

Article

Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in *Caryopteris* × *clandonensis*

Manfred Ritz [†] , Nadim Ahmad [†] , Thomas Brueck ^{*}  and Norbert Mehlmer ^{*} 

Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich (TUM), 85748 Garching, Germany; manfred.ritz@tum.de (M.R.); nadim.ahmad@tum.de (N.A.)

* Correspondence: brueck@tum.de (T.B.); norbert.mehlmer@tum.de (N.M.)

† These authors contributed equally to this work.

Abstract: Enzymatic terpene functionalization is an essential part of plant secondary metabolite diversity. Within this, multiple terpene-modifying enzymes are required to enable the chemical diversity of volatile compounds essential in plant communication and defense. This work sheds light on the differentially transcribed genes within *Caryopteris* × *clandonensis* that are capable of functionalizing cyclic terpene scaffolds, which are the product of terpene cyclase action. The available genomic reference was subjected to further improvements to provide a comprehensive basis, where the number of contigs was minimized. RNA-Seq data of six cultivars, Dark Knight, Grand Bleu, Good as Gold, Hint of Gold, Pink Perfection, and Sunny Blue, were mapped on the reference, and their distinct transcription profile investigated. Within this data resource, we detected interesting variations and additionally genes with high and low transcript abundancies in leaves of *Caryopteris* × *clandonensis* related to terpene functionalization. As previously described, different cultivars vary in their modification of monoterpenes, especially limonene, resulting in different limonene-derived molecules. This study focuses on predicting the cytochrome p450 enzymes underlying this varied transcription pattern between investigated samples. Thus, making them a reasonable explanation for terpenoid differences between these plants. Furthermore, these data provide the basis for functional assays and the verification of putative enzyme activities.

Keywords: terpene biosynthesis; cytochrome p450; *Caryopteris* × *clandonensis*; long read sequencing; transcriptomics; chemical diversity; volatile compound



Citation: Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Differential RNA-Seq Analysis Predicts Genes Related to Terpene Tailoring in *Caryopteris* × *clandonensis*. *Plants* **2023**, *12*, 2305. <https://doi.org/10.3390/plants12122305>

Academic Editors: Andreas W. Ebert and Nam-Soo Kim

Received: 25 April 2023

Revised: 17 May 2023

Accepted: 7 June 2023

Published: 13 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Caryopteris × *clandonensis* is an ornamental plant, also known as “Bluebeard”, which is phylogenetically classified in the *Lamiaceae* family. It is easily cultivated and rich in volatile compounds. These, and other molecules detected and described, are terpenes, e.g., α -copaene, limonene, or δ -cadinene [1], terpene derivatives, e.g., keto-glycosides, clandonosides, and harpagides [2], as well as the pyrano-juglon derivative α -caryopteron [3]. The species’ essential oil was found to display mosquito-repellent activity; however, the active agent for this mode of action was not yet detected [4]. The *Lamiaceae* family is known to harbor an interesting and valuable profile in secondary metabolites, including terpenoids, flavonoids, and phenylpropanoids [5–7]. These compounds play important roles in the plant’s interaction with its environment [8,9] as for the defense against abiotic and biotic stresses [10]. They also harbor potential in pharmaceutical or industrial applications, as seen for taxol [11], menthol [12], malvidin [13], isoliquiritigenin [14] or umbelliferone [15]. In general, terpenes and terpenoids are a molecule class, which is produced in vast varieties by flowering plants [16] and is involved in a wide range of biological activities. Essential oils and their monoterpenes, such as α -pinene and limonene, were investigated in terms of their anti-inflammatory and virucidal activity in recent studies [17–19]. Moreover, other terpenoids employ antibacterial properties [20] while others act as insecticides [4], are used

as allelochemicals [21], or as attractants for pollinators [22]. The backbone of plant-derived terpenes is produced via the mevalonate pathway. For this, the precursors dimethylallyl diphosphate (DMAPP) and the functional isomer isopentenyl pyrophosphate (IPP) can be connected via isoprenyl diphosphate synthases (IDS) to form larger units of terpenes. IPP consists of five C-atoms (hemiterpene) whereas, through condensation of IPP and DMAPP via IDS monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), and higher terpene structures are built [23]. Further tailoring of these basic terpenes is conducted by terpene synthases (TPS) and cytochrome p450 enzymes (CYPs). Plant TPS mediate complex carbocation reactions, resulting in various cyclic structures of higher terpenes [24,25]. These can be divided into eight subfamilies (TPSa-h) which can be clade- or even species-specific [26]. The first step in tailoring monoterpenes is hydroxylation. Subsequently, CYPs are mediating a plethora of further reactions to enhance the functionalization (carboxylation, acetylation or forming peroxides) [27,28]. Due to their promiscuity towards substrates, only a few enzymes are necessary to yield various terpenoid structures and, therefore, differences in their functions and modality [29]. Multiple sequences of different source organisms are available in curated databases [30,31]. These allow easy access to the genetic information on these enzymes. With CYPs occurring in all living organisms [30], the enzymes, similarly to TPS, are divided for better identification, whereas specific CYP families are reserved for each type of organism. Plant CYP families can be found in CYP71-99 and CYP701-999, and in a four-digit scheme from CYP7001-9999 [32]. The categorization into these classes is dependent on sequence similarity. The same family (Arabic number) needs matching amino acids $\geq 40\%$ and the subfamily (Arabic letter) $\geq 55\%$ [33]. Therefore, the CYP76S40 [34] is the 40th individual enzyme from the CYP76S subfamily and the CYP76 family. This way, after annotation, contaminating sequences can be discarded solely due to their classification in a non-plant CYP family.

