



The Draft Transcriptome of *Arctostaphylos Glauca*: De Novo Assembly and Analysis

SAVANAH SENN^{1*}, MAX MORVAN¹, ARTEM CHEREVATENKO¹,
MAYA REFUA¹, STEVEN CARREL², GERALD PRESLEY³.

¹ Los Angeles Pierce College, Department of Agriculture Sciences, Plant Science Program,
6201 Winnetka Ave., Woodland Hills, CA 91371

² Oregon State University, Center for Quantitative Life Sciences, 1500 SW Jefferson Way, Corvallis, OR
97331

³ Oregon State University, Department of Wood Science and Engineering, 1500 SW Jefferson Way,
Corvallis, OR 97331

*stclais@piercecollege.edu

Abstract: *Arctostaphylos glauca* (bigberry manzanita) is a California native plant species adapted to drought and fire-prone environments. As extreme weather events intensify abiotic stressors such as heat and water limitation, understanding the molecular mechanisms behind plant resilience is increasingly critical. This study aimed to investigate secondary metabolite gene expression and functional diversity in *A. glauca* using RNA sequencing and whole-genome similarity searches. A de novo transcriptome assembly was generated and functionally annotated using Trinotate and BLAST-based homology, revealing high expression of genes associated with the phenylpropanoid and flavonoid pathways. Chalcone synthase (CHS1), a key biosynthetic enzyme in the formation of flavonoid precursors, showed elevated transcript abundance and strong alignment to conserved orthologs in related species. Additional candidates included phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), and anthocyanidin reductase (ANR), enzymes known for roles in antioxidant, lignin, and proanthocyanidin biosynthesis. Whole-genome BLAST searches further confirmed homology of key transcripts to chromosomes and scaffolds in *Ericaceae* species including *A. glauca*, *Vaccinium corymbosum*, and *Rhododendron simsii*. Notably, multiple heat shock protein (HSP) isoforms were identified, supporting a model in which stress-response genes and secondary metabolism jointly contribute to the plant's survival strategy. These findings underscore the ecological and physiological importance of specialized metabolites and molecular stress responses in native woody shrubs. The identification of conserved biosynthetic pathways and stress-response genes lays the groundwork for future studies on resilience traits and conservation genomics in California chaparral ecosystems. The study provides valuable insights into the genetic makeup of manzanita and its adaptive traits.

Keywords: RNAseq; plant diversity; conservation; antioxidants

© 2026 under the terms of the J ATE Open Access Publishing Agreement

Introduction

Genomic resources for the conservation of California native plants have gained increased attention recently; however, prior to our study, there were no *Arctostaphylos glauca* whole transcriptome data published that documented leaf and fruit RNAseq. The purpose of this study was to study the functions of biotic and abiotic stress tolerance, including secondary metabolite production, that may make these plants more resilient.



Arctostaphylos glauca (Ericaceae) is a species of manzanita, a California native plant well-known for its wildfire regeneration abilities and adaptations, with *A. glauca* adapted specifically for infrequent, higher-intensity fires. These include the production of obligate seeds with hard coatings that germinate post-fire [1],[2],[3], a longer maturation period and overall lifespan in comparison with other genus members [1],[4], and the production of allelopathic chemicals such as arbutin and phenolic acids which inhibit understory growth [4].

High-severity fire events are known to cause profound and long-lasting changes in soil physical, chemical, and biological properties, including nutrient availability, redox conditions, microbial community structure, and organic matter composition. These post-fire conditions persist for years to decades and directly influence seedling establishment, metabolism, and stress physiology. As documented in our prior work [5], post-fire landscapes in the Angeles National Forest exhibit altered nitrogen cycling, shifts in microbial functional potential, and delayed soil recovery—factors that exert sustained selective pressure on regenerating plants.

Arctostaphylos glauca is a well-established fire-adapted chaparral species whose life history includes obligate post-fire germination, prolonged juvenile development, and persistence in nutrient-limited, drought-prone environments. Because the sampled individual germinated shortly after a high-intensity fire, it developed entirely within a post-fire soil and climatic context. Under these conditions, the expression of genes associated with oxidative stress mitigation, secondary metabolite biosynthesis, and cellular stress response is biologically expected.

Many of *A. glauca*'s abilities can be attributed to secondary metabolites, organic compounds produced by plants that are not the products of photosynthesis, such as terpenoids, phenolic compounds, alkaloids, and sulfur-containing compounds[6]. Manzanita has a rich history of traditional use among Native American cultures in California. Indigenous peoples utilized different parts of the plant for medicinal purposes, including the leaves, bark, and berries. Manzanita was employed in remedies for ailments such as colds, sore throats, and digestive issues. Ethnobotanical studies [7],[8] have documented the traditional knowledge and practices surrounding manzanita use among Native American tribes, providing valuable insights into its cultural significance and medicinal properties.

Methods

Tissue was collected from the leaf and fruit of one individual identified in the field at Gold Creek Preserve, Angeles National Forest, and immediately snap-frozen in liquid nitrogen using a portable dewar (Table 1). The Red Trail is a high fire intensity history area which was affected by the Station Fire in 2009 and the Creek Fire in 2018. The manzanita trees sprouted from seed in the understory of previous manzanitas following the Creek Fire in 2018, and tissue was sampled from one tree on May 12, 2022. The tree germinated during the spring following the Creek Fire. Samples were shipped to BGI Americas on dry ice for RNA extraction, library construction, and sequencing on the DNBSeg platform. Library construction at BGI Americas used the approach of Oligo(dT) magnetic beads for mRNA enrichment, mRNA fragmentation, cDNA synthesis, size selection, and PCR [9].

Table 1. The metadata for the manzanita individual that was sampled is shown in the table

Tissue	Location	Trail	Latitude	Longitude	Elevation
leaf	Gold Creek Preserve	Red Trail	34.321667	-118.31111	2200
fruit	Gold Creek Preserve	Red Trail	34.321667	-118.31111	2200

Following RNA sequencing of the fruit and leaf tissues separately, the sample data were combined for quality control, assembly, and quantification, and the plant transcripts were annotated with candidate genes from the best SPROT (SwissPROT) BLASTX results using default settings for each package. Quality control was performed with MultiQC [10] and SOAPnuke [11]. De novo assembly was performed with Trinity [12]. Trinity uses an approach of partitioning sequence data into numerous batches in order to capture multiple



isoforms in the absence of reference data; please see Chunk 1 in supplemental materials for source code. This is unlike genome assembly, which has greater connectivity between contigs [12]. The transcriptome assembly was evaluated with BUSCO [13]. Quantification was performed using Salmon [14], and transcripts were annotated with Trinotate [15]. Summary statistics for the transcriptome were computed using the script `nano.py` available in Chunk 3 under Supplemental Materials. In order to assist with interpretation of transcripts with relevance to secondary metabolites, the shell script `find_transcript_ids.sh` was used to filter transcripts with $\text{TPM} > 10$ and $\text{length} > 1000$ [16]. It is a publicly available script on GitHub that was developed by Pierce Plant Science student John Hsieh. This made it possible to filter non-plant transcripts for additional quality control, narrow the number of transcripts of interest to 6,840, and hand-curate 391 secondary metabolite and stress-related candidate genes in *Arctostaphylos glauca* leaf and fruit transcripts. The 391 candidate genes were curated manually based on their known function as secondary metabolite and stress response genes, according to the SPROT best blastx hits.

