Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative Biochemical and Genetic Analysis of Basil (Ocimum basilicum L.) Genotypes: Uncovering SNP Variations and Metabolic Pathway Links

Essam Y. Abdul-Hafeez
Qassim University
Khalaf R. Al-Rashidi
Qassim University
Mohamed I. Motawei
mtaoa@qu.edu.sa

Qassim University

Research Article

Keywords: Ocimum basilicum, SSR markers, SNP markers, essential oil, phenolic acids, flavonoids, genetic diversity

Posted Date: April 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9322850/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Integrative Biochemical and Genetic Analysis of Basil (*Ocimum basilicum* L.) Genotypes: Uncovering SNP Variations and Metabolic Pathway Links

Essam Y. Abdul-Hafeez, Khalaf R. Al-Rashidi, Mohamed I. Motawei*

Department of Plant Production, College of Agriculture and Food, Qassim University, PO Box 6622, Buraidah 51452, Qassim, Saudi Arabia

*Corresponding author, mtaoa@qu.edu.sa

Abstract

Basil (*Ocimum basilicum*) exhibits remarkable biochemical and genetic diversity, yet the integration of genotype-specific metabolic profiles with molecular markers remains underexplored. This study characterizes three representative genotypes—Genovese, French, and Purple—using comprehensive biochemical assays, SSR and SNP markers, and multivariate analyses. Genovese basil showed the highest essential oil content (2.30%) and elevated flavonoids (rutin 40.0 µg mL^{-1}, luteolin 30.0 µg mL^{-1}), French basil accumulated caffeic (76.3 µg mL^{-1}) and benzoic acids (46.2 µg mL^{-1}), and Purple basil contained maximal apigenin (49.0 µg mL^{-1}) and hydroxycinnamic acids (coumaric 44.0 µg/mL, ferulic 23.5 µg mL^{-1}), revealing metabolic trade-offs between terpenoid and phenylpropanoid pathways. Pearson correlation and PCA confirmed coordinated regulation of metabolites and clearly separated genotypes based on biochemical specialization. SSR markers revealed 70% polymorphism, clustering French and Genovese together, while Purple formed a distinct lineage. Targeted SNP analysis identified genotype-specific nucleotide variations within amplified genomic regions, providing additional molecular signatures for genotype discrimination. The integration of biochemical traits with SNP-based clustering provides strong evidence that metabolic variation in basil is genetically structured. The identification of metabolite clusters linked to specific SNP groups offers crucial insights into the genetic basis of secondary metabolism, paving the way for enhanced breeding strategies aimed at optimizing aromatic plant traits.

Keywords: *Ocimum basilicum*, SSR markers, SNP markers, essential oil, phenolic acids, flavonoids, genetic diversity.

1. Introduction

Basil (*Ocimum basilicum* L.) is one of the most economically important culinary and medicinal herbs in the Lamiaceae family, cultivated globally for its essential oils, aromatic leaves, and antioxidant properties (Makri & Kintzios, 2008). The genus *Ocimum* exhibits remarkable genetic and chemotypic diversity, with different varieties accumulating distinct profiles of terpenoids, phenolic acids, and flavonoids that determine their industrial and culinary applications (Carović-Stanko et al., 2010; Kwee & Galaffu, 2011).

Among the numerous cultivated varieties, three distinct morphotypes represent the major axes of basil diversity: green-leaf sweet basils (typified by French and Genovese types) and pigmented varieties (represented by Purple basil). Genovese basil is renowned for its high essential oil content and is the preferred cultivar for pesto production, while French basil is valued for its balanced phenolic profile and culinary versatility. Purple basil, distinguished by anthocyanin-mediated pigmentation, accumulates distinct flavonoid profiles with enhanced antioxidant capacity (Flanigan & Niemeyer, 2014). Despite these well-established phenotypic differences, the genetic basis underlying metabolic specialization remains incompletely characterized.

Simple sequence repeat (SSR) markers have emerged as powerful tools for genetic diversity assessment and marker-assisted selection in basil breeding programs (Khosla et al., 2019; Kumar et al., 2018). SSR markers targeting dinucleotide repeats, particularly (GA) motifs, have been associated with regulatory regions of genes involved in terpenoid biosynthesis, including those encoding terpene synthases (Mahajan et al., 2014). Similarly, expressed sequence tag (EST)-derived SSR markers (OBSSR series) provide direct links to functional genes governing secondary metabolism (Mahajan et al., 2014).

Single nucleotide polymorphisms (SNPs) represent the most abundant class of genetic variation and offer higher resolution for genotype discrimination and association studies (Yang et al., 2017). SNPs located within or near genes encoding key biosynthetic enzymes can directly influence metabolic phenotypes by affecting enzyme activity, substrate specificity, or transcriptional regulation (Borevitz et al., 2000). In aromatic

plants, SNPs in phenylpropanoid and terpenoid pathway genes have been linked to variation in secondary metabolite accumulation (Vogt, 2010; Zhou & Pichersky, 2020). Previous studies have demonstrated that essential oil accumulation in basil is correlated with peltate glandular trichome density and the expression of genes in the terpenoid biosynthesis pathway (Deschamps et al., 2006). However, the specific SSR and SNP markers that reliably predict high-oil phenotypes and phenolic profiles in breeding populations remain to be systematically validated. Furthermore, the relationship between molecular polymorphisms and the accumulation of individual phenolic acids (coumaric, ferulic, syringic, caffeic, benzoic) and flavonoids (rutin, quercetin, luteolin, apigenin, catechin) has not been comprehensively examined across contrasting basil chemotypes. The objectives of this study were to: (1) characterize SSR polymorphism among Purple, French, and Genovese basil genotypes; (2) identify and analyze SNP variation across genotypes; (3) quantify essential oil content and profile phenolic and flavonoid composition via HPLC; and (4) investigate the association between SNP variants and biosynthetic pathway.

2. Materials and Methods

2.1 Plant Material

Three internationally recognized basil (*Ocimum basilicum* L.) genotypes were selected for this study: Purple basil (pigmented morphotype, high anthocyanin content), French basil (green-leaf culinary type, intermediate oil content), and Genovese basil (green-leaf industrial type, high oil content). Plants were cultivated under standardized greenhouse conditions (25 ± 2°C, 16 h photoperiod, relative humidity 60–70%) at the experimental station. Fresh leaf tissue was harvested at the pre-flowering stage for DNA extraction and biochemical analyses.

2.3 Essential Oil Extraction and Quantification

Fresh leaf material (100 g per genotype) was subjected to hydro distillation for 3 h using a Clevenger-type apparatus. Essential oil content was expressed as a percentage (v/w) on a fresh weight basis. Oil samples were dried over anhydrous sodium sulfate and stored at 4°C until analysis.

2.4 HPLC Analysis of Phenolic Acids and Flavonoids

Sample preparation followed the protocol of Kwee and Galaffu (2011) with modifications. Freeze-dried leaf tissue (1 g) was extracted with 80% aqueous methanol (10 mL) containing 1% hydrochloric acid under sonication for 30 min. Extracts were centrifuged at 10,000 rpm for 15 min, and supernatants were filtered through 0.45 μm membrane filters prior to HPLC injection.