One approach to elucidate variations in the enzymatic makeup and investigate the sequences underlying terpene diversity is to compare differentially expressed gene (DEG) products at a quantitative level using modern bioinformatics tools. Differences in the metabolite profile exist during different stages of plant growth [35]. Different genes are regulated from seedlings to mature plants to translate their genomic information into proteins and interact in plant differentiation, protection or communication, depending on their developmental state [36]. During plant breeding, deletion, duplications, mutations or fragmentations can occur. Therefore, a distinct set of genes varies in its nucleotide code and their transcription or translation rate, resulting in different phenotypes in the mature plant [37]. The data can be levied and evaluated regarding efficacy to investigate these differences. The number of transcripts does not solely result in higher protein outcome, but also in, respectively, higher concentrations of secondary metabolites. Therefore, differential expression analysis can identify genes or gene products responsible for either the stress response mechanisms observed for abiotic stressors, such as drought or radiation, or as has been shown for biotic stressors, such as pests and plant reactions to herbivores [38]. Typical DEG experiments harness the up- and down-regulation of genes after induction or shock, e.g., during exposure to chemicals [39] or different environments [40]. Another possibility is the investigation of specific traits of plant cultivars due to their variations between hybrid plants [41]. Previously, the variations in *Caryopteris* \times *clandonensis* volatile compound setup was investigated, and a difference in the synthesis of limonene-derived molecules (LDM) was observed [1]. The cultivar Dark Knight was detected to harbor a low amount, whereas Pink Perfection shows high amounts of LDM. These variations were discovered without a distinct change in their TPS or CYP makeup.

To that end, we show that the identification of terpene variety between different plant cultivars can be pursued on a molecular level using a quantitative bioinformatics method such as RNA-Seq analysis. Furthermore, we focus on terpene functionalization enzymes, especially cytochrome p450 enzymes, to elucidate the mechanisms behind the variations in monoterpene modifications as seen for limonene [1].

2. Results and Discussion

2.1. RNA Sequencing and Mapping Quality

Samples subjected to short-read sequencing were taken from leaves of six *Caryopteris × clandonensis* cultivars known to show differences in their LDM profile, Dark Knight (DK), Grand Bleu (GB), Good as Gold (GG), Hint of Gold (HG), Sunny Blue (SB), and Pink Perfection (PP). Sequencing was performed using an Illumina NovaSeq platform, which generated about 20 million raw reads in bases for each sample. The reads were processed to remove low-quality reads, bases, and adapter sequences, resulting in the clean reads used for downstream analysis. After this purification step, a loss of 5.0 to 14.9 million bases was seen between the samples. In Table 1, the run as well as cleaning and mapping statistics are summarized. The Q20 and Q30 scores indicate the sequencing quality, with Q30 indicating a lower error rate than Q20. This experiment's high Q20 and Q30 scores suggest that the sequencing quality was highly sufficient, with only a few sequencing errors. Moreover, the clean reads exhibit a slight increase in quality scores, persistent throughout all samples.

Table 1. Statistics of short Illumina reads used for mapping on the reference genome (NCBI SAMN32308290 (Pink Perfection, PP)). A paired-end run was employed on a NovaSeq6000 SP (2 × 150 bp) for sequencing.

<i>Caryopteris × clandonensis</i> Cultivar		Raw Reads in Bases		Q20 in %	Q30 in %	Clean Reads in Bases		Q20 in %	Q30 in %	Totally Mapped in %	Uniquely Mapped in %
		Unique	Duplicate			Unique	Duplicate				
Dark Knight	R1	24,501,785	19,238,555	99.95	94.76	13,072,273	29,945,355	99.99	95.08	87.8	79.3
	R2	26,380,719	17,359,621	99.25	87.90	16,204,470	26,813,158	99.46	88.25		
Grand Bleu	R1	17,917,215	51,971,129	99.85	93.75	11,552,426	57,659,626	99.98	94.21	85.8	76.8
	R2	18,808,258	51,080,086	99.51	92.12	13,260,446	55,951,606	99.68	92.42		
Good as Gold	R1	22,797,327	27,074,692	99.60	94.52	13,160,322	31,359,084	99.95	95.08	86.7	75.4
	R2	25,142,438	24,729,581	99.35	89.41	16,112,061	28,407,345	99.54	89.76		
Hint of Gold	R1	20,547,645	20,953,229	99.89	94.67	15,165,071	33,935,373	99.98	95.06	86.4	77.2
	R2	23,044,700	18,456,174	99.38	88.40	18,084,814	31,015,630	99.56	88.81		
Sunny Blue	R1	20,535,582	25,022,034	99.96	94.96	13,908,152	27,181,140	99.99	95.35	87.0	80.5
	R2	22,771,085	22,786,531	99.39	88.30	16,573,745	24,515,547	99.56	88.60		
Pink Perfection	R1	25,751,312	28,610,858	99.96	94.23	12,295,625	26,046,846	99.99	94.60	87.7	82.0
	R2	29,512,685	24,849,485	99.40	90.42	14,649,539	23,692,932	99.58	90.70		

The available genome sequences from *Caryopteris × clandonensis* PP [1] were subjected to further cleaning and improvement steps to curb the influence of contamination. A binning algorithm, MetaBAT2 [42], usually used for metagenomic data, was used on the long-read assembly of the genome and differentiated into 40 bins. The completeness and contiguity were checked and, in summary, the 782 scaffolds/848 contigs, which add up to 344 Mb with a genome completeness score of 96.8%, were reduced to 53 scaffolds/88 contigs, which add up to 298 Mb and a BUSCO score of 96.5%. The utilized BUSCO gene sets belonged to the closest affiliate *Eudicotidae*. Detailed information can be found in Table S1. This refined genome was used as a reference for mapping the short-read sequences. A preliminary mapping of DK transcripts on the respective long-read genomic data, compared to mapping the transcripts on the PP genomic data, revealed an increased assignment of unique reads. Thus, the genome of *Caryopteris × clandonensis* PP was chosen as a mapping reference for both cultivars, DK and PP, resulting in a more comprehensive downstream analysis. The exact mapping counts for the different methods can be found in Table S2.