To maximize annotation accuracy and biological interpretability in a non-model species, we employed a two-stage transcript annotation strategy. First, functional annotation was performed using Trinotate with Swiss-Prot (SPROT), a manually curated protein database that enables reliable identification of conserved enzymatic functions. This step provides a high-confidence functional baseline for pathway-level interpretation. Second, annotated transcripts were mapped to closely related taxa using the Core nucleotide and Whole Genome Shotgun (WGS) databases, allowing phylogenetic validation and confirmation of transcript origin within the *Ericaceae* family. Please see Chunk 3 and Tutorial 1 in Supplemental Materials for full workflow and source code. Together, these complementary approaches integrate functional reliability with evolutionary relevance, strengthening confidence in the biological interpretation of the *Arctostaphylos glauca* transcriptome.

Gene mapping to closely related species was carried out next. The BLAST analysis began by filtering assembled Trinity FASTA sequences using a Python script to retain only transcripts of interest. These filtered sequences were then divided into smaller batches for efficient processing using a `fasta-splitter` Perl script [17]. Each batch underwent a megablast search (`blastn`) against nucleotide databases restricted to closely related species within the appropriate plant family *Ericaceae*. The resulting CSV hit tables were concatenated and filtered, keeping only matches with the highest percent identity. Finally, gene names and annotations were retrieved using batch ENTREZ, providing essential information for subsequent biological interpretation.

Results and Discussion

The NCBI SRA for *A. glauca* transcriptomic data is available under BioProject PRJNA1046570. As shown in Table 2, the total number of sequences in the *Arctostaphylos glauca* draft transcriptome was 157,151. The average length was 966.9, the L50 was 30,163, and the N80 was 649. Out of a total of 157 transcripts, 151 transcripts, 15,004 sequences were successfully annotated in Trinotate with the best SPROT blastx hit. After a non-plant gene filter was applied and transcripts with $\text{length} < 1000$ and $\text{TPM} < 10$ were excluded, 6,840 transcripts were validated based on those criteria. Figure 4 presents a size component distribution analysis, indicating that most transcripts were between 1001- 2500 bases in length. Of these validated transcripts, 391 isomers related to plant stress response and secondary metabolite production were selected for downstream analysis.

Table 2. The summary statistics for the draft transcriptome are given in the table

Number of Sequences	Total Length	Average Length	Min Length	Max Length	L50	N50	N80
157,151	1.52E+08	966.9	178	14,749	30,163	1,605	649

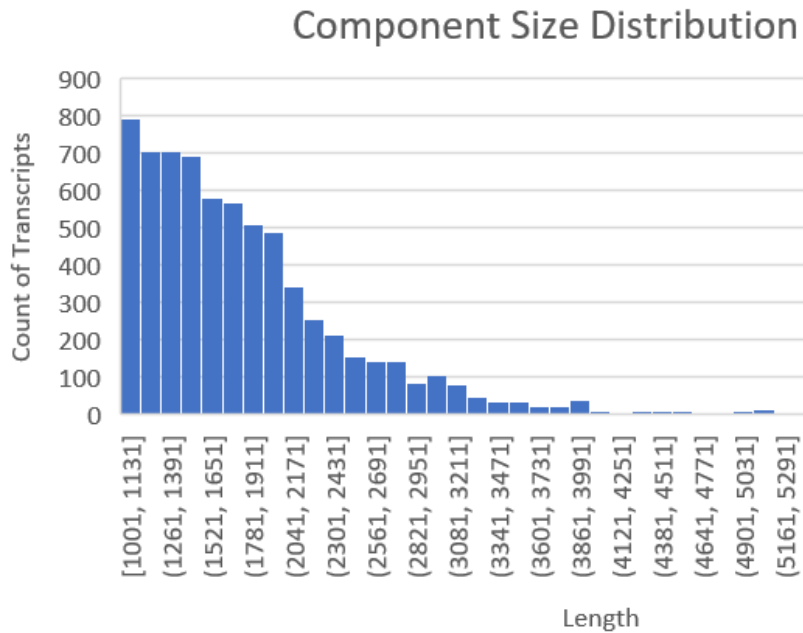


Fig. 1. The component size distribution analysis of 6,840 transcripts that were validated with annotation, length, and TPM metrics, and a non-plant gene filter revealed that most of the annotated transcripts were between 1000-2500 bases.

According to the BUSCO assessment results, the transcriptome was 92.18% complete (Figure 2). BUSCO assesses the completeness of genomic assemblies by comparing the assembly to a highly conserved set of 2,326 genes from OrthoDB [13].

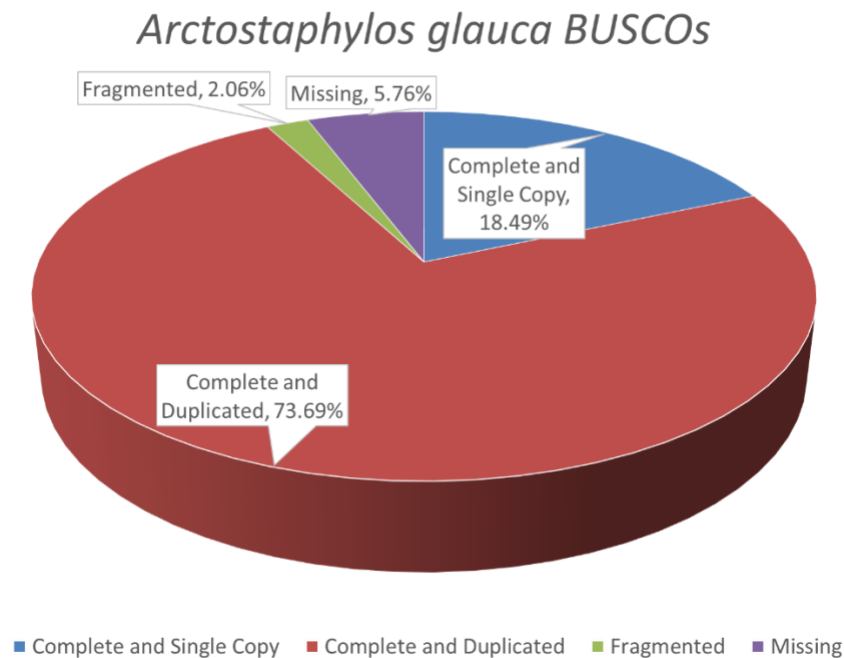




Fig. 2. According to the BUSCO assessment, the transcriptome assembly was 92.18%, indicating that it was in a draft state (completeness<95%).