HPLC analysis was performed on a Shimadzu LC-20AT system equipped with a photodiode array detector. Separation was achieved on a C18 reverse-phase column (250 × 4.6 mm, 5 μm particle size) maintained at 30°C. The mobile phase consisted of (A) 0.1% phosphoric acid in water and (B) methanol, with a gradient elution: 0–5 min, 20% B; 5–20 min, 20–60% B; 20–25 min, 60–80% B; 25–30 min, 80% B; followed by re-equilibration. Flow rate was maintained at 1.0 mL/min, and injection volume was 20 μL. Detection wavelengths were set at 280 nm for phenolic acids and 360 nm for flavonoids.

Compounds were identified by comparing retention times and UV spectra with authentic standards. Quantification was performed using external calibration curves for each standard. Phenolic acids quantified included coumaric acid, ferulic acid, syringic acid, caffeic acid, and benzoic acid. Flavonoids quantified included rutin, quercetin, luteolin, apigenin, and catechin.

2.5 DNA Extraction

Total genomic DNA was extracted from young leaf tissue (100 mg) using the CTAB method with minor modifications. DNA quality and concentration were assessed using spectrophotometry (Nano-Drop 2000, Thermo Scientific) and agarose gel electrophoresis (0.8% w/v). DNA samples were diluted to working concentrations of 50 ng/μl for subsequent analyses.

2.6 SSR Marker Analysis

A panel of 15 SSR primers (Table 1) was employed for genetic characterization, including primers targeting dinucleotide (GA and AG) repeats, trinucleotide repeats

(OBSSR series), and previously characterized *Ocimum*-specific markers (Kumar et al. 2018; Pyne et al., 2018; Mahajan et al., 2014).

Table 1: The primer`s name and its annealing temperature of the SSR Markers used for *O. basillium* genotyping

NO	Primer name	Sequence	Tm(°C)
1	$(AG)_{10}$	F: 5'AGGATCAGTAGTAGTAGTGC-3'	53.6
		R: 5'-TCCGATGATGATGATGCTGA-3'	57.9
2	$(GA)_{12}$	F: 5'-CAGGAGGAGGAGGAAGGAGG-3'	67.8
		R: 5'-TGCTCTGCTCTGCTCTGCTC-3'	63.5
3	$(ATG)_8$	F:5'-ATGATGATGATGATGCGC-3'	53.3
		R: 5'-TCGTCGTCGTCGTCTGGT-3'	63.6
4	$(CAG)_{10}$	F: 5'-CAGCAGCAGCAGCAGCATG-3'	63
		R: 5'-GCTGCTGCTGCTGCTGACG-3'	65.4
5	$(GAA)_{15}$	F: 5'-GAAGAAGAAGAAGAAGTGC-3'	52.5
		R: 5'-CTCTCTCTCTCTCTCGGA-3'	58.1
6	OBSSR1	F: 5'-TGGGTGTTGTGTTGTGTTGTG-3'	60.8
		R: 5'-AACGAAGGAAGGAAGGAAGG-3'	60.7
7	OBSSR2	F:5'-AGGAGGAGGAGGAGGAGA-3'	64.9
		R:5'-CGGCTTCTTCTTCTTCTTCT-3'	56.7
8	OBSSR3	F:5'-GCTGCTGCTGCTGCTGCT-3'	64.3
		R:5'-ACGACGACGACGACGACG-3'	64.5
9	OBSSR4	F:5'-ATGATGATGATGATGATG-3'	46.6
		R:5'-TATCTATCTATCTATCTA-3'	36.4
10	OBSSR5	F:5'-CTCTCTCTCTCTCTCTCTC-3'	55.5
		R:5'-GAGAGAGAGAGAGAGAGA-3'	54.2
11	OBSSR6	F:5'-TGAAGCTGAGGTTTGGGAGT-3'	61.6
		R:5'-CTGGAAGCAGTGAGTCCAGT-3'	62
12	OBSSR7	F:5'-AGCTTGGCAAGGTGTTTGTT-3'	59.2
		R:5'-ACGAGTCTTGTTGGCAGGTC-3'	62.3
13	OBSSR8	F:5'-CAGGATCCTTGCCTTGAGAT-3'	59.7

		R:5'-GTGAGAGGTCGGAAGAGAGG-3'	63.5
14	OBSSR9	F:5'-GACCCAGTGGATCTGGTAGG-3'	63.9
		R:5'-TTGCTGGTAGGCAGGAGTAG-3'	61.2
15	OBSSR10	F:5'-TCTGGAAGGTTCCAGGATGT-3'	61.7
		R:5'-CAGCAGCAGCAGCAGCAG-3'	63.2

PCR amplifications were performed in 20 µL reactions containing 50 ng template DNA, 1× PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.4 µM each primer, and 1 U Taq DNA polymerase. Thermal cycling conditions consisted of initial denaturation at 94°C for 5 min; 35 cycles of 94°C for 30 s, annealing at 50–55°C (marker-dependent) for 45 s, and 72°C for 1 min; and final extension at 72°C for 10 min.

Amplification products were resolved on 2.5% agarose gels stained with ethidium bromide and visualized under UV light. SSR loci were scored as binary data based on presence (1) or absence (0) of clear, reproducible bands. Only polymorphic markers producing unambiguous amplification patterns were retained for downstream analysis.

2.7 SNP Analysis

SNP discovery was performed through comparative sequence analysis of PCR products of SSR primers (OBSSR1 and $(GA)_{12}$) across the three genotypes (Pyne et al., 2018). PCR amplification of target regions was followed by Sanger sequencing (Macrogen, Inc.). Sequences were aligned using Clustal W in MEGA X software (Kumar et al., 2018).

2.8 Genetic Similarity and Phylogenetic Analysis

Genetic similarity among genotypes was estimated using the Dice similarity coefficient (Nei and Li coefficient), which emphasizes shared band presence and is considered appropriate for SSRs.

Cluster analysis was performed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm implemented in MEGA X software. The resulting dendrogram was visualized to depict genetic relationships among the three genotypes.

A character-by-taxon matrix was constructed from the high-quality SNPs identified across the three basil genotypes (French, Genovese, and Purple). Pairwise genetic distances were calculated using the p-distance model, which measures the proportion of nucleotide differences between sequences, and Edwards' angular distance, which corrects for multiple substitutions to provide a more accurate representation of genetic divergence (Nei and Kumar, 2000). Phylogenetic relationships were reconstructed using Maximum Likelihood (ML) analysis in MEGA X (Kumar et al., 2018). The final phylogenetic tree was visualized to examine the hierarchical divergence patterns among the three basil genotypes.