The percentages of reads mapped to the reference genome, as seen in Table 1, indicate the data accuracy and low presence of contaminating DNA. The amount of uniquely mapped reads is also an important metric, as it indicates the proportion of reads that map to a unique location in the reference genome. A high percentage of uniquely mapped reads (greater than 70%) is desirable, reducing the possibility of mapping errors or ambiguous mapping locations [43]. In our setting, we were able to accurately map between 85.8% and 87.8% of the sequences, indicating that a large proportion of the reads were successfully located on the provided genome. Furthermore, the percentage of uniquely mapped reads ranged from 75.4% to 82.0%, which is reasonably high and suggests that the quality of the sequencing reads was sufficient to allow for exact mapping and is suited for downstream

analysis. The observed duplication rates varied between 5.7% and 11.3%, and are well-known in plant transcript mapping due to transcript isoforms [44].

2.2. Identification of DEG

To identify the mechanism behind the modification of LDM, we wanted to focus on the DEGs between the cultivars of *Caryopteris × clandonensis*. Therefore, the mapping data were subset and pooled into highly LDM-positive (SB, PP) and highly LDM-negative (DK, GB) cultivars. The cultivars GG and HG were neither highly LDM-positive nor highly LDM-negative, therefore both were disregarded during the initial DEG analysis. From the 29,210 predicted genes in the mapping reference, 23,477 were observed to map in all investigated sets. The DEGs were filtered using a log₂ fold-change cutoff of absolute values greater than 1, and an adjusted *p*-value of a minimum of 0.05, thereby the values for each cultivar were transcribed at least two-fold. The values fitting these parameters are highlighted in green; those which were disregarded during further analysis, because of not fitting the parameters, are shown in red. Compared to the genes close to the middle, there are a few genes with high fold-changes in LDM-positive plants, compared to LDM-negative and those with significantly higher or lower transcript abundance. After filtering the DEGs between LDM-positive and LDM-negative cultivars, 3305 genes were identified, as seen in Figure 1A. For 100 genes, no Pfam class [45] and, for a further 168, no EggNOG [46] description, could be assigned. Regarding the DEGs, a closer look reveals the 20 most diverged genes, which can be seen in Figure 1B,C. Half of the annotated genes are still uncharacterized, or their distinct function is unknown, according to the cluster of orthologous groups. Interestingly, the genes associated with metal transport and metal binding are differentially transcribed, as seen for g4372, g9694, and g8497. These functions are known to be responsible for catalyzing redox reactions in plants [47,48]. Examining DEGs further, g14432 is associated with the protein argonaute family and g1887 is a zinc finger-like protein, whereas g3464 is a thioredoxin/disulfide isomerase. These proteins regulate biological processes [49], as well as responses to abiotic stresses such as drought stress [50,51]. In general, these DEGs describe the effects on the primary metabolism and stress response of plants; however, they do not show any direct participation in tailoring secondary metabolites within the plants. CYPs, in particular, are iron-binding; however, a connection between the upregulation of metal-transporting proteins and CYPs cannot be drawn from this data. The biosynthesis of LDM is not artificially induced in one cultivar or silenced in the other. Thus, a specific and significant transcription of related terpenetailoring genes cannot be observed. To elucidate these mechanisms, it is necessary to take a closer look into the DEGs of CYPs [28,52].

2.3. Terpene Tailoring through CYPs between Plant Cultivars

The identified 3305 DEGs can be further filtered into genes related to CYPs due to conserved domains and the corresponding CYP Pfam class. Here, the domain PF00067 was integrated into IPR001128. Both domains are indicators for sequences associated with the cytochrome p450 superfamily (IPR036396) [53]. This homology-based search allowed the identification of 70 putative sequences with different total lengths. Assuming a minimum size of 29 kDa for a CYP, 61 genes remain. From a statistical point of view, the average size of this pool amounts to a median of 1485 nucleotides, corresponding to the average size of a translated protein of 54.5 kDa. This is also reported in the literature, with an average plant CYP molecular mass between 45 and 62 kDa [54,55]. In regards to the identification of LDM-modifying enzymes, this subset is necessary to obtain a detailed overview into CYPs. These enzymes are known to play a huge part in terpene diversity in plants [56]. They are able to catalyze the hydroxylation of different backbones due to their substrate promiscuity [29,57]. Therefore, the transcript abundance of specific CYPs may reveal the mechanism behind LDM variances in this plant.