Trinotate Annotation with best SPROT blastx hits

With the transcriptome assembly quality established, we next examined gene function, focusing on stress-related and secondary metabolite genes identified through Trinotate annotation. Among the elevated secondary metabolite transcripts, Phenylalanine ammonia lyase (PAL) was closely related to the *Camellia* homolog (88.94% ID, TPM=29.59, length=2534), indicating conserved functionality. PAL catalyzes the conversion of phenylalanine to trans-cinnamic acid, a key step in the phenylpropanoid pathway, which leads to the biosynthesis of various secondary metabolites such as flavonoids, lignin, and coumarins. This enzyme's role in the phenylpropanoid pathway and secondary metabolite biosynthesis is well-established [18].

Chalcone synthase 1 (CHS1) was related to *Camellia* (TPM=312.72, 95.61% ID, length=1596). There was also an indication that at least one candidate Chalcone synthase transcript was related to *Vitis* (94.19% ID, length=1103, TPM=17.93). CHS1 is an enzyme involved in the biosynthesis of flavonoids, particularly chalcones, which serve as precursors for various flavonoids, including flavones, flavanols, and anthocyanins. Chalcones also possess antioxidant and antimicrobial properties. CHS1 is a highly conserved enzyme that catalyzes the condensation of one molecule of 4-coumaroyl-CoA with three molecules of malonyl-CoA to form naringenin chalcone, a precursor of flavonoids (Figure 3). Its structure typically comprises a catalytic domain and a substrate-binding pocket [19]. The elevated expression of this candidate gene highlights the potential role of flavonoid biosynthesis in this plant's resiliency.

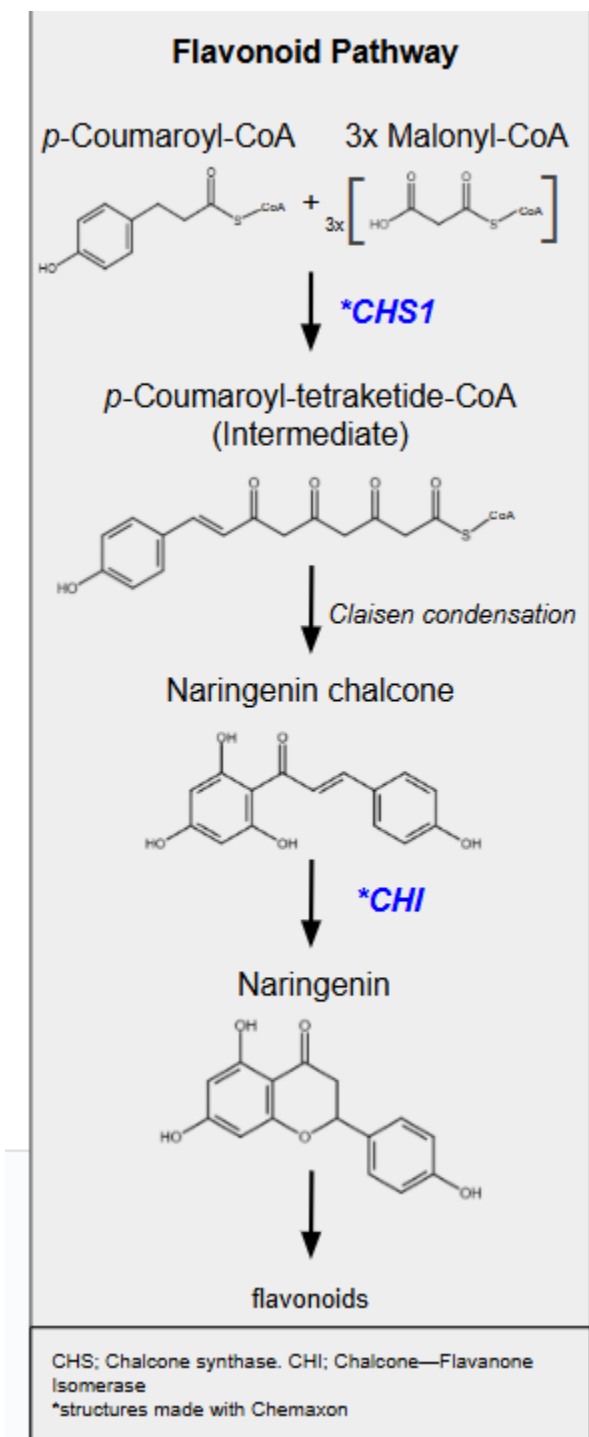


*Fig. 3. Proposed biosynthetic pathway catalyzed by chalcone synthase (CHS1), illustrating the condensation of 4-coumaroyl-CoA and malonyl-CoA to form naringenin chalcone, a key intermediate in flavonoid biosynthesis. CHS1 activity contributes to the production of antioxidant and antimicrobial flavonoids, which may support stress resilience in *Arctostaphylos glauca*.*

Chalcone—flavanone isomerase was 81.28% similar to *Camellia* (TPM=179.18, length=1059). Chalcone—Flavanone Isomerase is an enzyme involved in the flavonoid biosynthesis pathway, catalyzing the conversion of chalcones to flavanones. This enzymatic reaction plays a pivotal role in the synthesis of various flavonoids



with diverse biological activities (Figure 3). Structurally, Chalcone—Flavanone Isomerase typically comprises a catalytic domain responsible for the isomerization reaction and a substrate-binding pocket. The enzyme's activity is essential to produce flavanones, which serve as precursors for a wide range of secondary metabolites in plants, including flavonols, flavones, and anthocyanins. This enzyme participates in the conversion of chalcones to flavanones, which is crucial in flavonoid metabolism [20].