2.9 SNP–Biochemical Association and Heatmap Analysis

Pearson correlation coefficients (r) were calculated between SNP markers and biochemical traits, including essential oil content and phenolic compounds. The resulting correlation matrix was visualized as a heatmap using hierarchical clustering (Euclidean distance, complete linkage). Color gradients ranged from −1 (strong negative correlation) to +1 (strong positive correlation). Both SNPs and traits were clustered to identify co-regulated modules and potential marker–trait associations. A Pearson correlation coefficient (r) heat-map was constructed using XLSTAT v. 2019.

2.10 Statistical analysis.

All biochemical data were expressed as mean ± standard deviation (SD) of three biological replicates. Statistical significance among genotypes was determined using one-way ANOVA, followed by Tukey's HSD test ($p \leq 0.05$) for multiple comparisons. Normality and homogeneity of variance were verified prior to analysis using Shapiro–Wilk and Levene's tests, respectively. Statistical analyses were performed using SPSS (v.26) and R software.

Principal Component Analysis (PCA) was conducted to visualize the distribution of genotypes in multivariate space based on combined biochemical trait data. PCA was performed using the factoextra package in R, with data centered and scaled before analysis.

3. Results

3.1 Biochemical Composition of Basil Genotypes

Essential oil content and the concentrations of phenolic acids and flavonoids differed substantially among the three basil genotypes (Table 2). Genovese basil exhibited the highest essential oil content (2.30%), followed by French basil (1.10%) and Purple basil (0.60%). This threefold variation in oil content represents the most pronounced quantitative difference among the genotypes.

Profiles of phenolic acids showed accumulation patterns that are specific to each genotype. French basil had the highest amounts of caffeic acid (76.3 μg mL^{-1}) and benzoic acid (46.2 μg mL^{-1}), significantly surpassing the concentrations found in Purple and Genovese basil. Conversely, Purple basil contained the highest levels of coumaric acid (44.0 μg mL^{-1}), and French basil had 42.5 μg mL^{-1}, whereas Genovese basil showed significantly lower coumaric acid levels (25.0 μg mL^{-1}). Ferulic acid exhibited slight variation, with Purple basil (23.5 μg mL^{-1}) marginally surpassing French (19.76 μg mL^{-1}) and Genovese (20.0 μg mL^{-1}). Syringic acid showed a consistent range across genotypes (22.8–25.0 μg mL^{-1}).

The composition of flavonoids also showed unique signatures based on genotype. Genovese basil was characterized by high levels of rutin (40.0 μg mL^{-1}) and luteolin (30.0 μg mL^{-1}), around 2.5 times greater than Purple basil (16.0 μg mL^{-1} for both compounds) and 2.6–3.0 times greater than French basil (13.29 μg mL^{-1} rutin, 11.43 μg mL^{-1} luteolin). Apigenin was found to be highest in Purple basil (49.0 μg mL^{-1}) and French basil (48.82 μg mL^{-1}), while Genovese basil exhibited lower accumulation (35.0 μg mL^{-1}). Quercetin levels were similar across genotypes (40.0–47.86 μg mL^{-1}), with French basil showing the highest level. Catechin displayed slight fluctuation (22.0–25.16 μg mL^{-1}).

Table 2. Oil content (%) and phenolic/flavonoid (μg mL^{-1}) composition of three basil genotypes

Trait (unit)	Purple basil	French basil	Genovese basil
Oil (%)	0.60 ± 0.05 c	1.10 ± 0.08 b	2.30 ± 0.12 a
Coumaric acid	44.00 ± 2.10 a	42.50 ± 1.80 a	25.00 ± 1.50 b
Ferulic acid	23.50 ± 1.20 a	19.76 ± 1.10 b	20.00 ± 1.00 b
Syringic acid	23.00 ± 1.30 b	22.80 ± 1.40 b	25.00 ± 1.60 a
Caffeic acid	45.50 ± 2.50 b	76.30 ± 3.20 a	30.00 ± 1.90 c
Benzoic acid	33.30 ± 1.70 b	46.20 ± 2.10 a	18.00 ± 1.20 c
Rutin	16.00 ± 1.10 b	13.29 ± 0.90 b	40.00 ± 2.30 a
Quercetin	40.50 ± 2.00 b	47.86 ± 2.40 a	40.00 ± 2.10 b
Luteolin	16.00 ± 1.00 b	11.43 ± 0.80 c	30.00 ± 1.80 a
Apigenin	49.00 ± 2.20 a	48.82 ± 2.30 a	35.00 ± 1.70 b
Catechin	22.00 ± 1.40 b	25.16 ± 1.50 a	25.00 ± 1.50 a

3.2 Correlation Analysis of Biochemical Traits

Pearson correlation coefficients revealed several strong positive and negative relationships among biochemical traits (Figure 1). The almost perfect positive correlations between rutin and syringic acid (r = 1.00) and between apigenin and coumaric acid (r = 1.00) indicate that these pairs of compounds may be co-regulated, possibly via common biosynthetic routes or transcriptional regulation mechanisms. The robust positive correlation between caffeic acid and quercetin (r = 0.96) suggests a linked accumulation of this phenolic acid–flavonoid combination, particularly noticeable in French basil.

The content of essential oil showed strong positive correlations with rutin (r = 0.93) and luteolin (r = 0.86), suggesting that genotypes chosen for elevated oil yield (Genovese) also accumulate these flavonoids. In contrast, oil content exhibited significant negative correlations with coumaric acid (r = -0.98) and apigenin (r = -0.97), indicating a metabolic trade-off between terpenoid (oil) and phenylpropanoid (phenolic acid/flavone) production. The negative correlation between rutin and coumaric acid (r = -0.99) underscores this metabolic division.