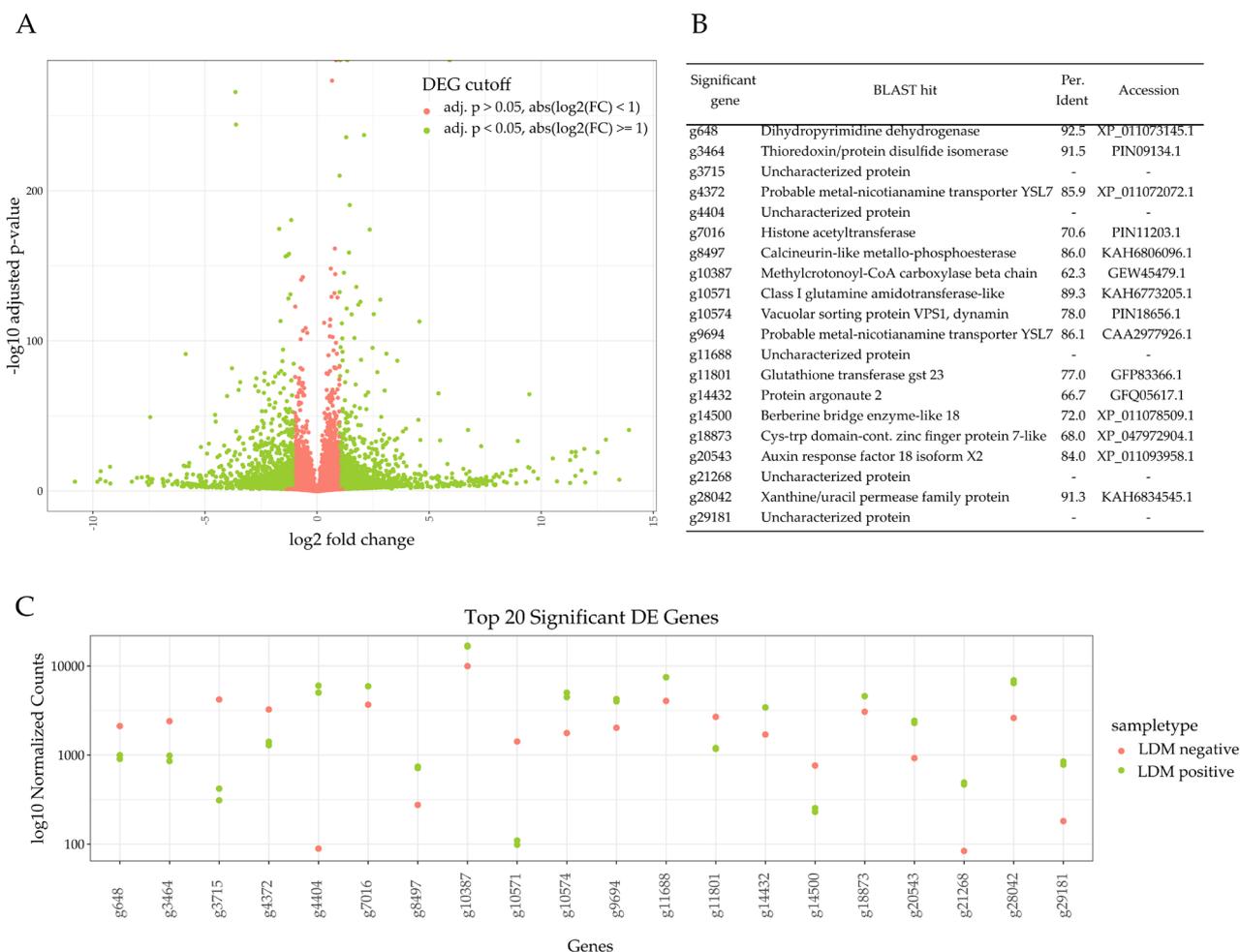


Figure 1. Differential expressed genes (DEGs) of *Caryopteris* × *clandonensis* cultivars highly producing limonene-derived molecules (LDM-positive) and cultivars which produce lower amounts of LDM (LDM-negative). Cultivars used for LDM-positive subset: Sunny Blue and Pink Perfection, and for LDM-negative subset Dark Knight and Grand Bleu. (A) The volcano plot of DEG was identified between the LDM-positive vs. LDM-negative plant cultivar subsets. Absolute log₂ fold-change cutoff was set to 1 and an adjusted *p*-value of 0.05 was used to assign the DEGs; values fitting these parameters are highlighted in green and those which were disregarded during further analysis are shown in red. (B) Top 20 most significantly transcribed genes and their respective description, including BLAST search percentage identity and determined accession for the putative assignment. (C) log₁₀ normalized counts of the top 20 significant DEG in this setup. Genes from LDM-positive samples are displayed in green, those corresponding to LDM-negative samples are highlighted in red.

Out of all the 23,477 mapped genes, 221 CYPs were detected, whereas 61 showed differences in transcript abundance. In Figure 2, all identified CYPs are visualized in an unrooted phylogenetic tree. CYPs with high transcript abundance in LDM-positive cultivars are highlighted in green, whereas CYPs with low transcript abundance are represented in red.

To allocate the putative CYPs to their distinct family or subfamily, the Pfam-classified CYP sequences were subjected to a BLAST search using a custom CYP database [54]. The sequences were assigned to the same subfamily if the percent identity was above 55%, and to the same family if greater than 40%. Eight CYP clans were highlighted within the found enzymes, CLAN51, CLAN71, CLAN72, CLAN74, CLAN710, CLAN85, CLAN86, and CLAN 97. This highlights that the major classes 71 and 72 are found to be involved, primarily, in the terpene tailoring of different terpene classes [28]. For CYP71, a variety of monoterpene modifications are described [34,58–61]. In our setting, most DEGs were observed in this clan. The enzymes related to CYP72 are described as tailoring triterpenoids

the transcript abundance between all samples. In, the results of a comparative approach are visualized. The phylogenetic distance between the identified CYPs is shown in 3A. Three clusters can be differentiated, with the first seen in the upper part consisting of 4 genes (g25953, g25443, g578, g8489), the second in the middle (g3273, g10380, g27034, g27468, g27861), and the third cluster with 14 genes (g24222, g2313, g27787, g24257, g20804, g3860, g10700, g16684, g24219, g9390, g14070, g28342, g8554, g2205) at the bottom. In Section 3B, the fold-change between the cultivars is visualized; boxes marked with X were transcripts with no mapping results in the respective cultivar. The clusters do not share a similar transcript abundance pattern, nor do the genes that are closely related. However, investigating the recurring, fixed-length patterns inside the sequences led to the discovery of five motifs shared among all sequences. Figure 3C visualizes the motifs and their distribution in the sequence. The exact motif sequences are presented in Table S3. A closer look also reveals distinct recurring, CYP-specific domains [63]. The conserved regions were reviewed extensively [38] and can be confirmed in this dataset. Starting with the proline-rich membrane hinge (motif 8), which is part of the membrane anchor, another conserved motif, which is important for the correct function of CYPs, is the site for oxygen binding and activation, A/G-G-X-E/D-T-T/S (motif 3). This is followed by the E-R-R triad and P(E)R(F) domain. Furthermore, the heme-binding site, with cysteine as the main ligand to the heme, C-X-G (motif 2), which is necessary for the typical redox reaction of CYPs [64], as well as the ERR triad (motif 6) and the (P(E)R(F)) sequence (motif 6), can be differentiated among the discovered 10 motifs.