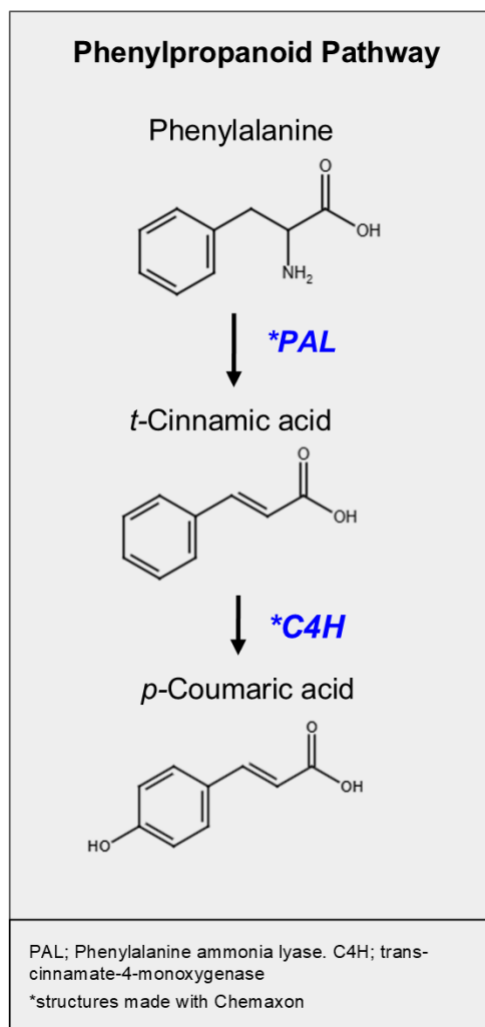


Fig. 4. Phenylalanine ammonia lyase and trans-cinnamate-4-monooxygenase are key enzymes in the phenylpropanoid pathway.

Anthocyanidin reductase ((2S)-flavon-3-ol-forming) was elevated, with its TPM (transcripts per million) =94.13 and length=1715. This transcript was most similar to *Vitis*. Anthocyanidin reductase catalyzes the conversion of anthocyanidins to flavan-3-ols, contributing to the biosynthesis of flavonoids with diverse physiological roles, such as pigmentation, UV protection, and defense against pathogens. Anthocyanidin reductase is a key enzyme in the biosynthesis of flavan-3-ols, which are important secondary metabolites with roles in plant defense and human health. Its structure typically includes NAD(P)H binding sites and catalytic residues [21],[22].

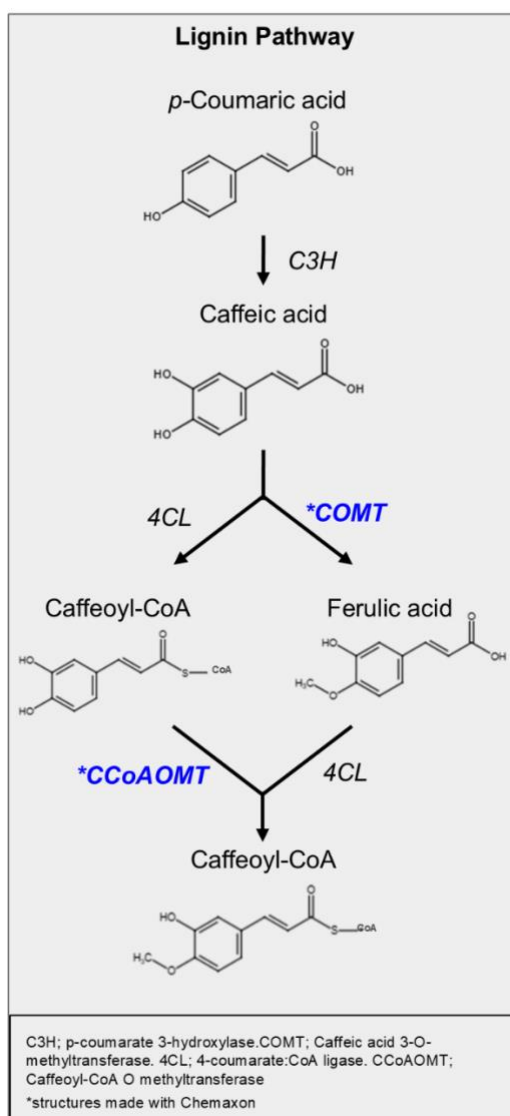


Fig. 5. The transformations of coumarate along the lignin pathway are depicted.

Several elevated transcripts were related to genes in *Populus*, including Trans-cinnamate-4-monooxygenase (TPM=68.19, length= 1979, 92.48% ID). Trans-cinnamate-4-monooxygenase is an enzyme involved in the phenylpropanoid pathway that catalyzes the hydroxylation of trans-cinnamate to *p*-coumarate (Figure 4). It plays a crucial role in lignin biosynthesis and defense against pathogens [23],[24]. Elevated transcripts related to *Populus* genes, including Trans-cinnamate-4-monooxygenase, suggest potential roles in lignin biosynthesis and defense mechanisms.

These results provide insights into the diversity of manzanita genes related to antioxidant production and stress resistance. While information on specific alkaloids, stilbenes, and essential oils in manzanita is scarce, phytochemical analyses of related species provide clues to the potential presence of these compounds. For example, [25] explored the phytochemical profile and pharmacological properties of *Arctostaphylos uva-ursi*, a close relative of manzanita, shedding light on its potential medicinal uses and active constituents.



Understanding the role of *A. glauca*'s secondary metabolites in fire adaptability provides insight into their role in overall fire ecology and underscores the importance of species with specialized fire-adaptive traits as the frequency and intensity of California's wildfires increases.

Transcript mapping to closely related species

To complement annotations from model species, we used MEGABLAST to map transcripts to *Ericaceae* genomes and NCBI core nucleotide sequences, allowing for evolutionary and taxonomic relevance in functional interpretation. In addition to identifying homologous genes involved in phenylpropanoid, flavonoid, and lignin biosynthesis, our BLAST results provided supporting evidence for the phylogenetic placement of *Arctostaphylos glauca* transcripts through top hits to closely related taxa in the *Ericaceae* family. A total of 78 core nucleotide results mapped to our *A. glauca* transcripts of interest; 25 WGS results in the NCBI database matched to our *A. glauca* transcripts of interest from the Trinity assembly. Several annotated transcripts from the core nucleotide and WGS datasets were most similar to genes from *Vaccinium*, *Rhododendron*, and *Erica*, reflecting conserved secondary metabolite biosynthesis pathways across the family. For instance, key enzymes in the phenylpropanoid pathway such as phenylalanine ammonia-lyase (PAL) and chalcone synthase (CHS) displayed high identity to sequences from *Vaccinium corymbosum* and *Rhododendron simsii*, respectively. These homologous genes have previously been implicated in lignin deposition, anthocyanin biosynthesis, and stress response in *Ericaceae* species, supporting their functional conservation in *A. glauca*.