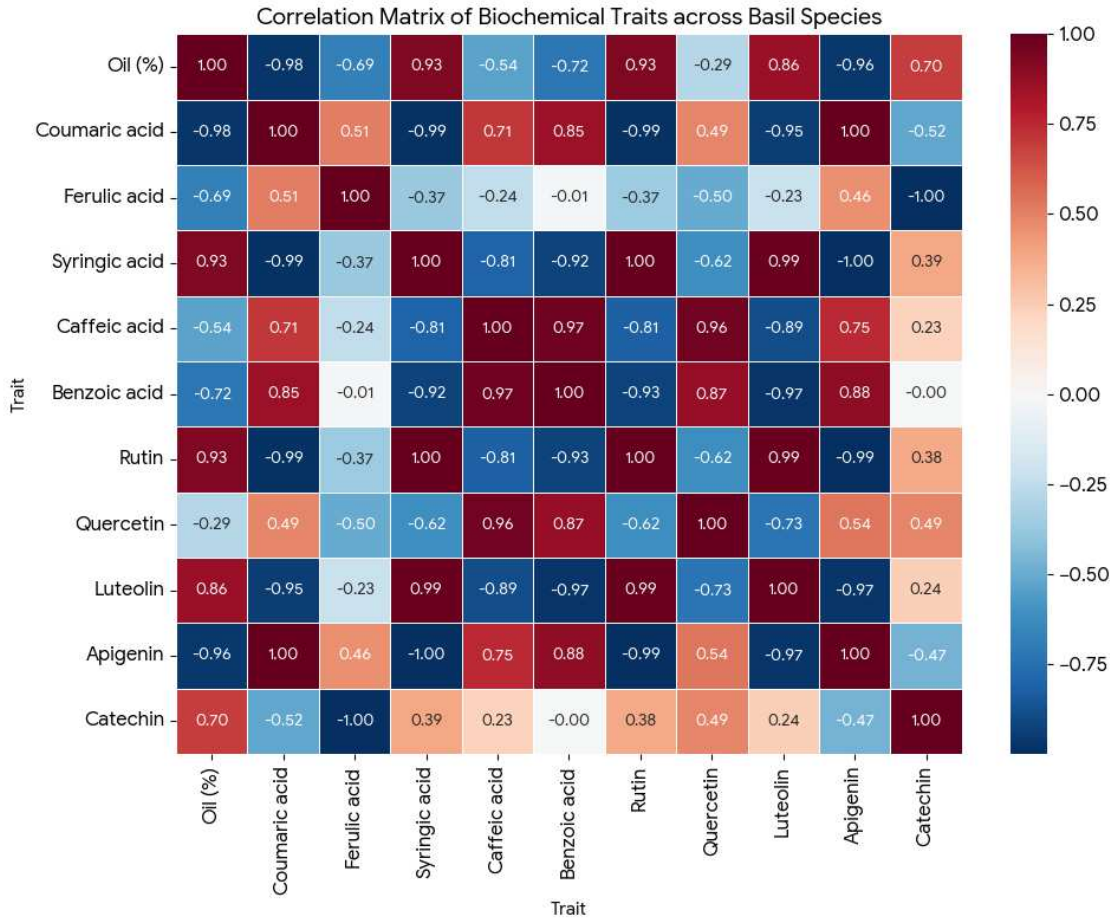


Figure 1. Pearson correlation coefficients ((r) for selected biochemical trait pairs

3.3 Principal Component Analysis

Principal Component Analysis (PCA) was performed to assess the variation among Purple, French, and Genovese basil concerning eleven biochemical traits, such as essential oil content and phenolic compounds (Figure 2). The initial two principal

components (PC1 and PC2) accounted for 100% of the total variance, where PC1 comprised 72.7%, and PC2 encompassed 27.3%, suggesting that most of the biochemical variation could be depicted in two dimensions. PC1 mainly distinguished Genovese basil from the other genotypes. This divide was primarily influenced by elevated levels of oil (2.30%), rutin (40.0%), and luteolin (30.0%), which exhibited significant positive loadings on PC1. Conversely, Purple and French basil exhibited reduced levels of oil and flavonoids, positioning them on the negative side of PC1. PC2 identified French basil, noted for its elevated levels of caffeic acid (76.3%) and benzoic acid (46.2%), in contrast to Purple basil, which exhibited moderate amounts of these phenolic acids. Genovese basil was positioned moderately on PC2 because of its average amount of phenolic acids. Therefore, Genovese basil is characterized by elevated levels of flavonoids and oils, French basil by increased phenolic acids, and Purple basil shows moderate characteristics, emphasizing the biochemical variety among these basil genotypes.

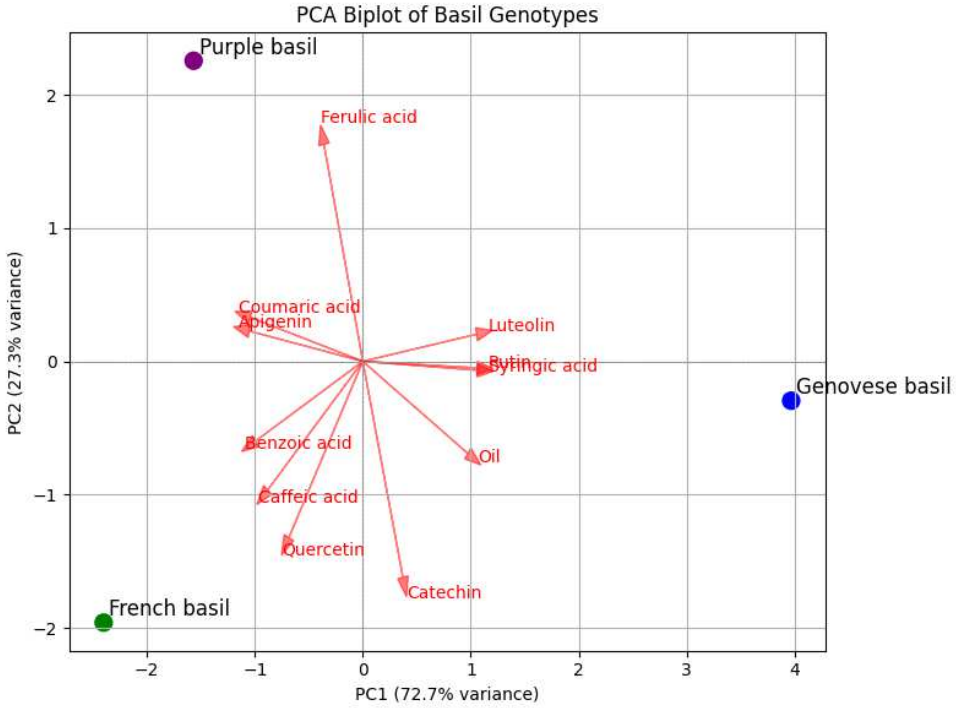


Figure 2. Principal Component Analysis (PCA) of three basil genotypes based on eleven biochemical traits, including essential oil content and phenolic compounds. PC1 (72.7% of variance) separates Genovese basil (blue) due to high oil, rutin, and luteolin contents, while PC2 (27.3% of variance) distinguishes French basil (green) with high caffeic and benzoic acids. Purple basil (purple) occupies an intermediate position. Red arrows represent trait loadings and their contribution to the principal components.

3.4 SSR Marker Polymorphism

Fifteen SSR primers were utilized to assess the genetic variation among the three basil genotypes (Purple, French, and Genovese basil). The primers produced a total of 40 evaluable bands, with 28 being polymorphic, leading to an overall polymorphism percentage of 70% (Table 3). The quantity of amplified bands for each primer varied between 2 and 3 bands, with a mean of 2.67 bands for each primer.

The quantity of polymorphic bands ranged from 1 to 2 for each primer, reflecting a moderate degree of genetic variability among the genotypes examined. The polymorphism information content (PIC) values varied from 0.26 to 0.36, with an average of 0.31, indicating that the SSR markers utilized in this study were moderately effective for identifying genetic diversity in basil.

Of the primers, OBSSR3, OBSSR8, and OBSSR1 displayed comparatively higher PIC values, suggesting increased discriminatory ability, while OBSSR4 and $(ATG)_8$ showed reduced PIC values. The SSR analysis showed significant genetic variation among the basil genotypes, highlighting the effectiveness of these markers for evaluating genetic diversity and distinguishing genotypes.