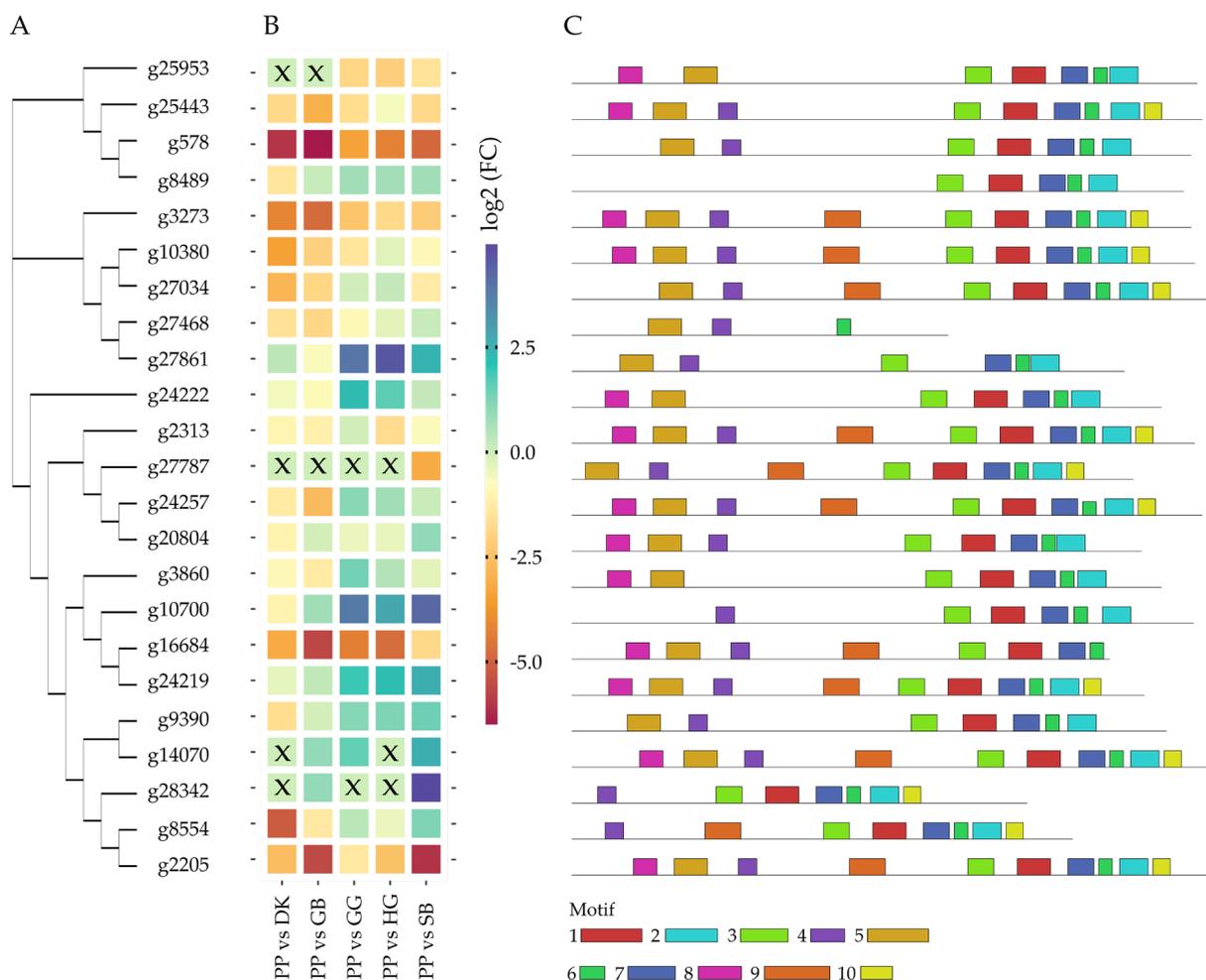


Figure 3. Analysis of differentially expressed cytochrome p450 enzymes (CYP) in different plant cultivars of *Caryopteris × clandonensis*, Dark Knight (DK), Grand Bleu (GB), Good as Gold (GG), Hint

of Gold (HG), Sunny Blue (SB), and Pink Perfection (PP). (A) Phylogenetic analysis of CYP sequences with highly abundant transcripts regarding limonene-derived molecules (LDM) within the cultivars, using Neighbor-joining method. (B) Heatmap of normalized transcript counts between distinct cultivars. X represents enzymes with no transcripts in respective cultivars. The color palette displays genes with high transcripts abundance in red to light-yellow colors, high transcript abundance is depicted in light-green to blue (C) Identification of recurring, fixed-length patterns (motifs) identified in LDM-positive transcripts. Motifs 1 to 10 are illustrated as colored boxes, to distinguish the motifs between the different genes. Sequences can be found in Table S3.

Regarding the production of LDM, the genes g8554, g27861, g10700, and g24222 show an interesting pattern compared to the highly LDM-positive cultivar PP, which makes them candidates for further functional characterization to prove their LDM-producing potential.