The presence of caffeic acid 3-O-methyltransferase (COMT) and cinnamate-4-hydroxylase (C4H) transcripts with strong similarity to *Erica carnea* and *Vaccinium myrtillus* further validates the active expression of lignin-related biosynthetic genes in fire-adapted manzanita species. Figure 5 shows an example pathway for lignin biosynthesis in manzanita involving these enzymes. These results are consistent with prior observations in other *Ericaceae* members that produce allelopathic phenolics and display fire-regeneration traits. Additionally, defense- and stress-related transcripts, including benzaldehyde dehydrogenase and heat shock cognate 70 kDa, showed strong identity to sequences from *Rhododendron* and *Vaccinium*, indicating shared mechanisms of oxidative stress tolerance within the family. By incorporating high-confidence hits to other *Ericaceae* taxa, this study not only validates the functional relevance of the *A. glauca* transcripts identified here but also highlights the potential for leveraging related species' genomic resources to enhance understanding of secondary metabolism, fire ecology, and evolutionary adaptation in California chaparral systems.

Whole genome mapping results

The top BLAST hits from the Whole Genome Shotgun (WGS) database for *A. glauca* transcripts are shown in Table 3. This table lists high-confidence matches, with annotations indicating genes potentially involved in fire resilience and antioxidant production. Several transcripts associated with stress response and secondary metabolite biosynthesis in *A. glauca* were identified through whole-genome shotgun (WGS) BLAST analysis and mapped to scaffolds of related *Ericaceae* species. For example, transcript TRINITY_DN2859_c0_g1_i1, annotated as a heat shock protein 83-like gene, was mapped to *Vaccinium corymbosum* scaffold 1476, supporting a role in protein folding and stress resilience. Another transcript, TRINITY_DN2564_c0_g1_i1, corresponding to a berberine bridge enzyme-like 15, was mapped to *Rhododendron ripense* scaffold RRI_r1.lsc00030, suggesting involvement in alkaloid biosynthesis or phenolic metabolism. The TRINITY_DN2365_c0_g1_i1 transcript, matching cytochrome P450 73A5 (also known as cinnamate-4-hydroxylase), was mapped to *Vaccinium corymbosum* scaffold 344 and plays a key role in the phenylpropanoid and lignin biosynthesis pathway. Additional transcripts such as TRINITY_DN2499_c0_g1_i1 (trans-cinnamate 4-monooxygenase) and TRINITY_DN2331_c0_g1_i1 (heat shock cognate 70 kDa protein) were mapped to scaffolds in *Rhododendron ripense* and *Vaccinium darrowii*, respectively. These findings emphasize the conservation of core stress and metabolic genes across *Ericaceae* and support the hypothesis that *A. glauca* possesses a rich repertoire of biochemical defenses, possibly underpinning its resilience in fire-prone environments.



Table 3. The BLAST annotation results for selected WGS sequences are shown in the table. Genes were mapped to scaffolds from related Ericaceae species. Functional annotations include roles in the phenylpropanoid pathway (e.g., cytochrome P450 73A5), alkaloid biosynthesis (berberine bridge enzyme-like), and heat stress tolerance (heat shock proteins), highlighting conserved biochemical pathways supporting plant resilience in fire-prone habitats.^{a,b}

Qseqid	Sseqid	Pident	Hit_Name
TRINITY_DN689_c0_g1_i2	JBHWAP010000063.1	100	Arbutus menziesii isolate PacificMadrone02 h2tg000067l_1, WGS
TRINITY_DN689_c0_g1_i3	JBHWAP010000063.1	100	Arbutus menziesii isolate PacificMadrone02 h2tg000067l_1, WGS
TRINITY_DN685_c0_g1_i2	JAHSXP010000004.1	96.8	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_0004, WGS
TRINITY_DN685_c0_g1_i5	JAHSXP010000004.1	96.454	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_0004, WGS
TRINITY_DN685_c0_g1_i9	JAHSXP010000004.1	96.528	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_0004, WGS
TRINITY_DN689_c0_g1_i6	JAHSXP010000014.1	100	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_0014, WGS
TRINITY_DN4868_c0_g1_i1	JAHSWP020000001.1	99.404	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_1, WGS
TRINITY_DN4868_c0_g1_i4	JAHSWP020000001.1	99.159	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_1, WGS
TRINITY_DN1448_c0_g1_i1	JAHSWP020000011.1	100	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_11, WGS
TRINITY_DN3508_c0_g1_i14	JAHSWP020000012.1	100	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_12, WGS
TRINITY_DN4853_c0_g1_i7	JAHSWP020000013.1	98.621	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_13, WGS
TRINITY_DN685_c0_g1_i1	JAHSWP020000035.1	96.549	Arctostaphylos glauca voucher Morrison 13 isolate GLNDR_RDG_Rd-6 SCAF_35, WGS
TRINITY_DN689_c0_g4_i1	JACYLG010002524.1	100	Rhododendron molle isolate GLZ1001 LG01, WGS

^aPident= Percent identity; ^bFull annotations results available in Table S2 in Supplemental Materials

Transcript mapping to core nucleotide accessions

The top BLAST hits from the Core nucleotide database for *Arctostaphylos glauca* transcripts include those related to stress resistance and secondary metabolite pathways, with annotation based on the best e-value matches and source species (Table 4). These included genes involved in flavonoid biosynthesis, such as *chalcone synthase* and *anthocyanidin reductase*, as well as cytochrome P450 enzymes associated with lignin and phenylpropanoid pathways. Several sequences also showed homology to genes from closely related species within the Ericaceae family, including *Rhododendron vialii*, *Vaccinium myrtillus*, and *Arbutus unedo*. These hits strengthen the phylogenetic relevance of the annotations and support the ecological significance of the expressed genes in *Arctostaphylos glauca*. This curated list offers insight into the biochemical strategies used by this species to defend against environmental stressors such as drought, UV exposure, and herbivory.