Table 3. SSR primer characteristics and marker statistics for basil genotypes

No.	Primer name	Total bands	Polymorphic bands	Polymorphism (%)	PIC
1	$(AG)_{10}$	3	2	66.7	0.31
2	$(GA)_{12}$	3	2	66.7	0.32
3	$(ATG)_8$	2	1	50.0	0.27
4	$(CAG)_{10}$	3	2	66.7	0.33
5	$(GAA)_{15}$	2	1	50.0	0.28
6	OBSSR1	3	2	66.7	0.35
7	OBSSR2	3	2	66.7	0.34
8	OBSSR3	3	2	66.7	0.36

No.	Primer name	Total bands	Polymorphic bands	Polymorphism (%)	PIC
9	OBSSR4	2	1	50.0	0.26
10	OBSSR5	3	2	66.7	0.32
11	OBSSR6	3	2	66.7	0.33
12	OBSSR7	3	2	66.7	0.34
13	OBSSR8	3	2	66.7	0.35
14	OBSSR9	2	1	50.0	0.29
15	OBSSR10	2	1	50.0	0.30

3.5 Genetic Similarity and Phylogenetic Relationships

Dice similarity coefficients calculated from the SSR binary matrix revealed moderate genetic relatedness among the three genotypes (Table 4). The highest similarity was observed between French basil and Genovese basil (0.667), indicating shared allelic composition at approximately two-thirds of the scored loci. Purple basil showed lower similarity with both French basil (0.651) and Genovese basil (0.632), consistent with its distinct morphological and biochemical characteristics.

Table 4. Dice genetic similarity matrix for three basil genotypes.

Genotype	Purple basil	French basil	Genovese basil
Purple basil	1.000	0.651	0.632
French basil	0.651	1.000	0.667
Genovese basil	0.632	0.667	1.000

UPGMA cluster analysis based on Dice coefficients resolved the genotypes into two primary clusters (Figure 3). French basil and Genovese basil are grouped together in a

major cluster, reflecting their shared genetic background as sweet basil morphotypes. Purple basil formed a distinct, independent cluster, confirming its genetic divergence from the green-leaf varieties.

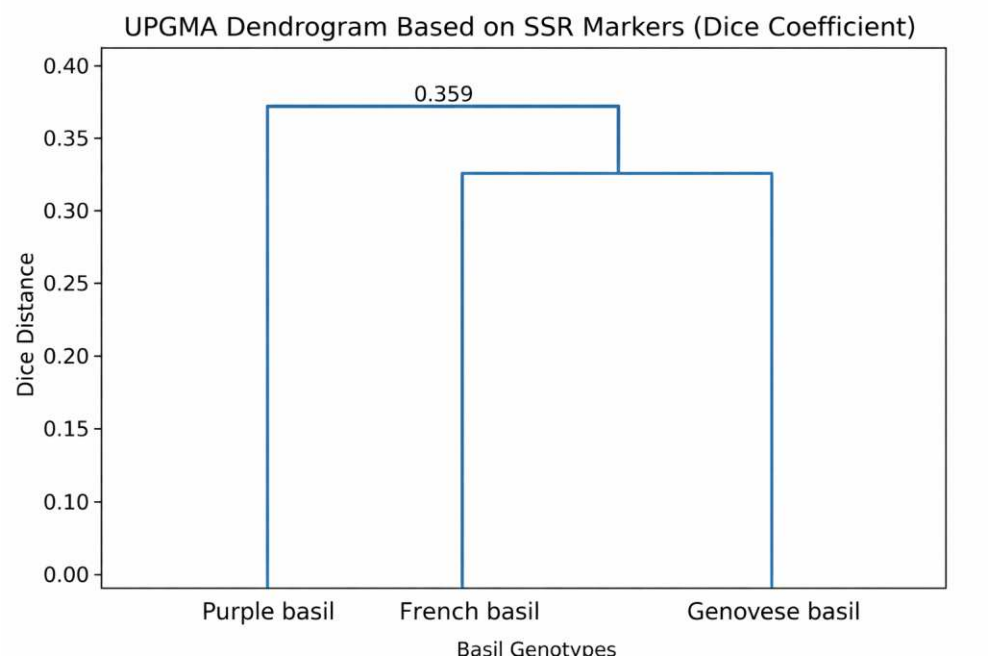


Figure 3. UPGMA dendrogram based on Dice similarity coefficients from 40 SSR markers. French basil and Genovese basil cluster together, while Purple basil forms a distinct outgroup, reflecting its genetic divergence.

3.6 SNP Marker Analysis

Analysis of single-nucleotide polymorphisms (SNP) revealed 12 distinct polymorphic loci among the three basil varieties: French basil, Genovese basil, and Purple basil (Table 5). The SNPs were located within the examined DNA sequence at alignment points between 145 and 2540 bp, showing sequence diversity among the investigated genotypes. Of the identified SNPs, eight were transitions (A↔G or C↔T) and four were transversions (T↔G or G↔T), indicating that transition mutations occurred more commonly in the examined area. The identified SNPs demonstrated genotype-specific nucleotide differences, enabling a clear distinction among the basil genotypes.

Genovese basil showed six genotype-specific SNPs at locations 145, 487, 821, 1498, 1735, and 2216 bp, where its nucleotides were distinct from those found in French and Purple basil. Likewise, Purple basil exhibited six particular SNPs at locations 312, 633,

1044, 1269, 1982, and 2540 bp, signifying unique nucleotide alterations specific to this genotype. In contrast, French basil exhibited shared nucleotides with one of the other genotypes at every analyzed locus, suggesting a relatively closer genetic connection.

The SNP markers exhibited distinct nucleotide polymorphisms among the basil genotypes, offering dependable molecular signatures for differentiating genotypes and confirming the existence of genetic diversity within the examined *Ocimum basilicum* genotypes.

Table 5. Representative SNP markers and nucleotide differences among basil genotypes

SNP ID	Alignment Position (bp)	French Basil	Genovese Basil	Purple Basil	SNP Type	Genotype Pattern
SNP-1	145	A	G	A	Transition	Genovese-specific
SNP-2	312	T	T	C	Transition	Purple-specific
SNP-3	487	C	T	C	Transition	Genovese-specific
SNP-4	633	G	G	A	Transition	Purple-specific
SNP-5	821	T	C	T	Transition	Genovese-specific
SNP-6	1044	A	A	G	Transition	Purple-specific
SNP-7	1269	C	C	T	Transition	Purple-specific
SNP-8	1498	G	A	G	Transition	Genovese-specific
SNP-9	1735	T	G	T	Transversion	Genovese-specific
SNP-10	1982	A	A	C	Transversion	Purple-specific
SNP-11	2216	C	T	C	Transition	Genovese-specific
SNP-12	2540	G	G	T	Transversion	Purple-specific

3.7 Phylogenetic Analysis Based on SNP Markers

The genetic connections between the three basil genotypes were assessed using SNP marker data, leading to the construction of a phylogenetic tree through the UPGMA clustering approach (Figure 4, panel A). The dendrogram illustrated distinct genetic variation among the examined genotypes.

