). The *P. dulcis* *DEAD-box ATP-dependent RNA helicase* encoded proteins suppress potyvirus accumulation through an interaction with the viral protein VPg that is required for virus infection (Huang *et al.* 2010).

A *FaSt* copy was present in the 5' UTR of plum *polygalacturonase ADPG1* gene, the protein product of which takes part in pectin degradation and fruit softening (Mahmood *et al.* 2021). The 3'UTR of *P. domestica* *ethylene-responsive transcription factor ERF053-like* and *beta-D-glucosyl crocetin beta-1,6-glucosyltransferase* genes also contained a *FaSt* copy. Ethylene response factor (ERF) transcription factors regulate both fruit ripening and resistance to biotic and abiotic stresses (Li *et al.* 2022), while the latter is an enzyme in the flavonoid and anthocyanin biosynthetic pathways (Wang *et al.* 2022). Since sequence alterations including MITE insertions in UTRs and introns have been frequently associated with regulatory roles (Dietrich *et al.* 2002; Hardy *et al.* 2024, Hu *et al.* 2024), future studies are required to determine if the presence of *FaSt* elements has any consequences on the expression of the respective genes.

The identification of genomic regions homologous to our SCoT-FaSt sequences in other *Prunus* genomes has provided further evolutionary inferences. We identified a *FaSt* element present in one *P. avium* and three *P. cerasus* sequences, which was missing from the homologous genomic regions of all other species in the *Prunus* subgenus. Two of those (Pav_CG_3_910 and Pcer_KJ_2_343) also lacked this *FaSt* element in *P. fruticosa*, and Pcer_KJ_4_1000 was missing from *P. × kanzakura*. The phylogenetic distribution of such *FaSt* copies helped determine the relative time of their insertion. The *FaSt* in Pcer_FM_2_462 was present in all species indicating it has been inserted between 47-52 Mya according to TimeTree 5 estimations based on data from (Chin *et al.* 2014), (Guo *et al.* 2008) and (Pouget *et al.* 2016). The *FaSt* in Pcer_FM_4_496 was present in all species of the *Cerasus* subgenus but missing from all other *Prunus*, so it must have emerged around 47 Mya. Three *FaSts* were inserted approx. 47-43 Mya on the phylogenetic lineage of *P. avium* and *P. cerasus*. In addition, the presence and size of other indels in Pcer_KJ_4_1000 and Pcer_KJ_4_662 also followed the known speciation events in *Prunus* genus (Chin *et al.* 2014). The distorted distribution of some of the *FaSt* elements between the *Cerasus* and *Prunus* subgenera raised the question if a burst of *FaSt* insertions 47-43 Mya was a major evolutionary event. To answer the question, we determined the proportion of complete and partial *FaSt* elements in the genome of several species in the *Cerasus* and *Prunus* subgenera. Our analysis indicated an unequal distribution of differently sized *FaSt* elements with species in the *Cerasus* subgenus having predominantly fragmented *FaSt* copies (smaller than the half of the 349 bp *FaSt*) while most species in the *Prunus* subgenus harbored considerably higher number of complete *FaSt* elements. The only exception was apricot with many partial *FaSt* elements but complete and presumably active *FaSt* copies were also identified in this species (Halász *et al.* 2014, Muñoz-Sanz *et al.* 2017). Polyploid species have 2 to 3-times more *FaSt* copies. The major trend seems to support the hypothesis that many of the *FaSt* elements of cherries are older than those of peach, almond and plums. A sudden burst of *FaSts* in one of the phylogenetic lineages may have resulted in an accelerated differentiation. MITE insertions have been described to be associated with an increased capacity for unequal crossing over (Wang *et al.* 2015) or alternative splicing (Xin *et al.* 2019) and modifications to gene regulation (Yu *et al.* 2024). Similar events have been identified over primate evolution due to a sudden increase of Alu repeats 40 Mya (Batzner and Deininger 2002).

The differential presence of *FaSt* copies in almond sequences was fully consistent with reported connections between germplasm and the pedigree of tested cultivars. Pdul_TK_22_1866 and Pdul_TK_4_761 had a *FaSt* also present in ‘Nonpareil’ and ‘Texas’, respectively. Hungarian almond breeding programs are reported to have used accessions from the

USA, which was also reflected by the self-incompatibility genotypes (Halász *et al.* 2010). The *FaSt* in ‘Tuono’ Pdul_TU_3_544 sequence was also found in ‘Lauranne’, which fits well with ‘Lauranne’ being an offspring of the cross ‘Ferragnés’ × ‘Tuono’ (López *et al.* 2006). Two unique *FaSt* copies in a ‘Tuono’ (Pdul_TU_33_342) and a ‘Tétényi keményhájú’ (Pdul_TK_22_901) sequence suggest the recent insertion of *FaSt*, which is also confirmed by the high proportion of complete *FaSt* elements in the almond genome (Fig. 21) and consistent with the activity of transposable elements in almond genome (Alioto *et al.* 2019). Recent *FaSt* insertions were also seen in the *P. domestica* sequences.

The combination of SCoT primers and a newly designed sequence-specific primer targeting the *Prunus* non-autonomous transposon, *FaSt*, provided a molecular marker with enhanced polymorphism detection, free from non-target amplification, and capable of delivering reliable phylogenetic information at both inter- and intraspecific taxonomic levels. The identification of variations in the genome-wide distribution of *FaSt* insertions enabled the detection of several genes and genomic regions exhibiting insertional polymorphisms. A more detailed analysis of these variations could provide valuable insights into the genetic diversity of breeding germplasm, the reconstruction of evolutionary relationships, and the genome-shaping impact of an active Mutator-type MITE in the *Prunus* genome.

6. CONCLUSIONS AND RECOMMENDATIONS

This thesis demonstrates that the *Falling Stones* (*FaSt*) Mutator-like miniature inverted-repeat transposable element (MITE) represents an active and biologically relevant component of *Prunus* genomes rather than a passive genomic relic. *FaSt* copies are widely distributed across all *Prunus* chromosomes, with notable enrichment on chromosomes 1 and 6, and are transcribed in multiple tissues of both diploid and polyploid species. The presence of full-length *FaSt* insertions within untranslated regions of numerous mRNAs, together with the discovery of a novel promoter insertion in the almond *PdCBF1* gene, provides compelling evidence that *FaSt* has the capacity to influence host gene regulation, particularly at the transcriptional and post-transcriptional levels.

Despite this regulatory potential, transcriptome analyses of SC and SI apricot reproductive tissues revealed only a limited number of differentially expressed genes associated with a *FaSt* insertion in a highly expressed locus. These results indicate that, although *FaSt* can disrupt coding regions such as the *S-haplotype F-box* gene ORF, it does not induce extensive transcriptional reprogramming under the tested conditions. The absence of widespread expression changes further suggests that *FaSt* is unlikely to play a central role in global regulatory pathways, such as miRNA-mediated gene expression control, and may instead exert localized or context-dependent effects.

Building on the widespread and gene-proximal distribution of *FaSt* elements, this work developed a combined SCoT–*FaSt* primer system as a novel molecular marker for *Prunus* species analysis. The SCoT–*FaSt* marker consistently outperformed conventional SCoT markers by delivering higher levels of polymorphism, improved band clarity, and enhanced resolving power, while avoiding non-target amplification. Sequencing of SCoT–*FaSt* amplicons revealed partial sequences of protein-coding genes and RNA-coding gene, with *FaSt* insertions located in 5' and 3' untranslated regions and introns of genes involved in essential physiological processes. These findings support the conclusion that *FaSt*-associated polymorphisms are enriched in regulatory genomic contexts and contribute to functional diversification rather than representing random insertional noise.

The SCoT–*FaSt* marker system also proved robust for phylogenetic reconstruction, accurately resolving evolutionary relationships among six *Prunus* species with strong statistical support and clustering European plum cultivars according to their documented breeding origins.

A biased distribution of *FaSt* elements between the *Cerasus* and *Prunus* subgenera characterized by increased fragmentation in cherry species, suggests that differential *FaSt* dynamics may have contributed to subgeneric divergence. Together, these results confirm the value of *FaSt*-based markers for both evolutionary inference and applied germplasm characterization, consistent with independent evaluations of the SCoT–*FaSt* strategy in *Prunus* species.

From an applied perspective, the findings of this thesis have clear implications for molecular breeding and germplasm management in stone fruits. Other studies using ISSR and SCoT markers have demonstrated their effectiveness in revealing genetic diversity and population structure in almond and other *Prunus* crops, particularly when wild and cultivated materials are jointly analyzed. The integration of *FaSt* transposon specificity into a SCoT framework extends these approaches by linking marker polymorphism to an active, gene-associated genomic element, thereby being biologically interpretable and having a breeding relevance. This is particularly important for breeding programs aiming to broaden the genetic base of cultivated germplasm through informed use of wild relatives.

Based on these results, it is recommended that SCoT–*FaSt* markers be further validated across geographically diverse breeding panels, landraces, and wild populations to assess allele stability, transferability, and informativeness under different genetic and environmental contexts. Conversion of the most informative SCoT–*FaSt* loci into high-throughput genotyping formats would facilitate their integration into marker-assisted selection (MAS) and marker-assisted backcrossing (MAB) schemes. Given their proximity to genes involved in key physiological processes, *FaSt*-associated markers also represent promising candidates for trait-linked studies, including QTL mapping and genomic selection frameworks.

Finally, more detailed functional analyses of *FaSt* insertion polymorphisms, particularly those located in untranslated regions or introns, could provide deeper insight into the long-term genome-shaping role of active MITEs in *Prunus* species. Such studies would not only advance our understanding of transposon-driven regulatory diversification but also support the development of biologically informed molecular tools that shorten selection cycles and increase efficiency in stone fruit breeding programs.

In conclusion, this thesis advances our understanding of *FaSt* as a dynamic genomic element, establishes a robust and functionally anchored marker system for *Prunus* species, and provides a clear foundation for future research and practical applications in molecular breeding, germplasm conservation, and evolutionary genomics of stone fruit trees.

7. NEW SCIENTIFIC RESULTS

I. The *Falling Stones* (*FaSt*) transposon is ubiquitously transcribed in all examined tissues (leaf, flower bud, calyx, petal, anther, and pistil) of diploid (*P. armeniaca* and *P. avium*), tetraploid (*P. cerasus*), and hexaploid (*P. domestica*) *Prunus* species.

II. Quantitative RT-PCR confirmed that *FaSt* expression was higher in the anther tissue of the SC apricot cultivar ‘Pannónia’ than in the anther of the SI European plum (*Prunus domestica* L.) cultivar ‘Hanita’.

III. A novel 358 bp *FaSt* insertion was discovered in the promoter region of the almond (*Prunus dulcis*) *PdCBF1* gene (in reverse orientation), sharing 84.2% identity with the previously described apricot *SFB FaSt* insertion, this *FaSt* element introduces new transcription factor binding sites (SBP/MADS-box, WRKY, SRS) not present in the homologous *PdCBF2* promoter.

IV. A new molecular marker was developed for *Prunus* analysis by the combination of SCoT primers and a newly designed, *Falling Stones* transposon-specific primer. The marker significantly enhanced PCR polymorphism, band clarity, and resolving power compared to SCoT analyses.

V. Partial sequences of a total of 17 protein-coding genes and one RNA-coding gene were identified in SCoT-*FaSt* amplicons where *FaSt* copies were located in the 5’ and 3’ UTRs, and the introns of genes critical for physiological processes, supporting a role for *FaSt* in diversifying gene regulation.

VI. The SCoT-*FaSt* marker analysis is a reliable tool for elucidating phylogenetic relationships. It accurately reconstructed the evolutionary relationships of six *Prunus* species (peach, almond, European plum, apricot, sweet and sour cherries) with 95–100% bootstrap support, and grouped 28 European plum cultivars into five clusters consistent with their breeding origins.

VII. Differential expression analysis on the RNA-Seq data from anther and style tissues of SC versus SI apricot cultivars revealed no statistically significant gene expression changes attributable to the presence of the *FaSt* transposon, confirming that *FaSt* insertion did not markedly alter transcript levels under the tested conditions.

8. SUMMARY

Transposable elements (TEs) are major drivers of genome evolution and variability, profoundly influencing gene regulation and phenotypic diversity in plants. Within the stone fruit genus *Prunus* (*Rosaceae*), which includes economically important species such as apricot, peach, cherry, plum, and almond, TEs have been implicated in key biological processes, including the breakdown of self-incompatibility (SI). However, their transcriptional activity and potential use in crop improvement remain underexplored. This thesis focuses on the characterization and practical exploitation of the *Falling Stones* (*FaSt*) miniature inverted-repeat transposable element (MITE) as both a genomic component and a molecular marker resource to accelerate *Prunus* breeding.

Comparative genomic analyses identified *FaSt* as a non-autonomous MITE characterized by two 82 bp terminal inverted repeats (TIRs) flanking a unique internal region of approximately 116 bp. *FaSt* elements were found to be highly abundant, particularly on chromosomes 1 and 6 and preferentially inserted into untranslated regions (UTRs) and introns of protein-coding genes. These structural features positioned *FaSt* as an excellent candidate for both evolutionary and applied genetic studies.

The first part of this work characterized *FaSt* transcriptional activity through PCR and qRT-PCR assays normalized to *ubiquitin10* (*UBQ10*). A 331 bp *FaSt* transcript was detected across multiple tissues: leaves, calyces, petals, anthers, and styles, in both diploid (*P. armeniaca* and *P. avium*) and polyploid (*P. cerasus* and *P. domestica*) species, with the highest expression observed in anthers. Complementary *in silico* analysis of eleven *Prunus* transcriptomes revealed 121–458 *FaSt* copies per genome, confirming its enrichment on chromosomes 1 and 6. BLASTN screening uncovered a novel 358 bp insertion in the *PdCBF1* promoter of almond (*P. dulcis*), introducing additional MADS-box and WRKY transcription factor binding motifs.

To assess *FaSt*'s transcriptional influence, RNA-seq datasets from SC and SI apricot anthers and styles were analyzed using DESeq2 with independent filtering. Among 27,655 and 29,914 expressed genes, 95 and 75 were differentially expressed in anthers and styles, respectively. Despite *FaSt*-mediated disruption of the *S-haplotype F-box* (*SFB*) gene associated with self-compatibility, no broad transcriptomic reprogramming was observed, indicating that *FaSt* exerts locus-specific rather than global regulatory effects under the tested conditions.

In the final phase of this research, a high-resolution SCoT–FaSt marker system was developed by pairing a 22-nt FaSt-R primer targeting FaSt’s internal loop with 19 SCoT primers. This dual-primer approach amplified *FaSt*-linked regions, yielding sharper and more polymorphic bands than SCoT or FaSt-R alone, with resolving power increases up to 339%. Across six *Prunus* species, the system achieved 95–100% bootstrap-supported clustering in neighbor-joining phylogenies, while analysis of 28 European plum cultivars revealed five distinct genetic groups consistent with known breeding origins. Sequencing of 32 representative amplicons confirmed *FaSt* insertion junctions within UTRs and introns of genes involved in fruit softening, stress response, and disease resistance, demonstrating both specificity and biological relevance.

In summary, this work establishes *Falling Stones* as an active, transcriptionally engaged MITE family within *Prunus*, highlighting its dual role in genome evolution and its practical utility in molecular breeding. Although dominant in nature, the SCoT–FaSt assay represents a reliable and cost-effective tool for germplasm characterization, phylogenetic analysis, and marker-assisted selection. Future research should explore cell-type-specific *FaSt* expression, the epigenetic control of *FaSt* loci, and functional assays using promoter–reporter systems or CRISPR-mediated excision to directly test regulatory effects. Collectively, this thesis bridges fundamental transposon biology with applied genetics, contributing both new scientific understanding and tangible resources for the advancement of stone-fruit breeding.