Several transcripts from *Arctostaphylos glauca* mapped with high identity to annotated genes from closely related Ericaceae species in the core nucleotide database, providing insight into conserved biosynthetic and stress response pathways. For instance, TRINITY_DN1494_c0_g1_i1 showed an 88% identity to *Rhododendron vialii* anthocyanidin reductase, a key enzyme in the flavonoid biosynthesis pathway associated with antioxidant production. TRINITY_DN1251_c0_g1_i2 matched a cytochrome P450 gene from *R. vialii*, indicating potential involvement in diverse secondary metabolite modifications. Two CHS (chalcone synthase) transcripts—TRINITY_DN335_c0_g2_i1 and TRINITY_DN335_c0_g1_i5—were mapped to *Rhododendron*



x hybridum and *Vaccinium myrtillus*, respectively, highlighting strong conservation in early flavonoid biosynthesis steps. Additionally, TRINITY_DN905_c0_g2_i2 aligned closely with a microsatellite sequence from *Arbutus unedo*, which had also been annotated with Chalcone isomerase from the SPROT best blastx hits. These matches underscore the evolutionary conservation of key defense and metabolic genes within Ericaceae and support their relevance in the adaptive physiology of *A. glauca*. The evolutionary relationships between *A. glauca* and closely related species are provided in Figure 6. The most distant genus is *Rhododendron*, and the closest is *Vaccinium*.

Table 4. The BLAST annotation results for selected core nucleotide sequences are shown in the table^{ab}

Qseqid	Sseqid	pident	Hit_Name
TRINITY_DN1494_c0_g1_i1	XM_058342174.1	88.01 4	PREDICTED: Rhododendron vialii anthocyanidin reductase ((2S)-flavan-3-ol-forming) (LOC131313714), mRNA
TRINITY_DN1251_c0_g1_i2	XM_058341023.1	86.96 1	PREDICTED: Rhododendron vialii cytochrome P450 71A1-like (LOC131312986), transcript variant X2, mRNA
TRINITY_DN905_c0_g2_i2	KF024464.1	97.46 8	Arbutus unedo clone 99638 microsatellite sequence
TRINITY_DN335_c0_g2_i1	MW358095.1	89.72 8	Rhododendron x hybridum chalcone synthase (CHS) gene, complete cds
TRINITY_DN335_c0_g1_i5	AY123765.1	91.25	Vaccinium myrtillus chalcone synthase (chs) mRNA, partial cds
TRINITY_DN1442_c0_g1_i1 2	MH321460.1	85.11 5	Vaccinium corymbosum chalcone synthase (CHS) mRNA, partial cds
TRINITY_DN10367_c0_g1_i 1	OZ228609.1	95.45 5	Calluna vulgaris genome assembly, chromosome: 6
TRINITY_DN4014_c0_g1_i2	KJ830748.1	90.50 3	Vaccinium corymbosum O-methyltransferase mRNA, complete cds
TRINITY_DN331_c0_g1_i2	XM_058326945.1	83.17 8	PREDICTED: Rhododendron vialii chitinase-like protein 1 (LOC131300915), transcript variant X1, mRNA
TRINITY_DN4853_c0_g1_i8	XM_058326354.1	85.15 5	PREDICTED: Rhododendron vialii small heat shock protein, chloroplastic-like (LOC131300481), transcript variant X1, mRNA
TRINITY_DN2237_c0_g1_i7	AM392492.1	84.49 8	Rhododendron kanahirai partial cpsHSP gene for chloroplast small heat shock protein
TRINITY_DN2237_c0_g1_i8	AM392497.1	84.49 8	Rhododendron noriakianum partial cpsHSP gene for chloroplast small heat shock protein
TRINITY_DN3508_c0_g1_i1 4	OZ224585.1	97.40 3	Calluna vulgaris genome assembly, chromosome: 6
TRINITY_DN1311_c0_g1_i5	XM_058338857.1	88.17 6	PREDICTED: Rhododendron vialii protein HEAT-STRESS-ASSOCIATED 32 (LOC131311405), mRNA
TRINITY_DN700_c2_g1_i1	XM_058346429.1	83.25 4	PREDICTED: Rhododendron vialii 18.1 kDa class I heat shock protein-like (LOC131316907), mRNA
TRINITY_DN293_c0_g1_i11	XM_058335467.1	82.84 5	PREDICTED: Rhododendron vialii 18.5 kDa class I heat shock protein-like (LOC131308529), mRNA
TRINITY_DN293_c0_g1_i14	FM172931.1	82.78 5	Rhododendron rubropunctatum partial CTsHSP2 gene for cytosolic class I small heat shock protein type 2
TRINITY_DN685_c0_g1_i4	KU522304.1	77.24 4	Vaccinium corymbosum small heat shock protein HSP17.7 mRNA, complete cds
TRINITY_DN348_c0_g1_i1	XM_058348784.1	88.20 6	PREDICTED: Rhododendron vialii heat shock cognate 70 kDa protein 2-like (LOC131318808), mRNA
TRINITY_DN689_c0_g4_i1	OZ228606.1	96.97	Calluna vulgaris genome assembly, chromosome: 3
TRINITY_DN689_c0_g1_i6	OZ228605.1	100	Calluna vulgaris genome assembly, chromosome: 2
TRINITY_DN0_c0_g1_i9	OZ224586.1	95.23 8	Calluna vulgaris genome assembly, chromosome: 7
TRINITY_DN2250_c0_g1_i2 0	OZ228610.1	97.36 8	Calluna vulgaris genome assembly, chromosome: 7



TRINITY_DN1078_c0_g2_i1	ON260905.1	89.65 4	Vaccinium corymbosum phenylalanine ammonia-lyase (PALd) mRNA, complete cds
TRINITY_DN1078_c0_g1_i1	MW735988.1	88.77	Rhododendron fortunei phenylalanine ammonia lyase (PAL) gene, complete cds
TRINITY_DN6767_c0_g1_i7	XM_058348634.1	88.02 7	PREDICTED: Rhododendron vialii phenylalanine ammonia-lyase-like (LOC131318698), mRNA
TRINITY_DN6767_c0_g1_i8	OK624454.1	87.83 6	Vaccinium dunalianum phenylalanin ammonia-lyase 3 (PAL3) mRNA, complete cds
TRINITY_DN1000_c0_g1_i5	XM_058328336.1	86.29 6	PREDICTED: Rhododendron vialii cell number regulator 2-like (LOC131301855), mRNA
TRINITY_DN1268_c0_g1_i1	XM_058342680.1	87.20 5	PREDICTED: Rhododendron vialii trans-resveratrol di-O-methyltransferase-like (LOC131314185), mRNA

^aPident= Percent identity; ^bFull annotation results available in Table S1 in Supplemental Materials

Convergent Findings from Core nucleotide and WGS Analyses

Together, the Core nucleotide transcriptome and WGS BLAST results provide complementary insights—while Core nucleotide highlights actively expressed genes under experimental conditions, the WGS mapping enables broader comparisons of gene conservation and genomic context across Ericaceae species. We searched for overlapping *Arctostaphylos glauca* transcripts related to secondary metabolism identified in both Whole Genome Shotgun (WGS) BLAST results and core nucleotide hits. Table 5 highlights transcripts with putative roles in secondary metabolite biosynthesis that were recovered in both the WGS and core nucleotide, annotated based on their closest Ericaceae homologs. Each gene is listed with its Trinity transcript ID, proposed function, and percent identity to the Ericaceae family hit. Overlapping hits across both datasets reinforce the reliability of these candidates for further study in the context of stress tolerance and ecological resilience.