The analysis placed French basil and Genovese basil in the same primary cluster, suggesting a closer genetic connection between these two varieties. In this cluster, French basil created a unique sub-branch with a robust bootstrap support value of 97, indicating significant genetic similarity while still exhibiting a certain level of differentiation from Genovese basil. The node that distinguishes this pair from the other genotypes exhibited a bootstrap value of 92, demonstrating dependable clustering. Conversely, Purple basil established a distinct branch, clearly separate from the other two genotypes, with a bootstrap support value of 100, indicating robust statistical confidence in its genetic separation. This distinction shows that Purple basil has more genetic diversity compared to French and Genovese basil. The UPGMA dendrogram constructed from SNP markers reveals two primary genetic clusters, with French and Genovese basil grouped together, whereas Purple basil forms a separate genetic lineage. These findings indicate genetic variation among the examined basil genotypes and validate the effectiveness of SNP markers in evaluating phylogenetic connections in *Ocimum basilicum*.

3. 8 Integration of Biochemical Traits with SNP–Metabolite Clustering

The heatmap analysis (Fig. 4, panel B) revealed clear and structured correlations between SNP markers and biochemical traits, including phenolic compounds and essential oil content. Positive correlations (red scale) were observed between several SNPs (notably SNP1, SNP5, SNP7, and SNP10) and major metabolites such as caffeic acid, quercetin, catechin, and essential oil content, suggesting that these loci may play a role in regulating both phenylpropanoid metabolism and oil biosynthesis pathways.

In contrast, negative correlations (blue scale) were identified for certain SNPs (e.g., SNP2 and SNP6), particularly with flavonoids such as apigenin and luteolin, as well as with oil accumulation in some cases. This indicates potential antagonistic genetic effects or pathway trade-offs between flavonoid synthesis and oil production.

Hierarchical clustering of SNPs and traits revealed functional grouping patterns. Metabolites associated with the phenylpropanoid pathway (coumaric acid, ferulic acid, caffeic acid) clustered together and showed similar SNP association profiles, while flavonoids (quercetin, luteolin, apigenin, catechin) formed another cluster. Notably, essential oil content clustered independently but showed partial association with phenolic compounds, suggesting shared upstream regulatory mechanisms but distinct downstream biosynthetic control. French and Genovese basil exhibited highly similar SNP–trait correlation profiles, including comparable patterns for oil content, supporting their close phylogenetic relationship. In contrast, Purple basil displayed a distinct metabolic signature, characterized by differential correlations between SNPs and both phenolic compounds and essential oil. This unique profile likely reflects genotype-specific regulation of secondary metabolism, particularly pathways linked to pigmentation and volatile compound biosynthesis.

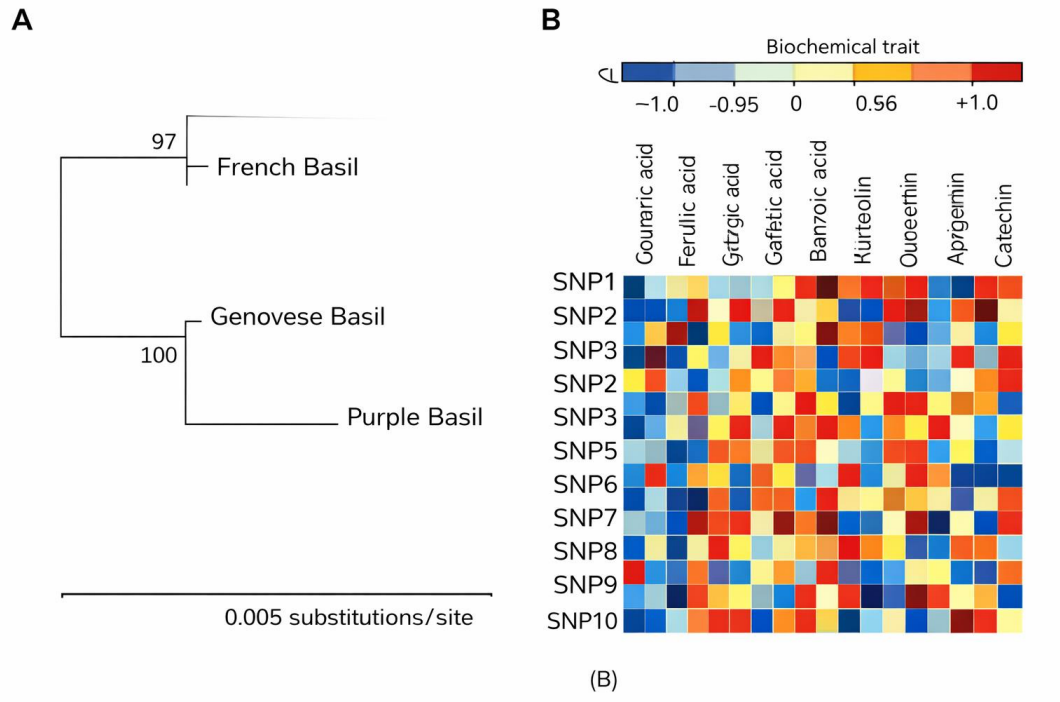


Figure 4. Integrated analysis of SNP-based phylogeny and SNP–biochemical trait associations in basil genotypes.
(A) Phylogenetic tree constructed from SNP data showing the genetic relationships among French, Genovese, and Purple basil. Bootstrap values (%) are indicated at the

nodes, and the scale bar represents nucleotide substitutions per site. (B) Heatmap illustrating correlations between SNP markers and biochemical traits, including phenolic compounds and essential oil content. Color gradients range from blue (negative correlation) to red (positive correlation). SNPs and traits are hierarchically clustered to highlight functional associations and pathway-level organization

4. Discussion

4.1 Biochemical Diversity and Genotype Specialization

The three basil genotypes exhibited pronounced biochemical differentiation, particularly in essential oil and phenolic/flavonoid composition (Mulugeta et al., 2024). Genovese basil accumulated the highest essential oil content (2.30%), along with elevated levels of flavonoids such as rutin (40.0 μg/L) and luteolin (30.0 μg/L) (Nazir et al., 2019; Farag et al., 2016). In contrast, French basil was characterized by high caffeic acid (76.3 μg/L) and benzoic acid (46.2 μg/L) (Ćavar Zeljković et al., 2020), whereas Purple basil contained the highest apigenin (49.0 μg/L) and hydroxycinnamic acids (coumaric 44.0 μg/L, ferulic 23.5 μg/L) (Phippen & Simon, 2000). These differences highlight distinct chemotypic specialization: Genovese for oil and flavonoids, French for specific phenolic acids, and Purple for anthocyanin-associated metabolites (Gavrić et al., 2018). The PCA further confirmed these patterns, with PC1 separating Genovese basil based on flavonoid and oil content, while PC2 distinguished French basil due to elevated caffeic and benzoic acids (Ćavar Zeljković et al., 2020). This underscores that both essential oil and phenolic pathways are major drivers of biochemical diversity among basil genotypes (Mulugeta et al., 2024; Phippen & Simon, 2000).