9. PUBLICATIONS SERVING AS THE BASIS OF THE DISSERTATION:

9.1 Papers related to the thesis:

- Ivanovska, Beti, Thanyarat Onlamun, Júlia Halász, and Attila Hegedűs. (2025). "The Combination of Start-Codon-Targeted (SCoT) and *Falling Stone* (FaSt) Transposon-Specific Primers Provides an Efficient Marker Strategy for *Prunus* Species" *International Journal of Molecular Sciences* 26, no. 9: 3972. <https://doi.org/10.3390/ijms26093972>
- Hegedűs, Attila, Pamulaklakin Eusebio Geronima, Ivanovska Beti, and Júlia Halász. (2023). "Molecular Characterization of the Eukaryotic Translation Initiation Factor 4E Isoform, eIF(iso)4E, in Traditional Cultivars of *Prunus domestica*." *TURKISH JOURNAL OF AGRICULTURE AND FORESTRY* 47 (6): 1155–1164. doi:10.55730/1300-011X.3154.
- Júlia, Halász, Szendy Gergő, Ivanovska Beti, György Tóth Endre, and Hegedűs Attila. (2023). "The Self-Incompatibility Locus and Chloroplast DNA Regions of *Prunus Domestica* Reflect the Origin and Genetic Diversity of Traditional Cultivars." *JOURNAL OF THE AMERICAN SOCIETY FOR HORTICULTURAL SCIENCE* 148 (5):230–239. doi:10.21273/JASHS05330-23.
- Peona, Valentina & Martelossi, Jacopo & Almojil, Dareen & Bocharkina, Julia & Brännström, Ioana & Brown, Max & Cang, Alice & Carrasco-Valenzuela, Tomas & DeVries, Jon & Doellman, Meredith & Elsner, Daniel & Espíndola-Hernández, Pamela & Friis, Guillermo & Gaspar, Bence & Zagorski, Danijela & Hałakuc, Paweł & Ivanovska, Beti & Laumer, Christopher & Lehmann, Robert & Suh, Alexander. (2023). Teaching transposon classification as a means to crowd source the curation of repeat annotation – a tardigrade perspective. *BioRxiv*; doi: <https://doi.org/10.1101/2023.11.06.565293>.

9.2 Other papers:

- Doneva, D., M. Pál, L. Brankova, G. Szalai, J. Tajti, R. Khalil, B. Ivanovska, et al. (2021). "The Effects of Putrescine Pre-Treatment on Osmotic Stress Responses in Drought-Tolerant and Drought-Sensitive Wheat Seedlings." *PHYSIOLOGIA PLANTARUM* 171 (2): 200–216. doi:10.1111/ppl.13150.
- Janda, Tibor, Judit Tajti, Kamirán Á. Hamow, Tihana Marček, Beti Ivanovska, Gabriella Szalai, Magda Pál, Ewa D. Zalewska, and Éva Darkó. (2021). "Acclimation of Photosynthetic

Processes and Metabolic Responses to Elevated Temperatures in Cereals.” *PHYSIOLOGIA PLANTARUM* 171 (2): 217–231. doi:10.1111/ppl.13207.

Pál, Magda, Beti Ivanovska, Tímea Oláh, Judit Tajti, Kamirán Áron Hamow, Gabriella Szalai, Radwan Khalil, et al. (2019). “Role of Polyamines in Plant Growth Regulation of Rht Wheat Mutants.” *PLANT PHYSIOLOGY AND BIOCHEMISTRY* 137: 189–202. doi:10.1016/j.plaphy.2019.02.013.

9.3 Publications:

Beti Ivanovska, Júlia Halász, and Attila Hegedűs. (2023). “Application of SCoT and MITE Insertion-Based Polymorphism Markers for Assessing the Genetic Diversity of *Prunus* Species.” In Abstract Book of XVI Eucarpia Symposium on Fruit Breeding and Genetics,

Beti Ivanovska, Onlamun Thanyarat, Júlia Halász, and Attila Hegedűs. (2023). “Assessment of Genetic Diversity of European Plum Cultivars Using Transposon-Based Markers.” In XXIX. Növénynevelési Tudományos Napok, 127.

Halász, J., B. Ivanovska, T. Onlamun, G. Szendy, E.Gy. Tóth, and A. Hegedűs. (2023). “Application and Development of Molecular Marker Systems for the Analysis of Polyploid *Prunus* Species.” In Abstract Book of Plant Biology 2023 Annual Meeting of the American Society of Plant Biologists.

Hegedűs, A., B. Ivanovska, T. Nagy, and J. Halász. (2023). “Steps toward Unraveling the Secrets of a Mysterious Transposon: The *Falling Stones* in Stone Fruits.” In Abstract Book of Plant Biology 2023 Annual Meeting of the American Society of Plant Biologists.

Ivanovska Beti, Attila Hegedűs, and Júlia Halász. (2022). “In Silico Analysis of *Falling Stones* (*FaSt*) Transposable Element Inserted in the Promoter Region of an Almond *CBF* Gene.” In A Lippay János – Ormos Imre – Vas Károly (LOV) Tudományos Ülésszak Tanulmányai [Proceedings of János Lippay – Imre Ormos – Károly Vas (LOV) Scientific Meeting], 365–369.

Ivanovska, B, J Halász, and A Hegedűs. (2020). “The Presence of *FaSt* Transposon in *Prunus* Transcriptome.” In SZIENTific Meeting for Young Researchers 2020, 204–208.

10. APPENDICES

Appendix A1. Bibliography

- Al-Hmoud, G. & Sözen, E., (2026). Genetic diversity and structure of cultivated and wild almond genotypes in Jordan revealed by inter-simple sequence repeat (ISSR) and start codon targeted (SCoT) markers. *Genetic Resources and Crop Evolution*, 73, 59. <https://doi.org/10.1007/s10722-025-02627-1>.
- Alioto, T., Alexiou, K.G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., Dhingra, A., Duval, H., Fernández i Martí, Á., Frias, L., Galán, B., García, J.L., Howad, W., Gómez-Garrido, J., Gut, M., Julca, I., Morata, J., Puigdomènech, P., Ribeca, P., Rubio Cabetas, M.J., Vlasova, A., Wirthensohn, M., Garcia-Mas, J., Gabaldón, T., Casacuberta, J.M. and Arús, P. (2020), Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *Plant J*, 101: 455-472. <https://doi.org/10.1111/tpj.14538>
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990), Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)803602](https://doi.org/10.1016/S0022-2836(05)803602).
- Amiryousefi, A., Hyvönen, J., Poczai, P. (2018), iMEC: Online marker efficiency calculator. *Appl. Plant Sci.* 6, e01159. <https://doi.org/10.1002/aps3.1159>
- Anders, S., Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* 11, R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Atapour, Y., Ghanbari, A., Estaji, A. and Haghjooyan, R., 2022. Evaluation of genetic diversity of twenty-eight sweet cherry genotypes by morphological traits and SCoT markers in the Northwest of Iran. *Journal of Genetic Resources*, 8(2), pp.138-146.
- Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., & Wishart, D. S. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic acids research*, 44(W1), W147–W153. <https://doi.org/10.1093/nar/gkw419>
- Bajashvili, E.I. (1991) Studies of some species of *Prunus* mill. genus. *Acta Hortic. (ISHS)* 283, 31–34.
- Baránek, M.; Meszáros, M.; Sochorová, J.; Čechová, J.; Raddová, J. (2012) Utility of retrotransposon-derived marker systems for differentiation of presumed clones of the apricot cultivar Velkopavlovická. *Sci. Hortic.-Amsterdam* , 143, 1-6. <https://doi.org/10.1016/j.scienta.2012.05.022>.
- Barros, P.M., Gonçalves, N., Saibo, N.J.M., Margarida Oliveira, M., (2012). Functional characterization of two almond C-repeat-binding factors involved in cold response. *Tree Physiol.* 32, 1113–1128. <https://doi.org/10.1093/treephys/tps067>

- Barros, P. M., Gonçalves, N., Saibo, N. J., & Oliveira, M. M. (2012). Cold acclimation and floral development in almond bud break: insights into the regulatory pathways. *Journal of experimental botany*, 63(12), 4585–4596. <https://doi.org/10.1093/jxb/ers144>
- Barteri, F., (2019). Impact of transposition on the generation of genetic variability in *Prunus* crop species. Candidate : Fabio Barteri <http://hdl.handle.net/10803/669550>. ISBN 9788449093098. <<https://ddd.uab.cat/record/240401>>
- Barteri, F., Duval, H., Fern, A., Frias, L., Gal, B., Howad, W., Cabetas, J.R., (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes : results from the almond genome sequence 455–472. <https://doi.org/10.1111/tpj.14538>
- Batzer, M.A.; Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nat. Rev. Genet.* , 3, 370-379. <https://doi.org/10.1038/nrg798>
- Baud, A., Wan, M., Nouaud, D., Francillonne, N., Anxolabéhère, D., Quesneville, H., (2019). Traces of transposable elements in genome dark matter co-opted by flowering gene regulation networks. *bioRxiv* 1–26. <https://doi.org/10.1101/547877>
- Bielenberg, D.G., Wang, Y.E., Li, Z., Zhebentyayeva, T., Fan, S., Reighard, G.L., Scorza, R., (2008). Sequencing and annotation of the evergrowing locus in peach [*Prunus persica* (L .) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation 495–507. <https://doi.org/10.1007/s11295-007-0126-9>
- Biro, J.C. (2008) Correlation between nucleotide composition and folding energy of coding sequences with special attention to wobble bases. *Theor. Biol. Med. Model.* 5, 14. <https://doi.org/10.1186/1742-4682-5-14>
- Blanchard, I., Bui, Q.T., Mergez, A., Denni, S., Cornille, A., Groppi, A. et al., (2025). Building pangenomes for domesticated and wild tree species: genomic complexity and strategies (Armeniaca section pangenome). *bioRxiv* preprint. doi: 10.1101/2025.07.22.665893.
- Bolger, AM, Lohse, M, Usadel, B. (2014). Trimmomatic: a Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30(15):2114–20. doi:10.1093/bioinformatics/btu170.
- Bourgon, R, Gentleman, R, Huber, W. (2010) Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci U S A.* 107(21):9546– 51. doi: 10.1073/pnas.0914005107. Epub. PMID: 20460310; PMCID: PMC2906865.
- Bray, NL, Pimentel, H, Melsted, P, Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34:525–27.. <https://doi.org/10.1038/nbt.3519>
- Bromham, L. Hua, X. Lanfear, R. Cowman, P.F. (2015) Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. *Am. Nat.* 185, 507-524. <https://doi.org/10.1086/680052>
- Buer, H., Rula, S., Wang, Z. Y., Fang, S., Bai, Y. E. (2022) Analysis of genetic diversity in *Prunus sibirica* L. in inner Mongolia using SCoT molecular markers. *Genet. Resour. Crop Ev.* 69, 1057-1068. <https://doi.org/10.1007/s10722-021-01284-4>
- Butac, M., Zagrai, I. and Botu, M. (2010) Breeding of New Plum Cultivars in Romania. *Acta Hortic.* 874, 51-58. DOI: 10.17660/ActaHortic.2010.874.5
- Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., Martin, C., (2012). Retrotransposons control fruit-specific, cold-dependent accumulation

- of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255. <https://doi.org/10.1105/tpc.111.095232>
- Cachi, A.M., Hedhly, A., Hormaza, J.I., Wünsch, A., (2014). Pollen tube growth in the self-compatible sweet cherry genotype, ‘Cristobalina’, is slowed down after self-pollination. *Annals of Applied Biology* 164, 73–84. <https://doi.org/10.1111/aab.12079>
- Campoy, J.A., Sun, H., Goel, M. *et al.* (2020). Gamete binning: chromosome-level and haplotype-resolved genome assembly enabled by high-throughput single-cell sequencing of gamete genomes. *Genome Biol* 21, 306 <https://doi.org/10.1186/s13059-020-02235-5>
- Canton, M., Forestan, C., Bonghi, C., & Varotto, S. (2021). Meta-analysis of RNA-Seq studies reveals genes with dominant functions during flower bud endo- to eco-dormancy transition in *Prunus* species. *Scientific reports*, 11(1), 13173. <https://doi.org/10.1038/s41598-021-92600-6>
- Canton, M., Forestan, C., Marconi, G., Carrera, E., Bonghi, C., & Varotto, S. (2022). Evidence of chromatin and transcriptional dynamics for cold development in peach flower bud. *The New phytologist*, 236(3), 974–988. <https://doi.org/10.1111/nph.18393>
- Cao, C., Xu, J., Zheng, G., Zhu, X., (2016). Evidence for the role of transposons in the recruitment of cis -regulatory motifs during the evolution of C4 photosynthesis. *BMC Genomics* 1– 11. <https://doi.org/10.1186/s12864-016-2519-3>
- Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., Jimenez Vieira, F. R., Bowler, C., Dorrell, R. G., Gitzendanner, M. A., et al. (2019). Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *GigaScience*, 8(10), giz126. <https://doi.org/10.1093/gigascience/giz126>.
- Carrasco, B, Meisel, L, Gebauer, M, Garcia-Gonzales, R, Silva, H.(2013) Breeding in peach, cherry and plum: from a tissue culture, genetic, transcriptomic and genomic perspective. *Biol Res.* ;46(3):219-30. doi: 10.4067/S0716-97602013000300001. PMID: 24346068.
- Castanera, R., Tomás, C., Ruggieri, V., Vicient, C., Eduardo, I., Aranzana, M., Arús, P., Casacuberta, J.(2024) A phased genome of the highly heterozygous ‘Texas’ almond uncovers patterns of allele- specific expression linked to heterozygous structural variants. *Hortic. Res.* 11, uhae106. <https://doi.org/10.1093/hr/uhae106>
- Chen, JW, Shrestha, L, Green, G, Leier, A, Marquez-Lago, TT. (2023). The hitchhikers’ guide to RNA sequencing and functional analysis. *Briefings in Bioinformatics* 24(1):bbac529. <https://doi.org/10.1093/bib/bbac529>.
- Chen, J., Xu, H., Zhang, J., Dong, S., Liu, Q., & Wang, R. (2021). Transcriptomic analysis reveals candidate genes for male sterility in *Prunus sibirica*. *PeerJ*, 9, e12349. <https://doi.org/10.7717/peerj.12349>
- Chin, S.-W., Shaw, J., Haberle, R., Wen, J., Potter, D. (2014). Diversification of almonds, peaches, plums and cherries - Molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Mol. Phylogenet. Evol.* 76, 34-48. <https://doi.org/10.1016/j.ympev.2014.02.024>

- Collard, B.C.Y.; Mackill, D.J. (2009) Start Codon Targeted (SCoT) Polymorphism: A Simple, Novel DNA Marker Technique for Generating Gene-Targeted Markers in Plants. *Plant Mol. Biol. Rep.* 27, 86–93. doi: 10.1007/s11105-008-0060-5.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology*, 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Cooke, J.E.K., Eriksson, M.E., Junttila, O., (2012). The dynamic nature of bud dormancy in trees: Environmental control and molecular mechanisms. *Plant, Cell Environ.* 35, 1707–1728. <https://doi.org/10.1111/j.1365-3040.2012.02552.x>
- Crane, M.B. and Lawrence, W.J.C. (1934) The genetic of garden plants. Mac Millan, London.
- Cronn, R., Small, R. L., Haselkorn, T., & Wendel, J. F. (2003). Cryptic repeated genomic recombination during speciation in *Gossypium gossypoides*. *Evolution; international journal of organic evolution*, 57(11), 2475–2489. <https://doi.org/10.1111/j.0014-3820.2003.tb01493.x>
- Darwin, C.R., (1843). Double flowers-their origin. *Gardeners' Chronicle*, 628.
- De Mori, G, Cipriani, G. (2023) Marker-Assisted Selection in Breeding for Fruit Trait Improvement: A review. *International Journal of Molecular Sciences*, 24(10), 8984. doi: 10.3390/ijms24108984. PMID: 37240329; PMCID: PMC10219056.
- Decroocq, V., Hagen, L., Favé, M.G. *et al.* (2004) Microsatellite markers in the hexaploid *Prunus domestica* species and parentage lineage of three European plum cultivars using nuclear and chloroplast simple- sequence repeats. *Mol. Breeding*, 13, 135–142. <https://doi.org/10.1023/B:MOLB.0000018761.04559.b3>
- Dietrich, C. R., Cui, F., Packila, M. L., Li, J., Ashlock, D. A., Nikolau, B. J., & Schnable, P. S. (2002). Maize Mu transposons are targeted to the 5' untranslated region of the *gl8* gene and sequences flanking Mu target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics*, 160(2), 697–716. <https://doi.org/10.1093/genetics/160.2.697>
- Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, J., Poizat, C., Zanetto, A., Arús, P., & Laigret, F. (2002). Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 105(1), 127–138. <https://doi.org/10.1007/s00122-002-0867-7>
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. Dobin A, Gingeras TR. (2015). Mapping RNA-Seq Reads with STAR. *Current Protocols in Bioinformatics*. doi:10.1002/0471250953.bi1114s51.
- Dudoit S, Yang YH, Callow MJ, Speed TP. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* 111–39. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.9702>.
- Endlich, J. von and Murawski, H. (1962) Contribution to breeding research on plum. III.

- Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., Vizzotto, G., Morgante, M., (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant Journal* 76, 175–187. <https://doi.org/10.1111/tpj.12283>
- Fang, Z.Z., Zhou, D.R., Ye, X.F., Jiang, C.C. & Pan, S.L., (2016). Identification of candidate anthocyanin-related genes by transcriptomic analysis of ‘Furongli’ plum (*Prunus salicina*) during fruit ripening using RNA-seq. *Frontiers in Plant Science*, 7, p.1338. doi: 10.3389/fpls.2016.01338.
- FAO. (2022). Available online: <https://www.fao.org/faostat/en/#data/QCL> (accessed on 05 February 2025).
- Food and Agriculture Organization of the United Nations (FAO) (2024) *FAOSTAT — QCL: Crops and Livestock Products* [online]. Available at: <https://www.fao.org/faostat/en/#data/QCL>
- United States Department of Agriculture, Foreign Agricultural Service (USDA FAS) (2024) *China: Stone Fruit Annual* [online]. Available at: <https://www.fas.usda.gov/data/china-stone-fruit-annual-8> (Accessed: 12 January 2026).
- United States Department of Agriculture, Foreign Agricultural Service (USDA FAS) (2024) *Production — Almonds* [online]. Available at <https://www.fas.usda.gov/data/production/commodity/0577400> (Accessed: 12 January 2026).)
- Faust, M., & Surányi, D. (2010). Origin and Dissemination of Plums.
- Feng, C., Wu, C., Wang, J., Wang, W., Yan, G., Zhou, Y., Zhang, K., Zhang, X., & Duan, X. (2025). Identification of Self-Incompatibility Related Genes in Sweet Cherry Based on Transcriptomic Analysis. *Biology*, 14(9), 1125. <https://doi.org/10.3390/biology14091125>
- Fernandez, L., Torregrosa, L., Segura, V., Bouquet, A., Martinez-Zapater, J.M., (2010). Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine 545–557. <https://doi.org/10.1111/j.1365-313X.2009.04090.x>
- Feschotte, C., Swamy, L., & Wessler, S. R. (2003). Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics*, 163(2), 747–758. <https://doi.org/10.1093/genetics/163.2.747>
- Feschotte, C., Pritham, E.J., (2007). DNA transposons and the evolution of eukaryotic genomes. Annual review of genetics 41, 331–368. <https://doi.org/10.1146/annurev.genet.40.110405.090448>
- Fiol, A., García, S., Dujak, C., Pacheco, I., Infante, R., Aranzana, MJ.. (2022). An LTR retrotransposon in the promoter of a *PsMYB10.2* gene associated with the regulation of fruit flesh color in Japanese plum. *Hortic Res.* 9:uhac206. 10.1101/2022.01.26.477575
- Foix, L., Nadal, A., Zagorščak, M., Ramšak, Ž., Esteve-Codina, A., Gruden, K., & Pla, M. (2021). *Prunus persica* plant endogenous peptides PpPep1 and PpPep2 cause PTI-like

- transcriptome reprogramming in peach and enhance resistance to *Xanthomonas arboricola* pv. Pruni. *BMC genomics*, 22(1), 360. <https://doi.org/10.1186/s12864-021-07571-9>
- Foster, T. M., & Aranzana, M. J. (2018). Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Horticulture research*, 5, 44. <https://doi.org/10.1038/s41438-018-0062-x>
- Fridborg, I., Kuusk, S., Moritz, T., Sundberg, E., (1999). The arabidopsis dwarf mutant shi exhibits reduced gibberellin responses conferred by overexpression of a new putative zinc finger protein. *Plant Cell* 11, 1019–1031. <https://doi.org/10.1105/tpc.11.6.1019>
- Fu, X.L., Xiao, W., Wang, D.L., Chen, M., Tan, Q.P., Li, L., De, X., (2014). Roles of Endoplasmic Reticulum Stress and Unfolded Protein Response Associated Genes in Seed Stratification and Bud Endodormancy during Chilling Accumulation in *Prunus persica* 9, 1–8. <https://doi.org/10.1371/journal.pone.0101808>
- Fu, W., da Silva Linge, C., Lawton, JM., Gasic, K. (2022). Feasibility of genomic prediction for brown rot (*Monilinia* spp.) resistance in peach. *Fruit Research* 2: 2 <https://doi.org/10.48130/FruRes-2022-0002>
- García-Gómez, B., Razi, M., Salazar, J.A. *et al.* (2018) Comparative Analysis of SSR Markers Developed in Exon, Intron, and Intergenic Regions and Distributed in Regions Controlling Fruit Quality Traits in *Prunus* Species: Genetic Diversity and Association Studies. *Plant Mol Biol Rep* 36, 23–35. <https://doi.org/10.1007/s11105-017-1058-7>
- García-Gómez, B.E., Ruiz, D., Salazar, J.A., Rubio, M., Martínez-García, P.J. & Martínez-Gómez, P., (2020). Analysis of Metabolites and Gene Expression Changes Relative to Apricot (*Prunus armeniaca* L.) Fruit Quality During Development and Ripening. *Frontiers in Plant Science*, 11, p.1269. doi:10.3389/fpls.2020.01269.
- Gingerich, D.J., Gagne, J.M., Salter, D.W., Hellmann, H., Estelle, M., Ma, L., *et al.* (2005). Cullins 3a and 3b assemble with members of the broad complex/tramtrack/bric-a-brac (BTB) protein family to form essential ubiquitin-protein ligases (E3s) in *Arabidopsis*. *J. Biol. Chem.* 280, 18810–18821. doi: 10.1074/jbc.M413247200
- Giri, M.S., Jingade, P., Huded, A.K.C. *et al.* (2024) Genetic characterization of *Paramyrothecium roridum* strains isolated from different host plants in coffee agroecosystem using SRAP, SCoT markers and ITS sequencing. *Indian Phytopathology* 77, 657–668. <https://doi.org/10.1007/s42360-024-00753-x>
- Glaus, P., Honkela, A., Rattray, M. (2012). Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics* 28:1721–28. <https://doi.org/10.1093/bioinformatics/bts260>
- Gómez, E.M., Buti, M., Sargent, D.J., Dicenta, F., Ortega, E., Martínez-García, P.J. *et al.*, (2019). Transcriptomic analysis of pollen–pistil interactions in almond (*Prunus dulcis*) identifies candidate genes for components of gametophytic self-incompatibility. *Tree Genetics & Genomes*, 15, article 53. doi: 10.1007/s11295-019-1360-7.
- González, M., Salazar, E., Castillo, J., Morales, P., Mura-Jornet, I., Maldonado, J., *et al.* (2016b). Genetic structure based on EST–SSR: a putative tool for fruit color selection in Japanese

- plum (*Prunus salicina* L.) breeding programs. *Mol. Breed.* 36, 1–15. doi: 10.1007/s11032-016-0491-x
- Goodall, G. J., & Filipowicz, W. (1991). Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *The EMBO journal*, 10(9), 2635–2644. <https://doi.org/10.1002/j.1460-2075.1991.tb07806.x>
- Guo, A.Y., Zhu, Q.H., Gu, X., Ge, S., Yang, J., Luo, J., (2008). Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene* 418, 1–8. <https://doi.org/10.1016/j.gene.2008.03.016>
- Guo, L., Li, H., Luoa, Z. (2018). New insights into the genetic diversity and species identification of the native apricots in Southern Xinjiang of China. *Genet. Mol. Res.* 17(1): gmr16039875. <https://doi.org/10.4238/gmr16039874>
- Guo, X., Fang, D., Sahu, S. K., Yang, S., Guang, X., Folk, R., Smith, S. A., Chanderbali, A. S., Chen, S., Liu, M., Yang, T., Zhang, S., Liu, X., Xu, X., Soltis, P. S., Soltis, D. E., & Liu, H. (2021). Chloranthus genome provides insights into the early diversification of angiosperms. *Nature communications*, 12(1), 6930. <https://doi.org/10.1038/s41467-021-26922-4>
- Halász, J., Fodor, A., Pedryc, A., Hegedűs, A. (2010). S-genotyping of Eastern European almond cultivars: identification and characterization of new (*S*₃₆–*S*₃₉) self-incompatibility ribonuclease alleles. *Plant Breeding* 129, 227–232. <https://doi.org/10.1111/j.1439-0523.2009.01686.x>
- Halász, J., Kodad, O., Hegedűs, A., (2014). Identification of a recently active *Prunus*-specific non- autonomous Mutator element with considerable genome shaping force. *Plant Journal* 79, 220–231. <https://doi.org/10.1111/tpj.12551>
- Halász, J., Kodad, O., Galiba, G.M., Skola, I., Ercisli, S., Ledbetter, C.A., Hegedűs, A. (2019). Genetic Variability is Preserved among Strongly Differentiated and Geographically Diverse Almond Germplasm: An Assessment by Simple Sequence Repeat Markers. *Tree Genet. Genomes* 15, 12. <https://doi.org/10.1007/s11295-019-1319-8>
- Halasz, J., Ossama, K., Hegedus, A., (2019). Self-compatibility in *Prunus*: Accidents in transposon traffic? *Acta Horticulturae* 1231, 123–130. <https://doi.org/10.17660/ActaHortic.2019.1231.21>
- Halász, J., Molnár, A. B., Ilhan, G., Ercisli, S., & Hegedűs, A. (2021). Identification and Molecular Analysis of Putative Self-Incompatibility Ribonuclease Alleles in an Extreme Polyploid Species, *Prunus laurocerasus* L. *Frontiers in plant science*, 12, 715414. <https://doi.org/10.3389/fpls.2021.715414>
- Halász, J., Szendy, G., Ivanovska, B., Tóth, E.Gy., Hegedűs, A. (2023). The Self- incompatibility Locus and Chloroplast DNA Regions of *Prunus domestica* Reflect the Origin and Genetic Diversity of Traditional Cultivars. *J. Amer. Soc. Hort. Sci.* 148, 230–239. <https://doi.org/10.21273/JASHS05330-23>
- Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

- Hammer, O. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 9.
- Hardy, E. C., & Balcerowicz, M. (2024). Untranslated yet indispensable-UTRs act as key regulators in the environmental control of gene expression. *Journal of experimental botany*, 75(14), 4314–4331. <https://doi.org/10.1093/jxb/erae073>
- Hartmann, W.; Neumüller, M. (2006). Breeding for resistance: breeding for Plum pox virus resistant plums (*Prunus domestica* L.) in Germany. *EPPO Bulletin* , 36, 332–336. <https://doi.org/10.1111/j.1365-2338.2006.01010.x>
- Hartmann, W., Neumüller, M. (2009). Plum Breeding. In: Breeding Plantation Tree Crops: Temperate Species. Springer, New York, NY. https://doi.org/10.1007/978-0-387-71203-1_6
- Hauck, N.R., Ikeda, K., Tao, R., Iezzoni, A.F., (2006a). The mutated S1-haplotype in sour cherry has an altered S-haplotype-specific F-box protein gene. *Journal of Heredity* 97, 514–520. <https://doi.org/10.1093/jhered/esl029>
- Hauck, N.R., Yamane, H., Tao, R., Iezzoni, A.F., (2006b). Accumulation of nonfunctional S-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid *Prunus*. *Genetics* 172, 1191–1198. <https://doi.org/10.1534/genetics.105.049395>
- Hayer, K. E., Pizarro, A., Lahens, N. F., Hogenesch, J. B., & Grant, G. R. (2015). Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. *Bioinformatics (Oxford, England)*, 31(24), 3938–3945. <https://doi.org/10.1093/bioinformatics/btv488>
- HEGEDUS, A, EUSEBIO, G, IVANOVSKA, B, & HALÁSZ, J (2023). Molecular characterization of the eukaryotic translation initiation factor 4E isoform,eIF(iso)4E, in traditional cultivars of *Prunus domestica*. *Turkish Journal of Agriculture and Forestry* 47 (6): 1155-1164. <https://doi.org/10.55730/1300-011X.3154>
- Heikrujam, M., Kumar, J., & Agrawal, V. (2015). Genetic diversity analysis among male and female Jojoba genotypes employing gene targeted molecular markers, start codon targeted (SCoT) polymorphism and CAAT box-derived polymorphism (CBDP) markers. *Meta gene*, 5, 90–97. <https://doi.org/10.1016/j.mgene.2015.06.001>
- Hawkins, J. S., Hu, G., Rapp, R. A., Grafenberg, J. L., & Wendel, J. F. (2008). Phylogenetic determination of the pace of transposable element proliferation in plants: copia and LINE-like elements in *Gossypium*. *Genome*, 51(1), 11–18. <https://doi.org/10.1139/g07-099>
- Hiroko, H., Takehiko, S., Hiroshi, F., Akiko, I., Yoshiki, K. (2006), Ethylene-regulation of fruit softening and softening-related genes in peach, *Journal of Experimental Botany*, Volume 57, Pages 4071–4077, <https://doi.org/10.1093/jxb/erl178>
- Hu, J., Liu, C., Du, Z., Guo, F., Song, D., Wang, N., Wei, Z., Jiang, J., Cao, Z., Shi, C., Zhang, S., Zhu, C., Chen, P., Larkin, R. M., Lin, Z., Xu, Q., Ye, J., Deng, X., Bosch, M., Franklin-Tong, V. E., ... Chai, L. (2024). Transposable elements cause the loss of self-incompatibility in citrus. *Plant biotechnology journal*, 22(5), 1113–1131. <https://doi.org/10.1111/pbi.14250>