A comparison of the WGS and core nucleotide transcriptome data revealed a shared set of genes that play essential roles in both primary and secondary metabolic processes. These findings reflect functional conservation rather than one-to-one transcript overlap. Notably, both datasets included heat shock proteins, which are indicative of conserved stress response mechanisms in *Arctostaphylos glauca*. The presence of Berberine Bridge Enzyme-like proteins (BBE13 and BBE15) highlights the plant's potential to synthesize benzylisoquinoline-like alkaloid precursors—an underexplored but ecologically important class of phenolic compounds. Among the most functionally significant shared annotations was trans-cinnamate 4-monooxygenase, a cytochrome P450 enzyme (CYP73A subfamily) involved in the hydroxylation of cinnamic acid, a key early step in the phenylpropanoid pathway. This overlap underscores the consistency between the reference-based and de novo approaches and supports the relevance of these genes in stress adaptation and secondary metabolite biosynthesis in this chaparral-adapted species.

Heat shock proteins (HSPs) were consistently recovered in both the transcriptome (core nucleotide) and genome (WGS) datasets of *Arctostaphylos glauca*, suggesting their potentially critical role in environmental stress adaptation. Among the annotated hits, several variants of the HSP70 family were detected, including both cytosolic and chloroplastic isoforms. The presence of distinct isoforms in different cellular compartments reflects the diverse roles HSPs play in protecting cellular structures under thermal and oxidative stress. HSP70 proteins function as molecular chaperones, stabilizing proteins and preventing aggregation during abiotic stress, such as heat and drought, which are characteristic of the fire-prone semi-arid chaparral habitat where *A. glauca* thrives. Chloroplastic small heat shock proteins were matched with known chloroplast-targeted HSPs (e.g., HS22C_PETHY, HS906_ARATH). Cytosolic isoforms, such as HSP83 and HSP80 (e.g., HSP80_SOLLC, HSP83_IPONI), were also observed in multiple high-confidence hits.

These findings suggest that *Arctostaphylos glauca* expresses multiple functional variants of heat shock proteins in different cellular compartments, which could enhance its resiliency to thermal stress in fire-prone, semi-arid ecosystems. The detection of multiple orthologs across species, including those with strong sequence similarity to *Rhododendron williamsianum*, *Vaccinium darrowii*, and *Camellia sinensis*, highlights conserved stress resilience mechanisms among Ericaceae and related lineages. The redundancy and diversity of these HSP genes



in *A. glauca* may be key to its ecological persistence in rapidly changing environments, offering a molecular basis for further exploration of climate-resilient traits in native California shrubs.

Table 5. The overlapping BLAST hits shared between core nucleotide and WGS are summarized in the table. %ID= Percent Identity

Qseqid	sseqid_WGS	%IDWGS	Hit_Name_WGS	sseqid_CoreNT	%IDCoreNT	Hit_Name_CoreNucleotide
TRINITY_DN689_c0_g1_i6	JAHSPX010000014.1	100	A. glauca SCAF_0014, WGS	OZ228605.1	100	Calluna vulgaris genome, chrom: 2
TRINITY_DN4868_c0_g1_i1	JAHSPW020000001.1	99.404	A. glauca SCAF_1, WGS	XM_058341007.1	86.987	PRED: Rhododendron vialii berberine bridge enzyme-like 15. mRNA
TRINITY_DN1448_c0_g1_i1	JAHSPW020000011.1	100	A. glauca SCAF_11, WGS	XM_058371471.1	93.023	PRED: Rhododendron vialii cytochrome P450 736A117-like , mRNA
TRINITY_DN3508_c0_g1_i14	JAHSPW020000012.1	100	A. glauca SCAF_12, WGS	OZ224585.1	97.403	Calluna vulgaris genome, chrom: 6
TRINITY_DN4853_c0_g1_i8	JAHSPW020000013.1	98.621	A. glauca SCAF_13, WGS	XM_058326354.1	85.155	PRED: Rhododendron vialii small heat shock protein, chloroplastic-like mRNA
TRINITY_DN685_c0_g1_i4	JAHSPW020000035.1	96.735	A. glauca SCAF_35, WGS	KU522304.1	77.244	Vaccinium corymbosum small heat shock protein HSP17.7 mRNA, complete cds
TRINITY_DN689_c0_g4_i1	JACYLG010002524.1	100	R. molle LG01, WGS	OZ228606.1	96.97	Calluna vulgaris genome assembly, chrom: 3

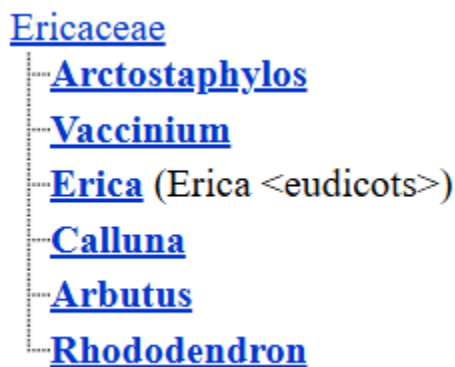


Fig. 6. Phylogenetic Tree generated vis NCBI Common Tree showing relationships between the genera with hits in Ericaceae for Core nucleotide and WGS hits.

Conclusion

This study provides new insight into the genetic basis of stress resilience and secondary metabolite biosynthesis in *Arctostaphylos glauca*, a keystone species in California's semi-arid chaparral ecosystem. By integrating transcriptome data with comparative BLAST analyses against whole genome scaffolds and curated nucleotide references, we identified candidate genes involved in flavonoid, phenylpropanoid, and lignin biosynthesis, as well as stress response pathways. Key enzymes such as CHS1, PAL, and C4H were not only highly expressed but also matched with conserved orthologs in closely related Ericaceae species, reinforcing their potential functional roles. The detection of multiple heat shock protein variants, including chloroplastic and cytosolic forms, highlights the importance of proteostasis under heat and drought stress. Together, these findings suggest



that the co-expression of specialized metabolite pathways and molecular stress response networks may underlie *A. glauca*'s ecological success in extreme environments. This work contributes foundational genomic knowledge for a non-model, fire-adapted species and supports ongoing efforts in conservation genomics, native plant propagation, and restoration strategies. By integrating curated functional annotation with phylogenetically informed genome and nucleotide comparisons, this transcriptome resource is designed for reuse across research, conservation, and instructional contexts. Further studies should compare expression in separate, replicated samples from different tissue sources and under different burn conditions where possible.