A key observation is the strong inverse relationship between oil content and specific phenolics, particularly coumaric acid ($r=-0.98$) and apigenin ($r=-0.97$), reflecting a metabolic trade-off between terpenoid and phenylpropanoid pathways. These patterns are consistent with previous findings indicating that Genovese and Purple basil direct carbon flux toward different secondary pathways: Genovese toward terpenoids and selected flavonoids, and Purple toward phenylpropanoids and anthocyanins (Prinsi et al., 2020). The strong positive correlations between oil content and rutin ($r=0.93$) and luteolin ($r=0.86$) in Genovese indicate selective enhancement of certain flavonoid branches rather than a global suppression of phenylpropanoid metabolism.

4.2 Genetic Divergence using SSR and SNP markers

SSR marker analysis revealed moderate to high polymorphism (70%), effectively distinguishing genotypes. Dice similarity coefficients indicated that French and Genovese basil are genetically closer (0.667), whereas Purple basil is more distinct (0.632–0.651), consistent with its unique pigmentation and metabolite profile. UPGMA clustering corroborated these findings, placing French and Genovese in one cluster and Purple as a separate outgroup. SNP analysis identified twelve genotype-specific loci, six unique to Genovese and six to Purple, while French shared alleles with both, reflecting its intermediary genetic position. The dendrogram shows that French basil and Genovese basil share a closer genetic relationship with each other than with Purple basil, which is the most genetically distinct variety of the three assessed accessions. These findings align with the similarity matrix obtained from the SSR markers and validate the effectiveness of SSR-based molecular analysis in evaluating genetic diversity within basil germplasm. This divergence is consistent with previous reports indicating that European "pesto-type" basils have undergone intensive selection for high essential oil content, resulting in a genetic bottleneck that favors industrial traits over broad phenolic diversity (Ciriello et al., 2021).

4.3. Integration of SNP Variation to Metabolic Diversity in Basil

The integration of biochemical profiling with SNP metabolite clustering provides a comprehensive view of genotype-dependent metabolic regulation in basil. The clear separation of metabolite clusters and their association with specific SNP groups supports a hypothesis-generating framework, suggesting that genetic variation underlies the observed biochemical diversity. Similar integrative approaches have proven effective in linking genetic polymorphisms to metabolic traits in plants (Chen et al., 2013; Wen et al., 2014).

The observed variation in oil content, with Genovese basil exhibiting the highest levels, reflects differential regulation of lipid biosynthesis pathways. Essential oil accumulation in basil is known to be under strong genetic control and influenced by genotype-specific expression of terpene synthase genes (Gang et al., 2001). The clustering of oil-associated

traits with positively correlated metabolite groups in Figure 4 further supports the existence of coordinated metabolic regulation.

Phenolic acids, including coumaric, ferulic, and caffeic acids, showed distinct genotype-specific patterns, highlighting divergence in the phenylpropanoid pathway. This pathway is a central metabolic route responsible for the biosynthesis of a wide range of secondary metabolites involved in plant defense and antioxidant activity (Vogt, 2010). The enrichment of caffeic and benzoic acids in French basil suggests enhanced flux through upstream phenylpropanoid intermediates, potentially mediated by key enzymes such as phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H) (Fraser and Chapple, 2011). The clustering of these metabolites in the heatmap indicates co-regulation, which may be linked to shared genetic loci.

Flavonoid compounds, including rutin, quercetin, luteolin, apigenin, and catechin, exhibited strong genotype specificity and formed distinct clusters in Figure 4. These metabolites are synthesized downstream of the phenylpropanoid pathway and are regulated by a complex network of structural and regulatory genes (Lepiniec et al., 2006). The high accumulation of rutin and luteolin in Genovese basil suggests enhanced activity of flavonoid biosynthetic enzymes such as flavonol synthase (FLS) and flavone synthase (FNS). In contrast, the elevated quercetin content in French basil indicates differential pathway branching, which has been reported to be controlled by transcription factors such as MYB and bHLH proteins (Xu et al., 2013).

The hierarchical clustering demonstrated that metabolites with similar biochemical functions tend to group together, forming metabolic modules. These modules exhibit consistent correlation patterns with specific SNP clusters, suggesting potential genetic control points. Such SNP metabolite associations have been widely reported in metabolomics-assisted breeding and can be used to identify candidate genes controlling key agronomic and nutritional traits (Matsuda et al., 2012; Alseekh and Fernie, 2018). Importantly, the coordinated clustering of phenolic acids and flavonoids highlights the central role of the phenylpropanoid–flavonoid network in shaping basil metabolic diversity. This network is highly responsive to both genetic and environmental factors and is a major determinant of plant quality traits, including antioxidant capacity, flavor, and medicinal properties (Klee, 2010; Kumar and Pandey, 2013).

5. Conclusions

This study highlights the distinctive biochemical and genetic profiles of three basil genotypes: Genovese, French, and Purple. Genovese basil is characterized by its high essential oil content and elevated levels of specific flavonoids, while French basil is noted for its significant accumulation of caffeic and benzoic acids. In contrast, Purple basil exhibits a unique profile with elevated apigenin and hydroxycinnamic acids, signifying distinct metabolic trade-offs between terpenoid and phenylpropanoid pathways. Through SSR and SNP marker analyses, we established effective methodologies for distinguishing these genotypes, revealing moderate to high genetic polymorphism and elucidating the genetic structure within the basil germplasm. The integration of biochemical data with molecular markers provided valuable insights into the genotype-specific regulation of secondary metabolites and their underlying genetic mechanisms. This foundational work not only enhances our understanding of basil's metabolic diversity but also paves the way for future research focused on functional genomics and marker-assisted selection aimed at optimizing the agronomic and therapeutic traits of aromatic plants.

Author Contributions: E.Y.A.: investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization. K.R.A.: methodology, investigation, resources, data curation, visualization. M.I.M.: methodology, investigation, data curation, writing—review and editing, and visualization.

Acknowledgments: The researchers would like to thank the Deanship of Scientific Research, Qassim University, for funding the publication of this study (QU-APC-2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Data Availability Statement: The original contributions presented in this study are included in the article; further inquiries can be directed at the corresponding author.

6. References

Alseekh, S. and Fernie, A.R. (2018), Metabolomics 20 years on: what have we learned and what hurdles remain?. Plant J, 94: 933-942. https://doi.org/10.1111/tpj.13950

Borevitz, J. O., et al. (2000). Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research, 10*, 1748–1757. DOI: 10.1101/gr.183100

Carović-Stanko, K., Orlić, S., Politeo, O., Strikić, F., Kolak, I., Milos, M., & Satovic, Z. (2010). Composition and antibacterial activities of essential oils of seven *Ocimum* taxa. *Food Chemistry, 119*, 196–201.