The candidate genes were further investigated in terms of their putative function. The initial estimates, using sequence and structural homology, consider g2422 and g8554 to be involved in the hydroxylation of cinnamic acid, whereas g27861 and g10700 display unknown activity towards flavonoids, sterols, and ferruginol. This substrate promiscuity is known for CYPs, as they are able to catalyze different ligands [57,65], thus making functional characterization using prokaryotic, yeast, or plant expression systems indispensable to support claims on putative functions.

3. Materials and Methods

3.1. Plant Material

Cultivars of *Caryopteris × clandonensis*, DK, GB, GG, HG, SB, and PP, were acquired from a local nursery (Foerstner Pflanzen GmbH, Bietigheim-Bissingen). DK and GB were investigated to show a highly LDM-negative profile, whereas SB and PP show a highly LDM-positive profile. GG and HG showed a non-conclusive profile in between. After growing to maturity in the open in a warm, moderate climate zone, healthy leaves were sampled and snap-frozen in liquid nitrogen and stored at -80°C until RNA preparation for RNA-Seq.

3.2. Genomic Resource

The reference genome of *Caryopteris × clandonensis* used in this study was obtained from NCBI SAMN32308290 (PP). The raw data were assembled as previously described [1] and subjected to further refinements. For further processing, the reference was cleared from possible contaminations, and scaled down from 783 contigs to 53 contigs using Metabat2 (v2.15) [42], keeping the genome completeness with 96.5% at a high level according to BUSCO (v5.3.2) [66] analysis (2326 BUSCO groups, lineage dataset: *Eudicotidae*). Gene model prediction was conducted using AUGUSTUS [67–70]. To detect repetitive sequences, such as tandem repeats or transposable elements, soft masking was employed using Red (v2018.09.10) [71].

3.3. RNA Preparation and Short Read Sequencing

High-quality RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's protocol. To ensure RNA integrity, the Bioanalyzer RNA 6000 assay kit (Agilent, Santa Clara, CA, USA) was employed to yield an average RNA Integrity Number of 7.7. The library preparation was performed using the Illumina stranded mRNA prep kit with IDT for Illumina UD Indexes, Plate A. Corresponding adapter was the Illumina Nextera Adapter (CTGTCTCTTATACATCT). Library preparation was performed according to the manufacturer's protocol with a shortened fragmentation time from 8 min (protocol) to 2 min (this study). Sequencing was performed at the Helmholtz Munich (HMGU) by the Genomics Core Facility on a NovaSeq6000 SP (2×150 bp). For each sample, two lanes were loaded and an average of 22 Mio fragments were yielded. The corresponding lanes of each sample were concatenated tail-to-head (v8.25) [72]. The combined short reads were subjected to comprehensive quality control steps. Every step was analyzed with FastQC (v0.11.9) [73] and the necessity of another trimming step was evaluated. Sequences shorter than 20 bp minimum length and with a

quality phred score beneath 20 were extracted from the paired-end read data. The Illumina Nextera Adapter was used to trim each read pair using Cutadapt (v.4.0) [74]. The first 10 bases were cut from the sequences, due to their sequence GC content, using Trimmomatic (v0.38) and headcrop parameter [75].

3.4. Mapping and Annotation of Aligned Reads

Refined short reads were mapped on the clean reference genome using STAR (v2.7.10b) [76], 140 bases were chosen as the length of the genomic sequence around annotated junctions. EggNOG (v2.1.5) [46,77] was employed to evaluate the function of the differentially expressed genes using Pfam, GO, and COG databases. MEME suite (v5.5.1) [78] was used for identification of motifs within sequences of interest. Visualizations were built in R. Except for STAR; all sequencing analyses were conducted using galaxy project [79]. Analysis was based on reference-based RNA-Seq data analysis [80,81]. The detection of CYPs was performed using a homology-based search, using the conserved domain PF00067, which was integrated to IPR001128. Both domains are indicators for a sequence association with the cytochrome p450 superfamily (IPR036396) [53]. CYP-family classification was performed using a BLAST search [82] and a custom database [83].

3.5. Evaluation of Differential Gene Expression between Aerial Plant Parts

Aligned transcripts were counted using FeatureCounts (v3.16) [84], normalized, and differentially investigated with DESeq2 (v1.34.0) [85–87]. An adjusted *p*-value below 0.05, and a fold-change greater than 2 and below 0.5, was used to determine the most differentially expressed genes in this dataset.

4. Conclusions

This study provides a basis for further CYP research in *Caryopteris × clandonensis*, especially regarding LDM. Furthermore, the reference genome was subjected to a cleaning step, resulting in a decrease from 782 scaffolds to 53 scaffolds. Six cultivars were subjected to an RNA analysis, which gradually neared the prediction of 4 possible LDM tailoring CYPs out of 24, which were differentially expressed, and showed high transcript abundance, compared to the other cultivars. Furthermore, the classification and phylogenetic analysis of all mapped CYPs were conducted and they showed a distinct clustering in CYP CLAN71 and 72. All essential and conserved motifs could be recognized within these sequences. However, experimentally focused research for functional characterization needs to be conducted in order to identify the exact predicted function of these enzymes. A further in silico step can include the prediction of docking and catalysis sites within a three-dimensional structural model, as well as through molecular dynamic techniques and free energy calculations [88,89].

In general, this approach can be used to detect further mechanisms and pathways in plants, which show valuable medicinal effects. The biotechnological production of artemisinin [90] and taxol [11] is a popular example of the possibilities in medicinal plant research. There are already several approaches used, which combine omics approaches to identify substances of interest [91–93].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12122305/s1>, Table S1: BUSCO assessment and assembly statistics, Table S2: Mapping statistics on the genomic reference of Pink Perfection, Table S3: Motif sequences of identified reoccurring patterns.