Acknowledgements. The authors would like to acknowledge the Gold Creek Committee, Los Angeles Community College District, the office of Career and Technical Education at Pierce College, and the Amgen Biotech Experience program for providing support. Thank you to Jillian M. Ford for technical writing assistance.

Disclosures. The authors declare no conflicts of interest.

Supplemental Materials. Link to be provided on publication

References

- [1] J. L. Howard, "Arctostaphylos glauca," *Fire Effects Information System*, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory, 1993. [Online]. Available: <https://www.fs.usda.gov/database/feis/plants/shrub/arcglu/all.html>
- [2] I. Abrahamson, "Arctostaphylos manzanita," *Fire Effects Information System*, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory, 2014. [Online]. Available: <https://www.fs.usda.gov/database/feis/plants/shrub/arcman/all.html>. Accessed: Apr. 1, 2024.
- [3] California Native Plant Society, "Arctostaphylos glauca shrubland alliance," *A Manual of California Vegetation Online*. [Online]. Available: <https://vegetation.cnps.org/alliance/117>. Accessed: Apr. 1, 2024.
- [4] L. Singer, "Arctostaphylos: The winter wonder," *The Poppy Print*, Theodore Payne Foundation, 2010. [Online]. Available: <https://www.theodorepayne.org/poppyprint/10-1.pdf>
- [5] S. Senn *et al.*, "The functional biogeography of eDNA metacommunities in the post-fire landscape of the Angeles National Forest," *Microorganisms*, vol. 10, no. 6, Art. no. 1218, Jun. 2022, doi: 10.3390/microorganisms10061218.
- [6] G. Guerriero *et al.*, "Production of plant secondary metabolites: Examples, tips and suggestions for biotechnologists," *Genes*, vol. 9, no. 6, Art. no. 309, Jun. 2018, doi: 10.3390/genes9060309.
- [7] M. K. Anderson, *Tending the Wild: Native American Knowledge and the Management of California's Natural Resources*. Berkeley, CA, USA: University of California Press, 2005.
- [8] S. A. Barrett and E. W. Gifford, *Miwok Material Culture*. Milwaukee, WI, USA: Public Museum of the City of Milwaukee, Bull. 2, no. 4, pp. 237–525, 1933.
- [9] Z. Peng and Z. Wang, *State-of-the-Art RNA Sequencing for Drug Discovery*. White Paper, BGI Global Genomics Services, May 2018. [Online]. Available: <https://www.bgi.com/wp-content/uploads/sites/2/2018/06/BGI-RNA-Transcriptome-Public-white-paper.pdf>



- [10] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: Summarize analysis results for multiple tools and samples in a single report," *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Oct. 2016, doi: 10.1093/bioinformatics/btw354.
- [11] Y. Chen, Y. Chen, C. Shi, Z. Huang, Y. Zhang, S. Li, *et al.*, "SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data," *GigaScience*, vol. 7, no. 1, pp. 1–6, Jan. 2018, doi: 10.1093/gigascience/gix120.
- [12] M. G. Grabherr, B. J. Haas, M. Yassour, *et al.*, "Full-length transcriptome assembly from RNA-Seq data without a reference genome," *Nature Biotechnology*, vol. 29, no. 7, pp. 644–652, Jul. 2011, doi: 10.1038/nbt.1883.
- [13] M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov, "BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring eukaryotic, prokaryotic, and viral genomes," *Molecular Biology and Evolution*, vol. 38, no. 10, pp. 4647–4654, Oct. 2021, doi: 10.1093/molbev/msab199.
- [14] R. Patro, G. Duggal, M. Love, *et al.*, "Salmon provides fast and bias-aware quantification of transcript expression," *Nature Methods*, vol. 14, no. 4, pp. 417–419, Apr. 2017.
- [15] D. M. Bryant, K. Johnson, T. DiTommaso, T. Tickle, M. B. Couger, D. Payzin-Dogru, T. J. Lee, N. D. Leigh, T. H. Kuo, F. G. Davis, J. Bateman, *et al.*, "A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors," *Cell Reports*, vol. 18, no. 3, pp. 762–776, Jan. 2017, doi: 10.1016/j.celrep.2016.12.063.
- [16] J. Hsieh, "find_transcript_ids," GitHub repository, Feb. 2024. [Online]. Available: https://github.com/PiercePlantScience/find_transcript_ids
- [17] K. Kryukov, "FASTA Splitter," 2024. [Online]. Available: <https://kirill-kryukov.com/study/tools/fasta-splitter/>. Accessed: Jan. 22, 2026.
- [18] R. A. Dixon and N. L. Paiva, "Stress-induced phenylpropanoid metabolism," *Plant Cell*, vol. 7, no. 7, pp. 1085–1097, Jul. 1995, doi: 10.1105/tpc.7.7.1085.
- [19] G. Forkmann and S. Martens, "Metabolic engineering and applications of flavonoids," *Current Opinion in Biotechnology*, vol. 12, no. 2, pp. 155–160, Apr. 2001, doi: 10.1016/S0958-1669(00)00198-2.
- [20] L. Ralston, *et al.*, "Structural basis for the activation of phenylalanine in the biosynthesis of cyanogenic glycosides," *Plant Cell*, vol. 17, no. 4, pp. 1098–1106, Apr. 2005, doi: 10.1105/tpc.104.029678.
- [21] C. Gomez, *et al.*, "Evolution of a novel phenolic pathway for pollen development," *Science*, vol. 325, no. 5948, pp. 1688–1692, Sep. 2009, doi: 10.1126/science.1174094.
- [22] J. Bogs *et al.*, "Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves," *Plant Physiology*, vol. 139, no. 2, pp. 652–663, 2005.
- [23] J. Huang, *et al.*, "Functional and evolutionary analysis of terpene synthases in magnoliids," *Journal of Biological Chemistry*, vol. 285, no. 30, pp. 23341–23349, Jul. 2010, doi: 10.1074/jbc.M110.122333.



[24] J. Wang, *et al.*, "Functional characterization of cinnamate 4-hydroxylase and caffeic acid O-methyltransferase in *Brachypodium distachyon*," *BMC Plant Biology*, vol. 14, Art. no. 274, Sep. 2014, doi: 10.1186/s12870-014-0274-1.

[25] C. Forni, *et al.*, "Phytochemical characterization of *Arctostaphylos uva-ursi* (L.) Sprengel leaves," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 149, pp. 403–409, Feb. 2018, doi: 10.1016/j.jpba.2017.11.042.