- Huang, S., Lee Hyun Sook, Karunanandaa, B., Kao Teh hui, (1994). Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen. *Plant Cell* 6, 1021–1028. <https://doi.org/10.1105/tpc.6.7.1021>
- Huang, T. S., Wei, T., Laliberté, J. F., & Wang, A. (2010). A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. *Plant physiology*, 152(1), 255–266. <https://doi.org/10.1104/pp.109.147983>
- Jafari, M., Shiran, B., Rabiei, G., Ravash, R., Sayed Tabatabaei, B.E., Martínez-García, P.J. & Martínez-Gómez, P., (2022). Identification of genes associated with kernel size in almond [*Prunus dulcis*] using RNA-Seq. *Plant Growth Regulation*, 97(2), pp.357–373. doi: 10.1007/s10725-022-00801-4.
- Janick, J.. (2005). The origin of fruits, fruit growing and fruit breeding. *Plant Breeding Reviews*. 25. 255-320. <https://doi.org/10.1002/9780470650301.ch8>
- Janik, J. (2008). Origins of Fruit Culture and Fruit Breeding. *Plant Breeding: The Arnel R. Hallauer International Symposium*, 269–282. Investigation on interspecific hybrids of *P. spinosa* L. *Züchter* 32, 121–133. <https://doi.org/10.1002/9780470752708.ch19>
- Jiang, H., & Wong, W. H. (2009). Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics* (Oxford, England), 25(8), 1026–1032. <https://doi.org/10.1093/bioinformatics/btp113>
- Joly-Lopez, Z., Forczek, E., Hoen, D.R., Juretic, N., Bureau, T.E., (2012). A Gene Family Derived from Transposable Elements during Early Angiosperm Evolution Has Reproductive Fitness Benefits in *Arabidopsis thaliana*. *PLoS Genetics* 8. <https://doi.org/10.1371/journal.pgen.1002931>
- Ivanovska, B., Hegedűs, A., (2021). *In silico* analysis of the occurrence of *Falling Stones* (*FaSt*) transposable elements in *Prunus* genome and transcriptome, in: 5th Uppsala Transposon Symposium. Virtual.
- Kajala, K., Brown, N.J., Williams, B.P., Borrill, P., Taylor, L.E., Hibberd, J.M., (2012). Multiple *Arabidopsis* genes primed for recruitment into C4 photosynthesis 47–56. <https://doi.org/10.1111/j.1365-313X.2011.04769.x>
- Kalendar, R., Antonius, K., Smýkal, P., & Schulman, A. H. (2010). iPBS: a universal method for DNA fingerprinting and retrotransposon isolation. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 121(8), 1419–1430. <https://doi.org/10.1007/s00122-010-1398-2>
- Karapetsi, L., Pantelidis, G., Pratsinakis, E. D., Drogoudi, P., & Madesis, P. P. (2006). P. in a Pomegranate Ex Situ (*Punica granatum* L.) Collection in Greece. *Agriculture*, 11(6), 482. <https://doi.org/10.3390/agriculture11060482>
- Karimi, M., Ghazanfari, F., Fadaei, A., Ahmadi, L., Shiran, B., Rabei, M. & Fallahi, H., (2016). The small-RNA profiles of almond (*Prunus dulcis* Mill.) reproductive tissues in response to cold stress. *PLOS ONE*, 11(6), e0156519. doi: 10.1371/journal.pone.0156519.
- Kejnovsky, E., Hawkins, J.S., McClintock, B., Brink, R.A., Emerson, R., (2012). Plant

- Transposable Elements: Biology and Evolution 17–35. <https://doi.org/10.1007/978-3-7091-1130-7>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Klein, J., Saedler, H., Huijser, P., (1996). A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene SQUAMOSA. *Mol. Gen. Genet.* 250, 7–16. <https://doi.org/10.1007/BF02191820>
- Ksouri, N., Jiménez, S., Wells, C.E., Contreras-Moreira, B. & Gogorcena, Y., (2016). Transcriptional responses in root and leaf of *Prunus persica* under drought stress using RNA sequencing. *Frontiers in Plant Science*, 7, 1715. doi: 10.3389/fpls.2016.01715.
- Kumar, S.; Stecher, G.; Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
- Lang, G.A., Early, J.D., Darnell, R., Martin, G.C., (1987). Endo-, Para-, and Eco-dormancy: Physiological Terminology and Classification for Dormancy Research. *Hortscience* 22, 371–377.
- Larionov, A., Krause, A., Miller, W. (2005). A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*. 6:62. doi: 10.1186/1471-2105-6-62. PMID: 15780134; PMCID: PMC1274258.
- Li, S., Wu, P., Yu, X., Cao, J., Chen, X., Gao, L., Chen, K., & Grierson, D. (2022). Contrasting Roles of Ethylene Response Factors in Pathogen Response and Ripening in Fleshy Fruit. *Cells*, 11(16), 2484. <https://doi.org/10.3390/cells11162484>
- Lisch, D., Girard, L., Donlin, M., & Freeling, M. (1999). Functional analysis of deletion derivatives of the maize transposon MuDR delineates roles for the MURA and MURB proteins. *Genetics*, 151(1), 331–341. <https://doi.org/10.1093/genetics/151.1.331>
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12, 323. <https://doi.org/10.1186/1471-2105-12-323>
- Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*, 41(10), e108. <https://doi.org/10.1093/nar/gkt214>
- Liao, Y., Smyth, G.K., Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. (7):923–30. doi: 10.1093/bioinformatics/btt656. Epub PMID: 24227677.
- Liu, Y., Dang, P., Liu, L., He, C., (2019). Cold acclimation by the CBF–COR pathway in a changing climate: Lessons from *Arabidopsis thaliana*. *Plant Cell Rep.* 38, 511–519. <https://doi.org/10.1007/s00299-019-02376-3>

- Liu, X., Chen, M., Wen, B., Fu, X., Li, D., Chen, X. et al., (2019). Transcriptome analysis of peach (*Prunus persica*) fruit skin and differential expression of related pigment genes. *Scientia Horticulturae*, 250, pp.271–277. doi: 10.1016/j.scienta.2019.02.058.
- Love, M.I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Love MI, Anders S, Kim V and Huber W. (2016). RNA-Seq workflow: gene-level exploratory analysis and differential expression [version 2; peer review: 2 approved]. *F1000Research*, 4:1070 <https://doi.org/10.12688/f1000research.7035.2>
- López, M., Vargas, F.J. & Batlle, I.(2006) self-(in)compatibility almond genotypes: A review. *Euphytica* 150, 1–16. <https://doi.org/10.1007/s10681-005-9009-z>
- Ma, K.F., Zhang, Q.X., Cheng, T.R., Yan, X.L., Pan, H.T., Wang, J., (2018). Substantial epigenetic variation causing flower color chimerism in the ornamental tree *Prunus mume* revealed by single base resolution methylome detection and transcriptome sequencing. *International Journal of Molecular Sciences* 19, 1–28. <https://doi.org/10.3390/ijms19082315>
- Ma, K., Zhang, Q., Cheng, T., Wang, J. (2018) Identification of transposons near predicted lncRNA and mRNA pools of *Prunus mume* using an integrative transposable element database constructed from Rosaceae plant genomes. *Mol Genet Genomics*. 293(5):1301- 1316. doi: 10.1007/s00438-018-1449-y. Epub 2018 May 26. PMID: 29804262.
- Ma, Y., Zhang, X., Zhang, W., Li, L., Cheng, S., Guo, M., Chen, G. (2023). Transcriptome analysis reveals the mechanism of delayed softening of ‘France’ prune (*Prunus domestica* L.) during storage at near-freezing temperature. *LWT*, 189, 115446. <https://doi.org/10.1016/j.lwt.2023.115446>
- Mahjbi, A., Baraket, G., Oueslati, A., and Salhi-Hannachi, A.,(2015) Start Codon Targeted (SCoT) markers provide new insights into the genetic diversity analysis and characterization of Tunisian *Citrus* species. , *Biochemical Systematics and Ecology*, vol. 61, Elsevier, pp. 390–398. doi:10.1016/j.bse.2015.07.017.
- Mahmood, U., Fan, Y., Wei, S., Niu, Y., Li, Y., Huang, H., Chen, Y., Tang, Z., Liu, L., Qu, C., Zhang, K., Li, J., Lu, K. (2021). Comprehensive analysis of polygalacturonase genes offers new insights into their origin and functional evolution in land plants. *Genomics*. Volume 113, Issue 1, Part 2, Pages 1096-1108, ISSN 0888-7543, <https://doi.org/10.1016/j.ygeno.2020.11.006>.
- Makovics-Zsohár, N., Tóth, M., Surányi, D., Kovács, S., Hegedűs, A., & Halász, J. (2017). Simple Sequence Repeat Markers Reveal Hungarian Plum (*Prunus domestica* L.) Germplasm as a Valuable Gene Resource. *HortScience*, 52(12), 1655–1660. <https://doi.org/10.21273/HORTSCI12406-17>
- Mariette, S., Wong Jun Tai, F., Roch, G., Barre, A., Chague, A., Decroocq, S., Groppi, A., Laizet, Y., Lambert, P., Tricon, D., Nikolski, M., Audergon, J. M., Abbott, A. G., & Decroocq, V. (2016). Genome-wide association links candidate genes to resistance to

- Plum Pox Virus in apricot (*Prunus armeniaca*). *The New phytologist*, 209(2), 773–784. <https://doi.org/10.1111/nph.13627>
- Martínez-Gómez, P., Arulsekhar, S., Potter, D., & Gradziel, T. M. (2003). An extended interspecific gene pool available to peach and almond breeding as characterized using simple sequence repeat (SSR) markers. *Euphytica*, 131(3), 313–322.
- Martínez-Gómez, P., Crisosto, C. H., Bonghi, C., & Rubio, M. (2011). New approaches to *Prunus* transcriptome analysis. *Genetica*, 139(6), 755–769. <https://doi.org/10.1007/s10709-011-9580-2>
- Mcclintock, B., (1950). THE ORIGIN AND BEHAVIOR OF MUTABLE LOCI IN MAIZE.
- Meena, N.K., Choudhary, K., Negi, N., Meena, V.S., Gupta, V. (2021). Nutritional Composition of Stone Fruits. In: Mir, M.M., Iqbal, U., Mir, S.A. (eds) Production Technology of Stone Fruits. Springer, Singapore. https://doi.org/10.1007/978-981-15-8920-1_9
- Morata, J., Marín, F., Payet, J., Casacuberta, J.M., (2018). Plant lineage-specific amplification of transcription factor binding motifs by miniature inverted-repeat transposable elements (MITEs). *Genome Biology and Evolution* 10, 1210–1220. <https://doi.org/10.1093/gbe/evy073>
- Muñoz-sanz, J.V., Zuriaga, E., Badenes, M.L., Romero, C., (2017). RESEARCH PAPER A disulfide bond A-like oxidoreductase is a strong candidate gene for self-incompatibility in apricot (*Prunus armeniaca*) pollen 68, 5069–5078. <https://doi.org/10.1093/jxb/erx336>
- Nettancourt, D., (2001). Incompatibility and Incongruity in Wild and Cultivated Plants. <https://doi.org/10.1007/978-3-662-04502-2>
- Nilo-Poyanco, R., Vizoso, P., Sanhueza, D., Balic, I., Meneses, C., Orellana, A. & Campos-Vargas, R., (2019). A *Prunus persica* genome-wide RNA-seq approach uncovers major differences in the transcriptome among chilling-injury sensitive and non-sensitive varieties. *Physiologia Plantarum*, 166(3), pp.772–793. doi: 10.1111/pppl.12831.
- Niu, J., Zhu, B., Cai, J., Li, P., Wang, L., et al. (2014) ‘Selection of Reference Genes for Gene Expression Studies in Siberian Apricot (*Prunus sibirica* L.) Germplasm Using Quantitative Real-Time PCR’, *PLoS ONE*, 9(8), e103900. <https://doi.org/10.1371/journal.pone.0103900>
- Oki, N., Yano, K., Okumoto, Y., Tsukiyama, T., Teraishi, M., & Tanisaka, T. (2008). A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. Japonica. *Genes & genetic systems*, 83(4), 321–329. <https://doi.org/10.1266/ggs.83.321>
- Olden, E.J.; Nybom, N. (1968). On the origin of *Prunus cerasus* L. *Hereditas* , 59, 327–345.
- Palacioğlu, G., Alkan, M., Derviş, S., Bayraktar, H., & Özer, G. (2023). Molecular phylogeny of plant pathogenic fungi based on start codon targeted (SCoT) polymorphism. *Molecular biology reports*, 50(10), 8271–8279. <https://doi.org/10.1007/s11033-023-08735-4>
- Pan, L., Zeng, W., Niu, L., Lu, Z., Liu, H., Cui, G., Zhu, Y., Chu, J., Li, W., Fang, W., Cai, Z., Li, G., & Wang, Z. (2015). *PpYUC11*, a strong candidate gene for the stony hard phenotype in peach (*Prunus persica* L. Batsch), participates in IAA biosynthesis during

- fruit ripening. *Journal of experimental botany*, 66(22), 7031–7044.
<https://doi.org/10.1093/jxb/erv400> Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14(4):417–419. doi: 10.1038/nmeth.4197.
- Peace, C.P., Crisosto, C.H., Gradziel, T.M., (2005). Endopolygalacturonase: A candidate gene for Freestone and Melting Flesh in peach. *Molecular Breeding* 16, 21–31.
<https://doi.org/10.1007/s11032-005-0828-34>
- Pérez de los Cobos, F., García-Gómez, B. E., Orduña-Rubio, L., Batlle, I., Arús, P., Matus, J. T. & Eduardo, I. (2024) ‘Exploring large-scale gene coexpression networks in peach (*Prunus persica* L.): a new tool for predicting gene function’, *Horticulture Research*. Oxford University Press. doi:10.1093/hr/uhad294.
- Pérez de los Cobos, F., Kandylas, I., Zaracho, N., Batlle, I., Reig, G., Asín, L., Alegre, S., Howad, W., Aranzana, M.J., Arús, P., & Eduardo, I. (2024). Marker-assisted breeding in the IRTA peach and almond scion breeding programs. *Acta Horticulturae*, 1412, 291–298. DOI: 10.17660/ActaHortic.2024.1412.44
- Piegu, B., Guyot, R., Picault, N., Roulin, A., Sanyal, A., Kim, H., Collura, K., Brar, D. S., Jackson, S., Wing, R. A., & Panaud, O. (2006). Doubling genome size without polyploidization: dynamics of retrotransposon-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome research*, 16(10), 1262–1269.
<https://doi.org/10.1101/gr.5290206>
- Pinosio, S., Marroni, F., Zuccolo, A., Vitulo, N., Mariette, S., Sonnante, G., Aravanopoulos, F.A., Ganopoulos, I., Palasciano, M., Vidotto, M., Magris, G., Iezzoni, A., Vendramin, G.G. and Morgante, M. (2020), A draft genome of sweet cherry (*Prunus avium* L.) reveals genome-wide and local effects of domestication. *Plant J*, 103: 1420–1432. <https://doi.org/10.1111/tbj.14809>
- Potter, D. (2012). Basic information on the stone fruit crops. In *Genetics, genomics and breeding of stone fruits*; Kole, C., Abbott, A.G., Eds.; CRC Press: Boca Raton, ; pp. 1–21. <https://doi.org/10.1201/b13104>
- Pouget, M., Youssef, S., Dumas, P.J., Baumberger, T., San Roman, A., Torre, F., Affre, L., Médail, F., Baumel, A. (2016). Spatial mismatches between plant biodiversity facets and evolutionary legacy in the vicinity of a major Mediterranean city. *Ecol. Indic.* , 60, 736–745. <https://doi.org/10.1016/j.ecolind.2015.07.017>
- Poczai, P. et al. (2013) ‘Advances in plant gene-targeted and functional markers: a review’, *Plant Methods*, 9, 6. <https://doi.org/10.1186/1746-4811-9-6>
- Ono, K., Akagi, T., Morimoto, T., Wünsch, A., & Tao, R. (2018). Genome Re-Sequencing of Diverse Sweet Cherry (*Prunus avium*) Individuals Reveals a Modifier Gene Mutation Conferring Pollen-Part Self-Compatibility. *Plant & cell physiology*, 59(6), 1265–1275.
<https://doi.org/10.1093/pcp/pcy068>
- Ortuño-Hernández, G., Sánchez, M., Ruiz, D., Martínez-Gómez, P., & Salazar, J. A. (2024). Monitoring Fruit Growth and Development in Apricot (*Prunus armeniaca* L.)