Author Contributions: Conceptualization, M.R., N.A. and N.M.; methodology, M.R. and N.A.; software, M.R., N.A. and N.M.; validation, M.R. and N.A.; formal analysis, M.R. and N.A.; investigation, M.R. and N.A.; resources, T.B.; data curation, M.R. and N.A.; writing—original draft preparation, M.R.; writing—review and editing, M.R., N.A., N.M. and T.B.; visualization, M.R.; supervision, N.M. and T.B.; project administration, N.M. and T.B.; funding acquisition, T.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry of Education and Research, grant number 031B0824A.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available in a publicly accessible repository. The refined genome data presented in this study are openly available at the National Center for Biotechnology Information (NCBI). BioSample accession number: Pink Perfection SAMN32308290.

Acknowledgments: The authors want to gratefully acknowledge the support of Christine Wurmser, (Chair of Animal Physiology and Immunology, TUM School of Life Sciences, Technical University of Munich) for her support in the library preparation and the handling of Illumina sequencing, and Foerstner Pflanzen GmbH, for providing plant materials. Furthermore, the authors want to acknowledge the support of the following colleagues at the Werner Siemens-Chair for Synthetic Biotechnology: Nathanael Arnold, Kevin Heieck, Zora Rerop, Selina Engelhart-Straub, and further colleagues for their support during conducting experiments and writing this manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Comparative Genome-Wide Analysis of Two Caryopteris × Clandonensis Cultivars: Insights on the Biosynthesis of Volatile Terpenoids. *Plants* **2023**, *12*, 632. [[CrossRef](#)] [[PubMed](#)]
2. Hannedouche, S.; Jacquemond-Collet, I.; Fabre, N.; Stanislas, E.; Moulis, C. Iridoid keto-glycosides from Caryopteris × Clandonensis. *Phytochemistry* **1999**, *51*, 767–769. [[CrossRef](#)]
3. Matsumoto, T.; Mayer, C.; Eugster, C.H. α -Caryopteron, ein neues Pyrano-juglon aus Caryopteris clandonensis. *Helv. Chim. Acta* **1969**, *52*, 808–812. [[CrossRef](#)]
4. Blythe, E.K.; Tabanca, N.; Demirci, B.; Bernier, U.R.; Agramonte, N.M.; Ali, A.; Baser, H.C.; Khan, I.A. Composition of the essential oil of Pink Chablis™ bluebeard (Caryopteris × clandonensis ‘Durio’) and its biological activity against the yellow fever mosquito *Aedes aegypti*. *Nat. Volatiles Essent. Oils* **2015**, *2*, 11–21.
5. Abdelaty, N.A.; Attia, E.Z.; Hamed, A.N.E.; Desoukey, S.Y. A review on various classes of secondary metabolites and biological activities of Lamiaceae (*Labiatae*) (2002–2018). *J. Adv. Biomed. Pharm. Sci.* **2021**, *4*, 16–31. [[CrossRef](#)]
6. Siciliano, T.; Bader, A.; Vassallo, A.; Braca, A.; Morelli, I.; Pizza, C.; De Tommasi, N. Secondary metabolites from Ballota undulata (*Lamiaceae*). *Biochem. Syst. Ecol.* **2005**, *33*, 341–351. [[CrossRef](#)]
7. Mimica-Dukic, N.; Bozin, B.; Mentha, L. Species (*Lamiaceae*) as Promising Sources of Bioactive Secondary Metabolites. *Curr. Pharm. Des.* **2008**, *14*, 3141–3150. [[CrossRef](#)]
8. Kliebenstein, D.J. Secondary metabolites and plant/environment interactions: A view through Arabidopsis thaliana tinged glasses. *Plant. Cell Environ.* **2004**, *27*, 675–684. [[CrossRef](#)]
9. Boncan, D.A.T.; Tsang, S.S.K.; Li, C.; Lee, I.H.T.; Lam, H.M.; Chan, T.F.; Hui, J.H.L. Terpenes and Terpenoids in Plants: Interactions with Environment and Insects. *Int. J. Mol. Sci.* **2020**, *21*, 7382. [[CrossRef](#)]
10. Holopainen, J.K.; Himanen, S.J.; Yuan, J.S.; Chen, F.; Stewart, C.N. Ecological functions of terpenoids in changing climates. *Nat. Prod.* **2013**, *1*, 2913–2940.
11. Wang, T.; Li, L.; Zhuang, W.; Zhang, F.; Shu, X.; Wang, N.; Wang, Z. Recent Research Progress in Taxol Biosynthetic Pathway and Acylation Reactions Mediated by Taxus Acyltransferases. *Molecules* **2021**, *26*, 2855. [[CrossRef](#)] [[PubMed](#)]
12. Kamatou, G.P.P.; Vermaak, I.; Viljoen, A.M.; Lawrence, B.M. Menthol: A simple monoterpene with remarkable biological properties. *Phytochemistry* **2013**, *96*, 15–25. [[CrossRef](#)]
13. Khoo, H.E.; Azlan, A.; Tang, S.T.; Lim, S.M. Anthocyanidins and anthocyanins: Colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr. Res.* **2017**, *61*, 1361779. [[CrossRef](#)]
14. Selvaraj, B.; Kim, D.W.; Huh, G.; Lee, H.; Kang, K.; Lee, J.W. Synthesis and biological evaluation of isoliquiritigenin derivatives as a neuroprotective agent against glutamate mediated neurotoxicity in HT22 cells. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127058. [[CrossRef](#)]
15. Mazimba, O. Umbelliferone: Sources, chemistry and bioactivities review. *Bull. Fac. Pharm. Cairo Univ.* **2017**, *55*, 223–232. [[CrossRef](#)]
16. Pichersky, E.; Raguso, R.A. Why do plants produce so many terpenoid compounds? *New Phytol.* **2018**, *220*, 692–702. [[CrossRef](#)] [[PubMed](#)]
17. Lešnik, S.; Furlan, V.; Bren, U. Rosemary (*Rosmarinus officinalis* L.): Extraction techniques, analytical methods and health-promoting biological effects. *Phytochem. Rev.* **2021**, *20*, 1273–1328. [[CrossRef](#)]