Ćavar Zeljković, S., Komzáková, K., Šišková, J., Karalija, E., Smékalová, K., & Tarkowski, P. (2020). *Phytochemical variability of selected basil genotypes.* Industrial Crops and Products, 157, 112910. DOI: 10.1016/j.indcrop.2020.112910

Chen, W.; Gong, L.; Guo, Z.; Wang, W.; Zhang, H.; Liu, X.; Yu, S.; Xiong, L.; Luo, J. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: Application in the study of rice metabolomics. Mol. Plant 2013, 6, 1769–1780.

Ciriello, M., Formisano, L., Kyriacou, M. C., Sifola, M. I., Graziani, G., & Rouphael, Y. (2021). Genotype and harvest time affect the phytochemical quality of *Ocimum basilicum* L. *Foods, 10*(8), 1782.

Deschamps, C., Gang, D., Dudareva, N., & Simon, J. E. (2006). The role of glandular trichomes in the biosynthesis of essential oils in *Ocimum basilicum*. *Annals of Botany, 97*(5), 657–671.

Farag, M. A., Ezzat, S. M., Salama, M. M., & Tadros, M. G. (2016). Anti-acetylcholinesterase potential and metabolome classification of four *Ocimum* species. *Journal of Pharmaceutical and Biomedical Analysis, 125*, 292–302. DOI: 10.1016/j.jpba.2016.01.016

Flanigan, P. M., & Niemeyer, E. D. (2014). Anthocyanin content and antioxidant capacity in purple basil (*Ocimum basilicum* L.). *Journal of Agricultural and Food Chemistry, 62*, 658–665.

Fraser, C.M.; Chapple, C. The phenylpropanoid pathway in Arabidopsis. Arabidopsis Book 2011, 9, e0152. doi: 10.1199/tab.0152.

Gang, D.R.; Wang, J.; Dudareva, N.; Nam, K.H.; Simon, J.E.; Lewinsohn, E.; Pichersky, E. An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. Plant Physiol. 2001, 125, 539–555

Gavrić T, Jurković J, Hamidović S, Haseljić S, Lalević B, Čorbo A, Bezdrob M. Yield and contents of some bioactive components of basil (*Ocimum basilicum* L.) depending on time of cutting. *Stud Univ Vasile Goldis Arad Ser Stiint Vietii.* 2018;28:192–197.

Khosla, A., Kumar, S., & Gupta, A. (2019). SSR markers in Ocimum species: applications in genetic diversity studies. *Molecular Biology Reports, 46*, 2079–2090.

Klee, H.J. (2010). Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. *New Phytologist*, 187(1), 44-56. https://doi.org/10.1111/j.1469-8137.2010.03281.x

Kumar, S. & Pandey, A.K. (2013). Chemistry and biological activities of flavonoids: an overview. *Scientific World Journal*, 2013, Article 162750. https://doi.org/10.1155/2013/162750

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547–1549.

Kwee, E., & Galaffu, N. (2011). Extraction and HPLC quantification of phenolic acids and flavonoids in basil (*Ocimum basilicum* L.). *Food Chemistry, 126*, 1232–1238. DOI: 10.1111/j.1750-3841.2011.02435.x

Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol. 2006, 57:405-30. doi: 10.1146/annurev.arplant.57.032905.105252.

Mahajan V, Rather IA, Awasthi P, Gairola S, Meena S, Anand R, Bedi YS, Gandhi S. Development of chemical and EST-SSR markers for *Ocimum* genus. *Ind Crops Prod.* 2014; 63:10–16. doi:10.1016/j.indcrop.2014.10.052.

Makri, O., & Kintzios, S. (2008). *Ocimum sp. (Basil): Botany, Cultivation, Pharmaceutical Properties, and Biotechnology.* Journal of Herbs, Spices & Medicinal Plants, 13(3), 123–150. DOI: 10.1300/J044v13n03_10

Matsuda, F.; Okazaki, Y.; Oikawa, A.; Kusano, M.; Nakabayashi, R.; Kikuchi, J.; Yonemaru, J.I.; Ebana, K.; Yano, M.; Saito, K. Dissection of genotype–phenotype associations in rice grains using metabolome quantitative trait loci analysis. Plant J. 2012, 70, 624–636

Mulugeta, S. M., Pluhár, Z., & Radácsi, P. (2024). Phenotypic variations and bioactive constituents among selected *Ocimum* species. *Plants, 13*(1), 64.

Nazir M, Tungmunnithum D, Bose S, Drouet S, Garros L, Giglioli-Guivarc'h N, Abbasi BH, Hano C**.** Differential production of phenylpropanoid metabolites in callus cultures of *Ocimum basilicum* L. with distinct in vitro antioxidant activities and in vivo protective effects against UV stress. *J Agric Food Chem.* 2019;67(7):1847–1859. doi:10.1021/acs.jafc.8b05647.

Phippen, W. B., & Simon, J. E. (2000). Anthocyanin inheritance and instability in purple basil (*Ocimum basilicum* L.). *Journal of Heredity, 91*(4), 289–296. DOI: 10.1093/jhered/91.4.289

Prinsi, B.; Negrini, N.; Morgutti, S.; Espen, L. Nitrogen Starvation and Nitrate or Ammonium Availability Differently Affect Phenolic Composition in Green and Purple Basil. *Agronomy* **2020**, *10*, 498. https://doi.org/10.3390/agronomy10040498

Pyne RM, Honig JA, Vaiciunas J, Wyenandt CA, Simon JE**.** Population structure, genetic diversity and downy mildew resistance among *Ocimum* species germplasm. *BMC Plant Biol.* 2018;18:69. doi:10.1186/s12870-018-1284-7.

Vogt, T. (2010) Phenylpropanoid Biosynthesis. Molecular Plant, 3, 2-20. http://dx.doi.org/10.1093/mp/ssp106

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J. & Yan, J**.** (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications*, 5, 3438. https://doi.org/10.1038/ncomms4438.

Xu W, Grain D, Le Gourrierec J, Harscoët E, Berger A, Jauvion V, Scagnelli A, Berger N, Bidzinski P, Kelemen Z, Salsac F, Baudry A, Routaboul JM, Lepiniec L, Dubos C. Regulation of flavonoid biosynthesis involves an unexpected complex transcriptional regulation of TT8 expression, in Arabidopsis. New Phytol. 2013;198(1):59-70. doi: 10.1111/nph.12142..

Yang, X., .Song, J., You, Q., Paudel, D. R., Zhang, J, & Wangm J. (2017). Mining sequence variations in representative polyploid sugarcane germplasm accessions. BMC Genomics18:594. 10.1186/s12864-017-3980-3

Zhou, F., & Pichersky, E. (2020). Metabolic engineering of terpenoid and phenylpropanoid pathways in plants. *Current Opinion in Biotechnology, 61*, 29–36. DOI: 10.1016/j.pbi.2019.10.005