- through Gene Expression Analysis. *International Journal of Molecular Sciences*, 25(16), 9081. <https://doi.org/10.3390/ijms25169081>
- Oshlack, A., Robinson, M. D., & Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome biology*, 11(12), 220. <https://doi.org/10.1186/gb-2010-11-12-220>.
- Qiu, Y., Köhler, C., (2020). Mobility connects: Transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochemical Society Transactions* 48, 1005–1017. <https://doi.org/10.1042/BST20190937>
- Rai M. K. (2023). Start codon targeted (SCoT) polymorphism marker in plant genome analysis: current status and prospects. *Planta*, 257(2), 34. <https://doi.org/10.1007/s00425-023-04067-6>
- Rehder, A. (1940). *Manual of cultivated trees and shrubs hardy in North America exclusive of the subtropical and warmer temperate regions*, 2nd ed.; MacMillan: New York, USA, pp. 1-996.
- Rohde, A., Bhalerao, R.P., (2007). Plant dormancy in the perennial context. *Trends Plant Sci.* 12, 217–223. <https://doi.org/10.1016/j.tplants.2007.03.012>
- Roberts, A., & Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature methods*, 10(1), 71–73. <https://doi.org/10.1038/nmeth.2251>
- Rodríguez-Robles, L., Devin, S. R., Ye, X., Sagbas, H. I., Mahdavi, S. M. E., Wettberg, E. B.-v., Feng, J., Rubio, M., & Martínez-Gómez, P. (2024). *Prunus* Movement Across the Silk Road: An Integrated Evolutionary and Breeding Analysis. *Horticulturae*, 10(12), 1381. <https://doi.org/10.3390/horticulturae10121381>
- Roy, N.S.; Choi, J.Y.; Lee, S.I.; Kim, N.S. (2015). Marker utility of transposable elements for plant genetics, breeding, and ecology: a review. *Genes Genom*, 37, 141-151.
- Rothkegel, K., Espinoza, A., Sanhueza, D., Lillo-Carmona, V., Riveros, A., Campos-Vargas, R., and Meneses, C. (2021) Identification of DNA Methylation and Transcriptomic Profiles Associated With Fruit Mealiness in *Prunus persica* (L.) Batsch. *Front. Plant Sci.* 12:684130. doi: 10.3389/fpls.2021.684130
- Rybin, W.A.(1936) Spontane und experimentell erzeugte Bastarde zwischen Schwarzdorn und Kirschkpflaume und das Abstammungsproblem der Kulturpflaume. *Planta* 25, 22–58. <https://doi.org/10.1007/BF01909303>
- Rushton, P.J., Somssich, I.E., Ringler, P., Shen, Q.J., (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. <https://doi.org/10.1016/j.tplants.2010.02.006>
- Sage, R.F., Christin, P., Edwards, E.J., 2011. The C 4 plant lineages of planet Earth 62, 3155–3169. <https://doi.org/10.1093/jxb/err048>
- Salazar, J. A., Ruiz, D., Zapata, P., Martínez-García, P. J., & Martínez-Gómez, P. (2022). Whole Transcriptome Analyses of Apricots and Japanese Plum Fruits after 1-MCP (Ethylene-Inhibitor) and Ethrel (Ethylene-Precursor) Treatments Reveal New Insights into the

- Physiology of the Ripening Process. *International Journal of Molecular Sciences*, 23(19), 11045. <https://doi.org/10.3390/ijms231911045>
- Shang, C., Cao, X., Tian, T., Hou, Q., Wen, Z., Qiao, G., & Wen, X. (2022). Cross-Talk between Transcriptome Analysis and Dynamic Changes of Carbohydrates Identifies Stage-Specific Genes during the Flower Bud Differentiation Process of Chinese Cherry (*Prunus pseudocerasus* L.). *International Journal of Molecular Sciences*, 23(24), 15562. <https://doi.org/10.3390/ijms232415562>
- Shen, Y., Ding, X., Fei, W., Cai, B., Gao, Z., Zhang, Z. (2011). Analysis of genetic diversity in Japanese apricot (*Prunus mume* Sieb. et Zucc.) based on REMAP and IRAP molecular markers. *Sci. Hortic.-Amsterdam*, 132, 50–58. <https://doi.org/10.1016/j.scienta.2011.10.005>
- Shi, S., Li, J., Sun, J., Yu, J., & Zhou, S. (2013). Phylogeny and classification of *Prunus sensu lato* (Rosaceae). *Journal of integrative plant biology*, 55(11), 1069–1079. <https://doi.org/10.1111/jipb.12095>
- Shi, T., Iqbal, S., Ayaz, A., Bai, Y., Pan, Z., Ni, X., Hayat, F., Saqib Bilal, M., Khuram Razzaq, M., & Gao, Z. (2020). Analyzing Differentially Expressed Genes and Pathways Associated with Pistil Abortion in Japanese Apricot via RNA-Seq. *Genes*, 11(9), 1079. <https://doi.org/10.3390/genes11091079>
- Šmarda, P., Bureš, P. (2012). The Variation of Base Composition in Plant Genomes. In: Wendel, J., Greilhuber, J., Dolezel, J., Leitch, I. (eds) *Plant Genome Diversity Volume 1*. Springer, Vienna. https://doi.org/10.1007/978-3-7091-1130-7_14
- Sonneveld, T., Tobutt, K.R., Vaughan, S.P., Robbins, T.P., (2005). Loss of Pollen- S Function in Two Self-Compatible Selections of *Prunus avium* Is Associated with Deletion / Mutation of an S Haplotype – Specific F-Box Gene 17, 37–51. <https://doi.org/10.1105/tpc.104.026963>
- Su, N., Hodel, R. G. J., Wang, X., Wang, J. R., Xie, S. Y., Gui, C. X., Zhang, L., Chang, Z. Y., Zhao, L., Potter, D., & Wen, J. (2023). Molecular phylogeny and inflorescence evolution of *Prunus* (Rosaceae) based on RAD-seq and genome skimming analyses. *Plant diversity*, 45(4), 397–408. <https://doi.org/10.1016/j.pld.2023.03.013>
- Surányi, D. Plum cultivars. In *Hungarian Fruit Cultivars*; Soltész, M., Ed. (2013).; Mezőgazda Kiadó: Budapest, Hungary, ; pp. 377–393.
- Surányi D. Nemzeti Szívügyünk III. A Pándy Meggy. *Tájökológiai Lapok*. 2016;14:135–154. doi: 10.56617/tl.3646.
- Tartarini, S., Tonutti, P., Arús, P., Orellana, A., Wells, C., Main, D., Vizzotto, G., Silva, H., Salamini, F., Schmutz, J., Morgante, M., Rokhsar, D.S., (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* 45, 487–494. <https://doi.org/10.1038/ng.2586>
- Tao, R., Watari, A., Hanada, T., Habu, T., Yaegaki, H., Yamaguchi, M., Yamane, H., (2007). Self-compatible peach (*Prunus persica*) has mutant versions of the S haplotypes found in self-

- incompatible *Prunus* species. *Plant Molecular Biology* 63, 109–123. <https://doi.org/10.1007/s11103-006-9076-0>
- Tatsuki, M., Soeno, K., Shimada, Y., Sawamura, Y., Suesada, Y., Yaegaki, H., Sato, A., (2018). Insertion of a transposon-like sequence in the 5' flanking region of the *YUCCA* gene causes the stony hard phenotype 815–827. <https://doi.org/10.1111/tpj.14070>
- Tian, F., Yang, D. C., Meng, Y. Q., Jin, J., & Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic acids research*, 48(D1), D1104–D1113. <https://doi.org/10.1093/nar/gkz1020>
- Thakur, M., Luharch, R., Sharma, V., Sharma, D.P. (2022). Molecular Markers Based Genetic Relatedness Studies in Tissue Culture Propagated Japanese Plum Cultivars Santa Rosa and Frontier. *Genet. Resour. Crop Evol.*, 69, 567–575. DOI: 10.21203/rs.3.rs-885836/v1.
- Tong, Z., Gao, Z., Wang, F., Zhou, J., Zhang, Z., (2009). Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol. Biol.* 10, 1–13. <https://doi.org/10.1186/1471-2199-10-71>
- Thodberg, S., Del Cueto, J., Mazzeo, R., Pavan, S., Lotti, C., Dicenta, F., Neilson, E.H.J., Møller, B.L., Sánchez-Pérez, R., (2018). Elucidation of the amygdalin pathway reveals the metabolic basis of bitter and sweet almonds (*Prunus dulcis*). *Plant Physiology* 178, 1096–1111. <https://doi.org/10.1104/pp.18.00922>
- Tong, Z., Gao, Z., Wang, F., Zhou, J., Zhang, Z., (2009). Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Mol. Biol.* 10, 1–13. <https://doi.org/10.1186/1471-2199-10-71>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5), 511–515. <https://doi.org/10.1038/nbt.1621>
- Tsukamoto, T., Hauck, N.R., Tao, R., Jiang, N., Iezzoni, A.F., (2006). Molecular characterization of three non-functional S-haplotypes in sour cherry (*Prunus cerasus*). *Plant Molecular Biology* 62, 371–383. <https://doi.org/10.1007/s11103-006-9026-x>
- Turro, E., Su, S. Y., Gonçalves, Â., Coin, L. J., Richardson, S., & Lewin, A. (2011). Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome biology*, 12(2), R13. <https://doi.org/10.1186/gb-2011-12-2-r13>
- Upton, R. N., Correr, F. H., Lile, J., Reynolds, G. L., Falaschi, K., Cook, J. P., & Lachowiec, J. (2023). *Design, execution, and interpretation of plant RNA-seq analyses*. *Frontiers in Plant Science*, 14, 1135455. <https://doi.org/10.3389/fpls.2023.1135455>.
- Van den Berge K, Hembach KM, Soneson C, Tiberi S, Clement L, Love MI, Robinson MD. (2019). RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annual Review of Biomedical Data Science* 2(1). doi:10.1146/annurev-biodatasci-072018-021255.
- Venkatesh and Nandini, B. (2020) 'Miniature inverted-repeat transposable elements (MITEs), derived insertional polymorphism as a tool of marker systems for molecular plant

- breeding', *Molecular Biology Reports*, 47, pp. 7393–7405.
<https://doi.org/10.1007/s11033-020-05365-y>
- Velasco, D., Hough, J., Aradhya, M., & Ross-Ibarra, J. (2016). Evolutionary Genomics of Peach and Almond Domestication. *G3 (Bethesda, Md.)*, 6(12), 3985–3993.
<https://doi.org/10.1534/g3.116.032672>
- Vendramin, E., Pea, G., Dondini, L., Pacheco, I., Dettori, M. T., Gazza, L., Scalabrin, S., Strozzi, F., Tartarini, S., Bassi, D., Verde, I., & Rossini, L. (2014). A unique mutation in a MYB gene cosegregates with the nectarine phenotype in peach. *PloS one*, 9(3), e90574.
<https://doi.org/10.1371/journal.pone.0090574>
- Vilanova, S., Badenes, M., Burgos, L., Martínez-Calvo, J., Llácer, G., Romero, C., (2006). Self-Compatibility of Two Apricot Selections Is Associated with Two Pollen-Part Mutations of Different Nature. *Plant physiology* 142, 629–641. <https://doi.org/10.1104/pp.106.083865>
- Villar L, Lienqueo I, Llanes A, Rojas P, Perez J, *et al.* (2020) Comparative transcriptomic analysis reveals novel roles of transcription factors and hormones during the flowering induction and floral bud differentiation in sweet cherry trees (*Prunus avium* L. cv. Bing). *PLOS ONE* 15(3): e0230110.
<https://doi.org/10.1371/journal.pone.0230110>
- Vimont, N., Fouché, M., Campoy, J.A., Tong, M., Arkoun, M., Yvin, J.-C., Wigge, P. and Dirlewanger, E., (2019). From bud formation to flowering: transcriptomic state defines the cherry developmental phases of sweet cherry bud dormancy. *BMC Genomics*, 20, p.974. <https://doi.org/10.1186/s12864-019-6348-z>
- Voytas, D.F., Naylor, G.J.P., (1998). Rapid flux in plant genomes. *Nature Genetics* 20, 6–7.
<https://doi.org/10.1038/1652>
- Walker, A.R., Lee, E., Bogs, J., McDavid, D.A.J., Thomas, M.R., Robinson, S.P., (2007). White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant Journal* 49, 772–785. <https://doi.org/10.1111/j.1365-313X.2006.02997.x>
- Wang, F., Tong, Z., Sun, J., Shen, Y., Zhou, J., Gao, Z., (2010). Genome-wide detection of Ty1-copia and Ty3-gypsy group retrotransposons in Japanese apricot (*Prunus mume* Sieb. et Zucc.) 9, 8583–8596. <https://doi.org/10.5897/AJB10.1259>
- Wang, J., Liu, W., Zhu, D., Zhou, X., Hong, P., (2020). A *de novo* assembly of the sweet cherry (*Prunus avium* cv. Tieton) genome using linked-read sequencing technology 1–18.
<https://doi.org/10.7717/peerj.9114>
- Wang, L., Peng, Q., Zhao, J., Ren, F., Zhou, H., (2016). Evolutionary origin of *Rosaceae*-specific active non-autonomous hAT elements and their contribution to gene regulation and genomic structural variation. *Plant Molecular Biology* 91, 179–191.
<https://doi.org/10.1007/s11033-016-0454-y>
- Wang, F., Tong, Z., Sun, J., Shen, Y., Zhou, J., Gao, Z., (2010). Genome-wide detection of Ty1-copia and Ty3-gypsy group retrotransposons in Japanese apricot (*Prunus mume* Sieb. et Zucc.) 9, 8583–8596. <https://doi.org/10.5897/AJB10.1259>
- Wang, J., Liu, W., Zhu, D., Zhou, X., Hong, P., (2020). A *de novo* assembly of the sweet cherry (*Prunus avium* cv. Tieton) genome using linked-read sequencing technology 1–18.
<https://doi.org/10.7717/peerj.9114>

- Wang, L., Peng, Q., Zhao, J., Ren, F., Zhou, H., (2016). Evolutionary origin of *Rosaceae*-specific active non-autonomous *hAT* elements and their contribution to gene regulation and genomic structural variation. *Plant Molecular Biology* 91, 179–191. <https://doi.org/10.1007/s11103-016-0454-y>
- Wang, D., Yu, C., Zuo, T., Zhang, J., Weber, D. F., & Peterson, T. (2015). Alternative Transposition Generates New Chimeric Genes and Segmental Duplications at the Maize p1 Locus. *Genetics*, 201(3), 925–935. <https://doi.org/10.1534/genetics.115.178210>
- Wang, G.; Weng, W.; Jia, Z.; Zhang, J.; Wang, T.; Xuan, J. (2022). Identification of candidate genes associated with pulp color by transcriptomic analysis of ‘Huaxiu’ Plum (*Prunus salicina* Lindl.) during fruit-ripening. *Curr. Issues Mol. Biol.* , 44, 6368- 6384. doi: 10.3390/cimb44120434
- Wang, G., Li, R., Wu, M. *et al.* Assessment of genetic diversity of *Prunus salicina* 'Shazikongxinli' by morphological traits and molecular markers. *Genet Resour Crop Evol* 70, 2727–2739 (2023). <https://doi.org/10.1007/s10722-023-01599-4>
- Wang, T., Hao, R., Pan, H., Cheng, T. and Zhang, Q. (2014) ‘Selection of Suitable Reference Genes for Quantitative Real-time Polymerase Chain Reaction in *Prunus mume* during Flowering Stages and under Different Abiotic Stress Conditions’, *Journal of the American Society for Horticultural Science*, 139(2), pp. 113–122. <https://doi.org/10.21273/JASHS.139.2.113>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). *RNA-Seq: a revolutionary tool for transcriptomics*. *Nature Reviews Genetics*, 10, 57–63. <https://doi.org/10.1038/nrg2484>.
- Wei, H., Chen, X., Zong, X., Shu, H., Gao, D. & Liu, Q., (2015). Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium*). *PLOS ONE*, 10(3), e0121164. doi: 10.1371/journal.pone.0121164.
- Wicker, T., Sabot, F., Aurélie Hua-Van, J.L.B., Pierre Capy, Boulos Chalhoub, Andrew Flavell, P.L., Michele Morgante, Olivier Panaud, Etienne Paux, P.S. and, Schulman, A.H., (2007). A unified classification system for eukaryotic transposable elements. <https://doi.org/10.1038/nrg2165>
- Witten.DM. (2011). "Classification and clustering of sequencing data using a Poisson model." *Ann. Appl. Stat.* 5 (4) 2493 – 2518, <https://doi.org/10.1214/11-AOAS493>
- Wöhner, T. W., Emeriewen, O. F., Wittenberg, A. H. J., Schneiders, H., Vrijenhoek, I., Halász, J., Hrotkó, K., Hoff, K. J., Gabriel, L., Lempe, J., Keilwagen, J., Berner, T., Schuster, M., Peil, A., Wünsche, J., Kropop, S., & Flachowsky, H. (2021). The draft chromosome-level genome assembly of tetraploid ground cherry (*Prunus fruticosa* Pall.) from long reads. *Genomics*, 113(6), 4173–4183. <https://doi.org/10.1016/j.ygeno.2021.11.002>
- Wu, B., Gao, L., Gao, J., Xu, Y., Liu, H., Cao, X., Zhang, B., & Chen, K. (2017). Genome-Wide Identification, Expression Patterns, and Functional Analysis of UDP Glycosyltransferase

- Family in Peach (*Prunus persica* L. Batsch). *Frontiers in plant science*, 8, 389. <https://doi.org/10.3389/fpls.2017.00389>
- Wu, T. D., & Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics (Oxford, England)*, 21(9), 1859–1875. <https://doi.org/10.1093/bioinformatics/bti310>
- Xiang-Ming Zhang (Zhang M), Cheng, W., Yuan, X., Wang, J., Cheng, T. & Zhang, Q., (2022). Integrated transcriptome and small RNA sequencing in revealing miRNA-mediated regulatory network of floral bud break in *Prunus mume*. *Frontiers in Plant Science*, 13, 931454. doi: 10.3389/fpls.2022.931454.
- Xiujun, W., Zhenqi, S., Yujing, T., Kaifeng, M. & Qingwei, L., (2022). Comparative transcriptome analysis linked to key volatiles reveals molecular mechanisms of aroma compound biosynthesis in *Prunus mume*. *BMC Plant Biology*, 22, article 395. doi: 10.1186/s12870-022-03779-3.
- Xie, X., Wu, N. (2002) Introns in higher plant genes. *Chin.Sci.Bull.* 47, 1409– 1415 . <https://doi.org/10.1360/02tb9311>
- Xin, Y., Ma, B., Xiang, Z., & He, N. (2019). Amplification of miniature inverted-repeat transposable elements and the associated impact on gene regulation and alternative splicing in mulberry (*Morus notabilis*). *Mobile DNA*, 10, 27. <https://doi.org/10.1186/s13100-019-0169-0>
- Xu, W., Bao, W., Liu, H., Chen, C., Bai, H., Huang, M. et al., (2022). Insights into the molecular mechanisms of late flowering in *Prunus sibirica* by whole-genome and transcriptome analyses. *Frontiers in Plant Science*, 12, 802827. doi: 10.3389/fpls.2021.802827
- Xue, S., Shi, T., Luo, W., Ni, X., Iqbal, S., Ni, Z., Huang, X., Yao, D., Shen, Z., Gao, Z., (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume* ., *Horticulture Research* 1–13. <https://doi.org/10.1038/s41438-019-0171-1>
- Ya-lin Zhao, Yong Li, Dan-dan Guo, Xue-jia Chen, Ke Cao, Jin-long Wu, Wei-chao Fang, Chang-wen Chen, Xin-wei Wang, Li-rong Wang (2025), Spatiotemporally transcriptomic analyses of floral buds reveal the high-resolution landscape of flower development and dormancy regulation in peach, *Horticulture Research*, Volume 12, Issue 5, uhaf029, <https://doi.org/10.1093/hr/uhaf029>
- Yamane, H., Ikeda, K., Hauck, N.R., Iezzoni, A.F., Tao, R., (2003). Self-incompatibility (S) locus region of the mutated S6- haplotype of sour cherry (*Prunus cerasus*) contains a functional pollen S allele and a non-functional pistil S allele. *Journal of Experimental Botany* 54, 2431–2437. <https://doi.org/10.1093/jxb/erg271>
- Yao, J., Dong, Y., & Morris, B. A. (2001). Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, 98(3), 1306–1311. <https://doi.org/10.1073/pnas.98.3.1306>
- Yeh, F.C.; Boyle, T.J.B. (1997). Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg. J. Bot.*, 129, 157-163.

- Yu, Z., Li, J., Wang, H., Ping, B., Li, X., Liu, Z., Guo, B., Yu, Q., Zou, Y., Sun, Y., Ma, F., & Zhao, T. (2024). Transposable elements in Rosaceae: insights into genome evolution, expression dynamics, and syntenic gene regulation. *Horticulture research*, 11(6), uhae118. <https://doi.org/10.1093/hr/uhae118>
- Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of computational biology : a journal of computational molecular cell biology*, 7(1-2), 203–214. <https://doi.org/10.1089/10665270050081478>
- Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., Tao, Y., Wang, J., Yuan, Z., Fan, G., Xing, Z., Han, C., Pan, H., Zhong, X., Shi, W., Liang, X., Du, D., Sun, F., Xu, Z., Hao, R., ... Wang, J. (2012). The genome of *Prunus mume*. *Nature communications*, 3, 1318. <https://doi.org/10.1038/ncomms2290>
- Zhang, M., Cheng, W., Yuan, X., Wang, J., Cheng, T., Zhang, Q. (2022). Integrated transcriptome and small RNA sequencing in revealing miRNA-mediated regulatory network of floral bud break in *Prunus mume*. *Front Plant Sci.* 2022;13:931454. doi:10.3389/fpls.2022.931454
- Zhang, X., Duan, X., Wang, J., Ran, J., Xue, Z., Zhang, X. et al., (2024). Combination of metabolome and transcriptome reveals flower color change candidate genes of *Prunus humilis*. *Scientia Horticulturae*, 336, 113364. doi: 10.1016/j.scienta.2024.113364.
- Zhao, S.P., Song, X.Y., Guo, L.L., Zhang, X.Z., Zheng, W.J., (2020). Genome-wide analysis of the shi-related sequence family and functional identification of Gmsrs18 involving in drought and salt stresses in soybean. *Int. J. Mol. Sci.*, 21, 1810. <https://doi.org/10.3390/ijms21051810>
- Zhebentyayeva, T.; Shankar, V.; Scorza, R.; Callahan, A.; Ravelonandro, M.; Castro, S.; DeJong, T.; Saski, C.A.; Dardick, C. (2019). Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. *Hortic. Res.* , 6, 12. <https://doi.org/10.1038/s41438-018-0090-6>
- Zhou, M., Tao, G., Pi, P., Zhu, Y., Bai, Y., & Meng, X. (2016). Genome-wide characterization and evolution analysis of miniature inverted-repeat transposable elements (MITEs) in moso bamboo (*Phyllostachys heterocycla*). *Planta*, 244(4), 775–787. <https://doi.org/10.1007/s00425-016-2544-0>
- Zhu, Y., Li, Y., Xin, D., Chen, W., Shao, X., Wang, Y. & Guo, W., (2015). RNA-Seq-based transcriptome analysis of dormant flower buds of Chinese cherry (*Prunus pseudocerasus*). *Gene*, 555(2), pp.362–376. doi: 10.1016/j.gene.2014.11.032.
- Zhu, H., Shi, Y., Zhang, J., Bao, M. & Zhang, J., (2022). Candidate genes screening based on phenotypic observation and transcriptome analysis for double flower of *Prunus mume*. *BMC Plant Biology*, 22, article 499. doi: 10.1186/s12870-022-03895-0.
- Zuriaga, E., Muñoz-Sanz, J. V., Molina, L., Gisbert, A.D., Badenes, M.L., Romero, C., (2013). An S-Locus Independent Pollen Factor Confers Self-Compatibility in “Katy” Apricot. *PLoS ONE* 8. <https://doi.org/10.1371/journal.pone.0053947>

Appendix A2. Bioinformatics Pipeline Script (from *REPORT.ipynb*)

This appendix provides the complete bioinformatics workflow used for the analysis of RNA-Seq data, as described in Section 4.7 (Analyses of RNA-Seq data) and corresponding Results Section 5.5 (Results for RNA-Seq analysis and read alignment).

The Jupyter notebook *REPORT.ipynb* documents the full computational pipeline, including all commands, parameters, and software versions used in the preprocessing and alignment of raw reads. Specifically, it includes:

- Quality control of raw reads using FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>);
- Adapter and quality trimming with Trimmomatic v1.0 (Bolger *et al.*, 2014);
- Read alignment to the haploid apricot reference genome (*Apricot Genome CUR v1*, cultivar ‘Rojo Pasión’; Campoy *et al.*, 2020) using *STAR* v2.6 (Dobin *et al.*, 2015);
- Alignment statistics summarizing mapping efficiency and read distribution, corresponding to Figure 8 in the Results chapter.

The notebook ensures reproducibility of all preprocessing steps and serves as the foundation for the gene expression quantification and statistical analyses described in Appendix A3.

...

Quality Control on Raw DATA

```
#!/usr/bin/env bash
mkdir -p ./raw_qc

#AM6
fastqc \
  -o ./raw_qc \
  -t 2 \
  ~/storage/rawdata/AM6_1.fq.gz \
  ~/storage/rawdata/AM6_2.fq.gz
```

```
#AMK
fastqc \
  -o ./raw_qc \
  -t 2 \
  ~/storage/rawdata/AMK_1.fq.gz \
  ~/storage/rawdata/AMK_2.fq.gz
```

```
#AHA
fastqc \
  -o ./raw_qc \
  -t 2 \
  ~/storage/rawdata/AHA_1.fq.gz \
  ~/storage/rawdata/AHA_2.fq.gz
```

```
#ARK
fastqc \
  -o ./raw_qc \
  -t 2 \
  ~/storage/rawdata/ARK_1.fq.gz \
  ~/storage/rawdata/ARK_2.fq.gz
```

```
#ACO
fastqc \
  -o ./raw_qc \
  -t 2 \
  ~/storage/rawdata/ACO_1.fq.gz \
  ~/storage/rawdata/ACO_2.fq.gz
```

```
#APA
fastqc \
  -o ./raw_qc \
  -t 2 \
  ~/storage/rawdata/APA_1.fq.gz \
```

```
~/storage/rawdata/APA_2.fq.gz
```

```
#SCO
```

```
fastqc \
```

```
-o ./raw_qc \
```

```
-t 2 \
```

```
~/storage/rawdata/SCO_1.fq.gz \
```

```
~/storage/rawdata/SCO_2.fq.gz
```

```
#SHA
```

```
fastqc \
```

```
-o ./raw_qc \
```

```
-t 2 \
```

```
~/storage/rawdata/SHA_1.fq.gz \
```

```
~/storage/rawdata/SHA_2.fq.gz
```

SortMeRNA – optional

```
#!/usr/bin/env bash
```

```
mkdir -p "$IDX_DIR"
```

```
# Define your reference sequence paths on the HPC server
```

```
REF_18S="/home/Ivanovska/storage/silva-euk-18s-id95.fasta"
```

```
REF_28S="/home/Ivanovska/storage/silva-euk-28s-id98.fasta"
```

```
IDX_DIR="/home/Ivanovska/storage/run/idx/kvdb"
```

```
for SAMPLE in ACO AHA ARK APA AM6 AMK SCO SHA SMK SPA SM6 SRK  
do
```

```
sortmerna \
```

```
--fastx \
```

```
--other ~/storage/rawdata/${SAMPLE}_sortmerna \
```

```
--out2 \
```

```
--paired_in \
```

```
--reads ~/storage/rawdata/${SAMPLE}_1.fq.gz \
```

```
--reads ~/storage/rawdata/${SAMPLE}_2.fq.gz \
```

```
--ref ${REF_18S} \
```

```
--ref ${REF_28S} \
```

```
--idx ${IDX_DIR} \
```

```
--threads 28 \
--kvdb kvdb_${SAMPLE} |& tee sortmerna_${SAMPLE}.log # key
value database
done
```

```
#SORTMERNA_DB =
/data/rRNA_databases/silva_ids_acc_tax.tar.gz/sortmerna/
```

Quality Control after SortMeRNA (if used)

```
#!/usr/bin/env bash

mkdir -p ./sortmerna_qc

#for SAMPLE in AHA ACO APA AM6 AMK ARK SCO SHA SM6 SMK SPA SRK
#do

fastqc \
-o ./sortmerna_qc \
-t 2 \
~/storage/rawdata/AHA_sortmerna.fq
```

Trimmomatic

```
#!/usr/bin/env bash

TRIMMOMATIC_JAR="/home/Ivanovska/.conda/envs/trimmomatic_env/trimming/Trimmomatic-0.39/trimmomatic-0.39.jar"

for SAMPLE in ACO AHA ARK APA AM6 AMK SCO SHA SMK SPA SM6 SRK
do
java -jar ${TRIMMOMATIC_JAR} PE \
-threads 28 \
~/storage/rawdata/${SAMPLE}_1.fq.gz \
~/storage/rawdata/${SAMPLE}_2.fq.gz \
./${SAMPLE}_trimmomatic_1.fastq.gz \
./${SAMPLE}_trimmomatic_1u.fastq.gz \
./${SAMPLE}_trimmomatic_2.fastq.gz \
./${SAMPLE}_trimmomatic_2u.fastq.gz \
```

```

ILLUMINACLIP:/home/Ivanovska/.conda/envs/trimmomatic_env/trimmin
g/Trimmomatic-0.39/adapters/TruSeq2-PE.fa:2:30:10 \
    SLIDINGWINDOW:5:5 \
    MINLEN:50
done

```

Quality Control after Trimmomatic

```

#!/usr/bin/env bash

mkdir -p ./trimm_qc
for SAMPLE in ACO AHA ARK APA AM6 AMK SCO SHA SMK SPA SM6 SRK
do
fastqc \
    -o ./trimm_qc \
    -t 2 \
    ~/work/rna/${SAMPLE}_trimmomatic_1.fastq.gz \
    ~/work/rna/${SAMPLE}_trimmomatic_2.fastq.gz
done

```

Multi Quality Control Report

```

#!/usr/bin/env bash

mkdir -p multiqc

multiqc \
    -o multiqc \
    .

```

STAR -Genome aligner

```

#!/usr/bin/env bash

STAR --runThreadN 28 \

```

```

--runMode genomeGenerate \
--genomeDir ~/storage/genome \
--genomeFastaFiles ~/storage/genome/parmeniaca.fasta \
--sjdbGTFfile ~/storage/genome/parmeniaca.gtf \
--sjdbOverhang 149

#!/usr/bin/env bash

for SAMPLE in ACO AHA APA AM6 AMK ARK SCO SHA SM6 SMK SPA SRK
do
STAR --runThreadN 12 --runMode alignReads --genomeDir
~/storage/genome --readFilesCommand zcat \
--readFilesIn ~/work/rna/${SAMPLE}_trimmomatic_1.fastq.gz
~/work/rna/${SAMPLE}_trimmomatic_2.fastq.gz \
--outSAMtype BAM SortedByCoordinate \
--quantMode GeneCounts --outFileNamePrefix
~/storage/star/${SAMPLE}aligned"
done

# FeatureCounts
#!/usr/bin/env bash
for SAMPLE in APA AMK ACO AHA AM6 ARK SCO SHA SM6 SMK SPA SRK
do
featureCounts -a /home/Ivanovska/storage/genome/apricot.gtf \
-o ${SAMPLE}_counts.txt \
-t exon \
-g gene_id \
-M \
-p \
-T 28 \

/home/Ivanovska/storage/star/${SAMPLE}alignedAligned.sortedByCoo
rd.out.bam # Add your BAM files here

done

```

Appendix A3. R Script for RNA-Seq Analysis

This appendix presents the R script used for gene-level quantification, normalization, and differential expression analysis of RNA-Seq data, as detailed in Sections 4.7.2–4.7.4 (Quantification, Differential Expression, Functional Annotation) and visualized in Results Section 5.5 (RNA-Seq analysis and read alignment).

The script, implemented primarily with the DESeq2 R package (Love *et al.*, 2014), performs the following major analytical steps:

- Import and normalization of count data generated by *featureCounts* v2.0.8 (Liao *et al.*, 2014);
- Variance-stabilizing transformations and visualization of sample relationships (PCA, sample distances; Figures 11–12);
- Differential expression testing using a negative binomial model with Benjamini–Hochberg correction ($p_{adj} \leq 0.05$ threshold);
- Generation of plots and summary tables (DEG lists) presented in Section 5.5.6;
- Export of DEG results for downstream annotation and discussion in Section 5.5.7.

All scripts are fully annotated and include R package versions and execution instructions to allow full reproducibility of the analyses presented in the Results and Discussion.

1. Load Required Libraries

```
# Suppress startup messages, load all needed packages...
suppressPackageStartupMessages({
  library(DESeq2)
  library(edgeR)
  library(here)
  ...
})
```

2. Define Directories & List BAM Files

```
# Set 'dir' to the folder containing aligned BAM files
dir <- "~/RNASewq/rojofc/bam"
list.files(dir)
```

3. Read Sample Metadata & Verify Filenames

```
# Read metadata CSV, set rownames, build full paths to BAMs,
check existence
datadir <- here("~/RNASewq/salmodata")
meta <- read.csv(file.path(datadir, "meta_anther.csv"))
rownames(meta) <- meta$names
filenames <- file.path(dir, paste0(meta$names,
"aligned18Aligned.sortedByCoord.out.bam"))
file.exists(filenames)
```

4. Load Count Data from featureCounts

```
# Read featureCounts output and summary statistics
fc <- read.csv("~/RNASewq/rojofc/merger_anther.csv",
row.names = 1)
stat <-
read.csv("/home/ljupco/RNASEWQ/rojofc/merged_18counts_summary.csv")
names(fc)
```

5. Plot Summary Statistics as Barplot

```
# Define color palette, convert 'stat' to matrix, and draw
barplot
stats <- stat %>%
column_to_rownames("Status") %>%
as.matrix()
# Define a color palette similar to the heatmap
pal <- colorRampPalette(c("blue", "white", "red"))(3)
```

```

# Adjust margins to make space for the legend below
par(mar = c(7, 4, 4, 2) + 0.1)
X11()

# Plot the barplot with the new palette and legend below
barplot(stats[rowSums(stats) > 0, ],
        beside = TRUE,
        col = pal,
        las = 2,
        cex.axis = 0.8,
        cex.names = 0.8,
        legend.text = rownames(stats)[rowSums(stats) > 0],
        args.legend = list(
            horiz = TRUE,
            cex = 0.8,
            x = "bottom",
            inset = c(0, -0.3), # move legend below the plot
            xpd = TRUE,        # allow drawing outside the
plot region
            bty = "n"
        ))

```

6. Prepare Metadata & Create DESeq2 Dataset

```

# Coerce condition to factor, filter low-count genes, run
DESeq2
dds <- DESeqDataSetFromMatrix(countData = fc, colData = meta,
design = ~ condition)
dds <- dds[rowSums(counts(dds)) > 1, ]
dds <- DESeq(dds)
res <- results(dds)

```

7. Transformation & Visualization

```

# rlog transformation, sample-distance heatmap, PCA, MDS,
Poisson distance

```

```

rld <- rlog(dds, blind = FALSE)

sampleDists <- dist(t(assay(rld)))

sampleDistMatrix <- as.matrix(sampleDists)

rownames(sampleDistMatrix) <- paste(rld$condition, rld$names,
sep = "-")

# rld should already exist:
# rld <- rlog(dds, blind = FALSE)

# Extract PCA data
pcaData <- plotPCA(rld, intgroup = c("condition", "tissue"),
returnData = TRUE)

# Add sample names from metadata (meta$names must match column
order of rld)
pcaData$sample <- colData(rld)$names

# Percent variance for axis labels
percentVar <- round(100 * attr(pcaData, "percentVar"))

# Color palette
pal <- brewer.pal(min(9, length(unique(pcaData$condition))),
"Set1")

# PCA plot with labels
ggplot(pcaData, aes(PC1, PC2, color = condition, shape =
tissue)) +
  geom_point(size = 4, alpha = 0.9) +
  geom_text_repel(aes(label = sample),
    size = 4,
    box.padding = 0.4,
    point.padding = 0.3,

```

```

        segment.size = 0.25,
        max.overlaps = Inf) +
  scale_color_manual(values = pal) +
  coord_fixed() +
  theme_bw(base_size = 13) +
  theme(panel.grid = element_blank(),
        legend.position = "right",
        plot.title = element_text(hjust = 0.5)) +
  xlab(paste0("PC1: ", percentVar[1], "% variance")) +
  ylab(paste0("PC2: ", percentVar[2], "% variance")) +
  ggtitle("PCA of rlog-transformed counts – Anther Samples")

```

```

  poisd <- PoissonDistance(t(counts(dds)))
  samplePoisDistMatrix <- as.matrix(poisd$dd)
  pheatmap(samplePoisDistMatrix, clustering_distance_rows =
poisd$dd, ...)
  mdsPoisData <- data.frame(cmdscale(samplePoisDistMatrix))
  mdsPois <- cbind(mdsPoisData, as.data.frame(colData(dds)))
  ggplot(mdsPois, aes(X1, X2, color = condition)) +
  geom_point(size = 3)

```

8. Differential Expression Results Exploration

```

  # Summarize res, apply padj < 0.05, log2FC thresholds, list
top genes
  summary(res)
  res.05 <- results(dds, alpha = 0.05)
  table(res.05$padj < 0.05)
  resLFC1 <- results(dds, lfcThreshold = 1)
  table(resLFC1$padj < 0.05)
  res.SC_vs_SI <- results(dds, contrast = c("condition", "SC",
"SI"))
  summary(res.SC_vs_SI)

```

```

sum(res$pvalue < 0.05, na.rm = TRUE)
sum(res$padj < 0.05, na.rm = TRUE)
resSig <- subset(res, padj < 0.05)
head(resSig[order(resSig$log2FoldChange), ])
head(resSig[order(resSig$log2FoldChange, decreasing = TRUE),
])

```

9. Plot Counts for the Top Gene

```

# Get gene with smallest padj, plot counts (scatter, dot,
line) on log scale

topGene <- rownames(res)[which.min(res$padj)]

data_top <- plotCounts(dds, gene = topGene, intgroup =
c("condition"), returnData = TRUE)

ggplot(data_top, aes(condition, count, color = condition)) +
scale_y_log10() + ...

ggplot(data_top, aes(condition, count, fill = condition)) +
scale_y_log10() + ...

ggplot(data_top, aes(condition, count, color = condition,
group = condition)) + scale_y_log10() + ...

```

10. Gene Clustering & Heatmap for Top Variable Genes

```

# Identify top 50 variable genes by rowVars, center rows,
annotate by condition

topVarGenes <- head(order(rowVars(assay(rld))), decreasing =
TRUE), 50)

mat <- assay(rld)[topVarGenes, ]
mat <- mat - rowMeans(mat)

df <- as.data.frame(colData(rld)[, c("condition")])

rownames(df) <- colnames(rld)

pheatmap(mat, annotation_col = df)

```

11. Export Results & Heatmap Data

```

# Order results by padj, output top 100 to a data frame,
prepare top 50 heatmap matrix

```

```

resOrdered <- res[order(res$padj), ]
resOrderedDF <- as.data.frame(resOrdered)[1:100, ]
topSigGenes <- head(order(res$padj), 50)
mat_sig <- assay(rld)[topSigGenes, ]
mat_sig <- mat_sig - rowMeans(mat_sig)
pheatmap(mat_sig, annotation_col = df)

# Optionally save CSVs for downstream analysis:
# write.csv(resOrderedDF, "DESeq2_top100_results.csv")
# write.csv(mat_sig, "heatmap_top50_matrix.csv")

# Select top 20 genes by padj
top20Genes <- head(order(res$padj), 20)
mat <- assay(rld)[top20Genes, ]

# Center rows for relative expression
mat <- mat - rowMeans(mat)

# Prepare column annotation
df <- as.data.frame(colData(rld)[, "condition", drop = FALSE])
rownames(df) <- colnames(rld)

# Generate heatmap
pheatmap(
  mat,
  annotation_col = df,
  color = colorRampPalette(c("blue", "white", "red"))(100),
  main = "Top 20 Genes by Adjusted P-value"
)

```

Appendix A4. PCR Amplification Patterns

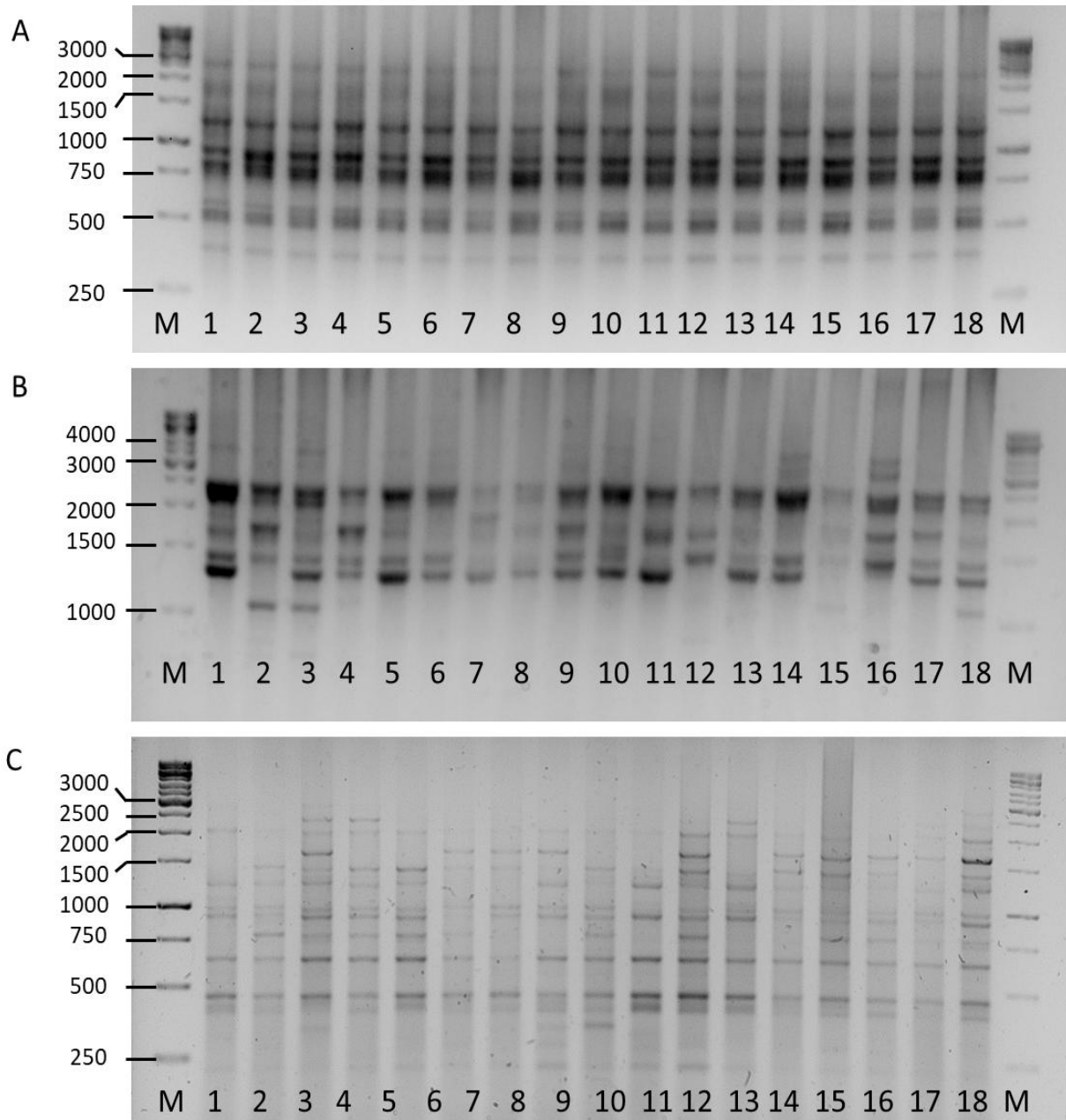


Figure S1: Patterns of amplified fragments in PCR of European plum (*Prunus domestica*) cultivars using the SCoT1 (A), FaSt-R (B) and a combination of SCoT1 and FaSt-R (C) primers. Labels are the followings: (M) GeneRuler 1 kb DNA ladder, (1) ‘Bühler Frühzwetschge’, (2) ‘Czar’, (3) ‘Grand Prize’, (4) ‘Fehérszilva’, (5) ‘Páczelt szilvája’, (6) Soviet 3, (7) ‘Tuleu timpuriu’, (8) ‘Vengerka virla’, (9) ‘Vörös nagygyümölcsű’, (10) ‘Besztercei Bt.2’, (11) ‘Sanctus Hubertus’, (12) ‘Wagenheim’, (13) ‘Giant’, (14) Soviet 2, (15) ‘Duránci’, (16) ‘Opal’, (17) ‘Haganta’ and (18) ‘Presenta’ (Ivanovska *et al.*, 2025).

Appendix A5. Sequence Data

Table 7. Accession numbers and labels of the sequenced amplicons from the SCoT-FaSt-R PCR assays with amplicon size, parameters of BlastN homology searches and identified regions (Ivanovska *et al.*, 2025).

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
SCoT1-FaSt-R, <i>Prunus domestica</i> cultivar ‘Topfive’							
PQ869779 <i>Pdom 1-TF-907</i>	907	P. mume polygalacturonase ADPG1 (LOC103339337), mRNA	XM_008242608.2	77	0.0	96.13	5’ UTR <1..293 <i>FaSt</i> <1..211 exon 294..>907
SCoT1-FaSt-R, <i>Prunus domestica</i> cultivar ‘Besztercei Bt.2’							
PQ869780 <i>Pdom_BB_1_908</i>	908	P. mume polygalacturonase ADPG1 (LOC103339337), mRNA	XM_008242608.2	77	0.0	96.14	5’ UTR <1..293 <i>FaSt</i> <1..211 exon 294..>908
SCoT2-FaSt-R, <i>Prunus cerasus</i> cultivar ‘Kántorjánosi 3’							
PQ869781 <i>Pcer_KJ_2_1227</i>	1,227	<i>P. avium beta-D-glucosyl crocetin beta-1,6-glucosyltransferase-like (LOC110763544), mRNA</i>	XM_021966336	100	0.0	99.92	exon <1..1101 3’ UTR 1102..>1227
SCoT2-FaSt-R, <i>Prunus cerasus</i> cultivar ‘Feketicsi meggy’							
PQ869782 <i>Pcer_FM_2_1227</i>	1,227	<i>P. avium beta-D-glucosyl crocetin beta-1,6-glucosyltransferase-like (LOC110763544), mRNA</i>	XM_021966336.1	100	0.0	99.84	exon <1..1101 3’ UTR 1102..>1227

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
SCoT2-FaSt-R, <i>Prunus cerasus</i> cultivar 'Kántorjánosi 3'							
PQ869783 <i>Pcer_KJ_2_343</i>	343	P. armeniaca genome assembly, chromosome: 1	OZ046360.2	48	1e-55	91.62	<i>FaSt</i> <1..>166 Intergenic 167.. >343
SCoT2-FaSt-R, <i>Prunus cerasus</i> cultivar 'Feketicsi meggy'							
PQ869784 <i>Pcer_FM_2_462</i>	462	P. armeniaca genome assembly, chromosome: 2	OZ046361.2	100	0.0	96.98	Intergenic <1..263 <i>FaSt</i> <264..>462
SCoT3-FaSt-R, <i>Prunus dulcis</i> cultivar 'Tuono'							
PQ869785 <i>Pdul_TU_3_544</i>	544	<i>P. dulcis</i> inositol 2- dehydrogenase/D-chiro-inositol 3- dehydrogenase (LOC117628739)	XM_034361251.1	48	4e- 132	100	5' UTR <1..444 <i>FaSt</i> <1..211 exon 445..>544
SCoT3-FaSt-R, <i>Prunus avium</i> cultivar 'Canada giant'							
PQ869786 <i>Pav_CG_3_910</i>	910	P. armeniaca genome assembly, chromosome: 8	OZ046367.2	56	6e- 153	94.93	<i>FaSt</i> <1..193 Intergenic 194.. >910
SCoT4-FaSt-R, <i>Prunus dulcis</i> cultivar 'Tétényi keményhéjú'							
PQ869787 <i>Pdul_TK_4_761</i>	761	<i>Prunus persica</i> uncharacterized LOC18771123 (LOC18771123), transcript variant X2, mRNA	OZ022203.1	48	2e- 157	94.57	5' UTR <1..210 exon 211..>368 intron 369.. >761 <i>FaSt</i> 550.. .>761

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
SCoT4-FaSt-R, <i>Prunus cerasus</i> cultivar 'Kántorjánosi 3'							
PQ869788 <i>Pcer_KJ_4_662</i>	662	P. avium peroxidase N1-like (LOC110751195), mRNA	XM_021951630.1	77	0.0	94.45	exon <1..>343 intron 344.. >662
SCoT4-FaSt-R, <i>Prunus cerasus</i> cultivar 'Kántorjánosi 3'							
PQ869789 <i>Pcer_KJ_4_965</i>	965	Similar to <i>At1g77260</i> Probable methyltransferase <i>PMT10</i> (<i>Arabidopsis thaliana</i> OX=3702)	Pcer_083304-RA	100	0.0	99.00	exon <1..555 3' UTR 556..>965
SCoT4-FaSt-R, <i>Prunus cerasus</i> cultivar 'Kántorjánosi 3'							
PQ869790 <i>Pcer_KJ_4_1000</i>	1,000	<i>P. avium</i> BTB/POZ domain- containing protein <i>At4g08455-like</i> (LOC110766395), mRNA	XM_021969719.1	13	6e-59	100.00	<i>FaSt</i> <1..201 exon <871..>1000 intron <1.. >870
SCoT4-FaSt-R, <i>Prunus cerasus</i> cultivar 'Feketicsi meggy'							
PQ869791 <i>Pcer_FM_4_1056</i>	1,056	P. dulcis U1 spliceosomal RNA (LOC117616929), ncRNA	XR_004584266.1	15	8e-68	96.91	intergenic <1..137 snRNA 138..299 intergenic 300..>1056 <i>FaSt</i> <862..>1056
SCoT4-FaSt-R, <i>Prunus cerasus</i> cultivar 'Feketicsi meggy'							
PQ869792 <i>Pcer_FM_4_496</i>	496	P. avium pectinesterase (LOC110749177), mRNA	XM_021949211.1	23	3e-49	100.00	<i>FaSt</i> <1..202 exon <384..>496 intron <1..383

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
SCoT4-FaSt-R, <i>Prunus domestica</i> cultivar 'Elena'							
PQ869793 <i>Pdom_EL_4_433</i>	433	P. armeniaca genome assembly, chromosome: 8	OZ046367.2	100	0.0	96.31	Intergenic <1..>220 <i>FaSt</i> <221..>433
SCoT4-FaSt-R, <i>Prunus domestica</i> cultivar 'Elena'							
PQ869794 <i>Pdom_EL_4_766</i>	766	P. mume 7-deoxyloganetin glucosyltransferase-like (LOC103319662), mRNA	XM_008221234.2	100	0.0	98.56	exon <1..711 3' UTR 712..>766
SCoT16-FaSt-R, <i>Prunus persica</i> cultivar 'Collins'							
PQ869795 <i>Pper_CO_16_454</i>	454	P. armeniaca genome assembly, chromosome: 2	OZ046361.2	99	1e-87	92.44	intergenic <1..249 <i>FaSt</i> 250..>454
SCoT19-FaSt-R, <i>Prunus armeniaca</i> cultivar 'Goldrich'							
PQ869796 <i>Parm_GR_19_251</i>	251	P. armeniaca genome assembly, chromosome: 4	OZ046363.2	100	2e-107	95.62	intergenic <1..43 <i>FaSt</i> 44..>251
SCoT21-FaSt-R, <i>Prunus domestica</i> cultivar 'Toptaste'							
PQ869810 <i>Pdom_TT_21_269</i>	269	<i>P. persica</i> WUSCHEL-related homeobox 9 (LOC18778807), transcript variant X2, mRNA	XM_020562726.1	48	5e-54	97.69	exon <1..>128 intron 129..>269

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
SCoT22-FaSt-R, <i>Prunus dulcis</i> cultivar ‘Tétényi keményhájú’							
PQ869797 <i>Pdul_TK_22_1866</i>	1,866	P. dulcis genome assembly, chromosome: 3	OZ022199.1	100	0.0	95.17	<i>FaSt</i> <1..211 intergenic 212.. >1866
SCoT22-FaSt-R, <i>Prunus dulcis</i> cultivar ‘Tétényi keményhájú’							
PQ869798 <i>Pdul_TK_22_497</i>	497	P. dulcis genome assembly, chromosome: 3	OZ022199.1	100	9e- 139	98.60	intergenic <1..211 <i>FaSt</i> 287..>497
SCoT22-FaSt-R, <i>Prunus dulcis</i> cultivar ‘Tétényi keményhájú’							
PQ869799 <i>Pdul_TK_22_901</i>	901	<i>P. dulcis</i> DEAD-box ATP-dependent RNA helicase 52C-like (LOC117633584), mRNA	XM_034367242.1	37	3e- 171	100.00	Intron <1..>569 <i>FaSt</i> <1..>189 exon <570..>901
SCoT22-FaSt-R, <i>Prunus dulcis</i> cultivar ‘Tétényi keményhájú’							
PQ869800 <i>Pdul_TK_22_263</i>	263	P. dulcis genome assembly, chromosome: 1	OZ022197.1	100	6e- 133	100.00	<i>FaSt</i> <1..>178 intergenic 179.. >263
SCoT22-FaSt-R, <i>Prunus persica</i> cultivar ‘Collins’							
PQ869801 <i>Pper_CO_22_644</i>	644	P. dulcis genome assembly, chromosome: 1	OZ022197.1	100	0.0	95.49	intergenic <1..432 <i>FaSt</i> 433..>644

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
SCoT22-FaSt-R, <i>Prunus domestica</i> cultivar ‘Nemtudom P3’							
PQ869802 <i>Pdom_NT_22_873</i>	873	<i>P. dulcis</i> ethylene-responsive transcription factor ERF053-like (LOC117623168), mRNA	XM_034354046.1	100	0.0	97.48	exon <1..390 3’ UTR 391..>873 <i>FaSt</i> <815..>873
SCoT22-FaSt-R, <i>Prunus domestica</i> cultivar ‘Nemtudom P3’							
PQ869803 <i>Pdom_NT_22_334</i>	334	<i>P. dulcis</i> glycosyltransferase BC10-like (LOC117629155), mRNA	XM_034361660.1	17	7e-14	94.83	intron <1..277 278.. >334 <i>FaSt</i> <1..221
SCoT32-FaSt-R, <i>Prunus domestica</i> cultivar ‘Toptaste’							
PQ869804 <i>Pdom_TT_32_418</i>	418	<i>P. armeniaca</i> genome assembly, chromosome: 1	OZ046360.2	100	2e-95	97.64	<i>FaSt</i> <1..212 intergenic 213.. >418
SCoT32-FaSt-R, <i>Prunus domestica</i> cultivar ‘Haroma’							
PQ869805 <i>Pdom_HR_32_551</i>	551	<i>P. armeniaca</i> genome assembly, chromosome: 3	OZ046362.2	100	3e-178	88.27	<i>FaSt</i> <1..204 intergenic 205.. >551
SCoT32-FaSt-R, <i>Prunus domestica</i> cultivar ‘Haroma’							
PQ869806 <i>Pdom_HR_32_1224</i>	1,224	<i>P. mume</i> beta-D-glucosyl crocetin beta-1,6-glucosyltransferase-like (LOC103331154), mRNA	XM_008233760.1	100	0.0	97.14	exon <1..1098 3' UTR 1099..>1224 <i>FaSt</i> 1154..>1224
SCoT33-FaSt-R, <i>Prunus dulcis</i> cultivar ‘Tuono’							

Accession Nu.	Size bp	BLAST homolog	Accession Nu. of BLAST hit	Coverage %	E value	Identity %	Regions (bp)
PQ869807 <i>Pdul_TU_33_590</i>	590	P. dulcis genome assembly, chromosome: 5	OZ022201.1	100	0.0	99.74	Intergenic <1..377 <i>FaSt</i> 378..>590
SCoT33-FaSt-R, <i>Prunus dulcis</i> cultivar 'Tuono'							
PQ869808 <i>Pdul_TU_33_342</i>	342	P. dulcis genome assembly, chromosome: 6	OZ022202.1	50	9e-23	94.81	<i>FaSt</i> <1..>75 Intergenic 76..>342
SCoT33-FaSt-R, <i>Prunus domestica</i> cultivar 'Haroma'							
PQ869809 <i>Pdom_HR_33_1460</i>	1,460	<i>P. dulcis leaf rust 10 disease-</i> <i>resistance locus receptor-like</i> <i>protein kinase-like 1.2</i> (LOC117613970), transcript variant X1, mRNA	XM_034342614.1	44	0.0	96.11	intron <1..338 exon 339..977 intron 978.. >1460 <i>FaSt</i> <1243..>1460

Appendix A6. Declaration of Students and Doctoral Candidates on the Use of Artificial Intelligence (AI)”

1. general information:

Name of the student: Beti Ivanovska

Neptun ID: AM4U5W

Level of program (mark with X): ☐ BSc/BA ☐ MSc/MA ☒ Doctoral School (PhD) ☐ Other:

.....

Title of the work: **Molecular characterization of the *Prunus Falling Stones* transposon and its application in marker development**

2. Declaration on the Use of AI

I, the undersigned, fully aware of my ethical responsibility, make the following declaration:

(Please choose one of the options below!)

☐ A) I have not used any artificial intelligence system or service. (If you selected this option, completing the subsequent tables is not required.)

☒ B) I have used an artificial intelligence system or service. (Please fill in the relevant tables!)

3. Details of Artificial Intelligence Usage

TABLE I: Assistant or Minor Usage (e.g., translation, language proofreading, brainstorming, etc.)

(For these uses, attaching the specific prompts and responses is not required.)

Purpose of Use	Name and Version of the AI Tool Used	Affected Section (if not applicable to the entire text)
Language proofreading	Microsoft Copilot (GPT-4, OpenAI)	
Extraction and clarification of abbreviations and technical terms	ChatGPT (GPT-5, OpenAI)	Methods, Results, and Abbreviations sections
Formatting and summarization of supporting material	ChatGPT (GPT-5, OpenAI)	Appendices and supplementary notes
Language editing, style refinement, and figure caption integration (merging and rephrasing of existing text)	ChatGPT (GPT-5, OpenAI)	Figure captions and related descriptions in Results section
Language refinement and minor stylistic proofreading	ChatGPT (GPT-5, OpenAI)	Results and Discussion

5. Declaration Applicable to All Students:

I declare that I have critically reviewed, edited, and incorporated any content potentially generated by AI in all cases. I take full responsibility for every element of the submitted work, including its originality and scientific validity. I acknowledge that the Hungarian University of Agriculture and Life Sciences may check the submitted work with an artificial intelligence detector and may initiate proceedings if my declaration is found to be false or incomplete.

Place and Date: Budapest, 2026. April 20

Blanovska

.....
Signature of the Student

.....
Signature of the Advisor/Supervisor

11. ACKNOWLEDGMENT

I would like to express my deepest gratitude to everyone who has supported me throughout this journey.

First of all, I am sincerely thankful to the members of my thesis committee for their insightful comments and encouragement, and for the questions that inspired me to explore my research from broader perspectives. Your guidance has truly shaped both this work and my growth as a researcher.

I am profoundly indebted to my supervisor at the Department of Plant Biotechnology, Institute of Genetics and Biotechnology, Hungarian University of Agriculture and Life Sciences (MATE) Prof. Dr. Attila Hegedűs and also for Prof. Dr. Júlia Halász for their unwavering guidance, thoughtful feedback, and constant encouragement. Your patience in answering my countless questions, your insightful critiques, and your steadfast belief in my potential have been the cornerstone of this work and a source of continuous inspiration.

Many thanks to Dr. Zsuzsanna Benyóné György and Dr. Iman Mirmazloun for their invaluable advice, and my laboratory colleagues: Aaqib Javid, Luiz Fernando Gonçalves Zanfelicci, Thanyarat Onlamun, and Pamulaklakin Eusebio Geronima, for their generous support, collaboration, and friendship. I am grateful to the departmental staff, Melisa Chikhi and Rita Hermán, for keeping our work running smoothly.

Special thanks go to Dr. Zoltán Havelda and Dr. Éva Várallyay for their expert help with RNA isolation and RNA-Seq preparation, which was crucial for this research.

This work was supported by the Flagship Research Groups Programme of MATE and the NKFI K_128874 project. I am thankful to MATE for the doctoral scholarship and to the Tempus Public Foundation for their financial support, which made this research possible.

I am especially grateful to Ana for her encouragement and steadfast support, and to all my friends and colleagues whose kindness and companionship sustained me along the way.

Finally, my deepest thanks go to my family. Your love, patience, and unwavering belief in me have been my greatest source of strength.