18. Furlan, V.; Bren, U. Helichrysum italicum: From Extraction, Distillation, and Encapsulation Techniques to Beneficial Health Effects. *Foods* **2023**, *12*, 802. [[CrossRef](#)]
19. Fadilah, N.Q.; Jittmittraphap, A.; Leaugwutiwong, P.; Pripdeevech, P.; Dhanushka, D.; Mahidol, C.; Ruchirawat, S.; Kittakoop, P. Virucidal Activity of Essential Oils From Citrus x aurantium L. Against Influenza A Virus H1N1: Limonene as a Potential Household Disinfectant Against Virus. *Nat. Prod. Commun.* **2022**, *17*, 1934578X211072713. [[CrossRef](#)]
20. Chassagne, F.; Samarakoon, T.; Porras, G.; Lyles, J.T.; Dettweiler, M.; Marquez, L.; Salam, A.M.; Shabih, S.; Farrokhi, D.R.; Quave, C.L. A Systematic Review of Plants With Antibacterial Activities: A Taxonomic and Phylogenetic Perspective. *Front. Pharmacol.* **2021**, *11*, 2069. [[CrossRef](#)]
21. Islam, A.K.M.M.; Suttiyut, T.; Anwar, M.P.; Juraimi, A.S.; Kato-Noguchi, H. Allelopathic Properties of Lamiaceae Species: Prospects and Challenges to Use in Agriculture. *Plants* **2022**, *11*, 1478. [[CrossRef](#)]
22. Byers, K.J.R.P.; Bradshaw, H.D.; Riffell, J.A. Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (*Mimulus*). *J. Exp. Biol.* **2014**, *217*, 614–623. [[CrossRef](#)] [[PubMed](#)]
23. Nagel, R.; Schmidt, A.; Peters, R.J. Isoprenyl diphosphate synthases: The chain length determining step in terpene biosynthesis. *Planta* **2018**, *249*, 9–20. [[CrossRef](#)]
24. Dickschat, J.S. Bacterial Diterpene Biosynthesis. *Angew. Chem. Int. Ed.* **2019**, *58*, 15964–15976. [[CrossRef](#)]
25. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4126–4133. [[CrossRef](#)] [[PubMed](#)]
26. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **2011**, *66*, 212–229. [[CrossRef](#)]
27. Karunanithi, P.S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.* **2019**, *10*, 1166. [[CrossRef](#)]
28. Liu, X.; Zhu, X.; Wang, H.; Liu, T.; Cheng, J.; Jiang, H. Discovery and modification of cytochrome P450 for plant natural products biosynthesis. *Synth. Syst. Biotechnol.* **2020**, *5*, 187. [[CrossRef](#)]
29. Foti, R.S.; Honaker, M.; Nath, A.; Pearson, J.T.; Buttrick, B.; Isoherranen, N.; Atkins, W.M. Catalytic vs. Inhibitory Promiscuity in Cytochrome P450s: Implications for Evolution of New Function. *Biochemistry* **2011**, *50*, 2387. [[CrossRef](#)]
30. Fischer, M.; Knoll, M.; Sirim, D.; Wagner, F.; Funke, S.; Pleiss, J.; Bateman, A. The Cytochrome P450 Engineering Database: A navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* **2007**, *23*, 2015–2017. [[CrossRef](#)] [[PubMed](#)]
31. Nelson, D.R. The Cytochrome P450 Homepage. *Hum. Genom.* **2009**, *4*, 59. [[CrossRef](#)] [[PubMed](#)]
32. Nelson, D.R. Cytochrome P450 nomenclature, 2004. *Methods Mol. Biol.* **2006**, *320*, 1–10. [[CrossRef](#)] [[PubMed](#)]
33. Rasool, S.; Mohamed, R. Plant cytochrome P450s: Nomenclature and involvement in natural product biosynthesis. *Protoplasma* **2015**, *253*, 1197–1209. [[CrossRef](#)] [[PubMed](#)]
34. Krause, S.T.; Liao, P.; Crocoll, C.; Boachon, B.; Förster, C.; Leidecker, F.; Wiese, N.; Zhao, D.; Wood, J.C.; Buell, C.R.; et al. The biosynthesis of thymol, carvacrol, and thymohydroquinone in Lamiaceae proceeds via cytochrome P450s and a short-chain dehydrogenase. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2110092118. [[CrossRef](#)] [[PubMed](#)]
35. Gupta, P.; Geniza, M.; Naithani, S.; Phillips, J.L.; Haq, E.; Jaiswal, P. Chia (*Salvia hispanica*) Gene Expression Atlas Elucidates Dynamic Spatio-Temporal Changes Associated With Plant Growth and Development. *Front. Plant Sci.* **2021**, *12*, 667678. [[CrossRef](#)]
36. Li, H.; Li, J.; Dong, Y.; Hao, H.; Ling, Z.; Bai, H.; Wang, H.; Cui, H.; Shi, L. Time-series transcriptome provides insights into the gene regulation network involved in the volatile terpenoid metabolism during the flower development of lavender. *BMC Plant Biol.* **2019**, *19*, 313. [[CrossRef](#)]
37. Lichman, B.R.; Godden, G.T.; Buell, C.R. Gene and genome duplications in the evolution of chemodiversity: Perspectives from studies of Lamiaceae. *Curr. Opin. Plant Biol.* **2020**, *55*, 74–83. [[CrossRef](#)]
38. Bak, S.; Beisson, F.; Bishop, G.; Hamberger, B.; Höfer, R.; Paquette, S.; Werck-Reichhart, D. Cytochromes P450. *Arab. Book* **2011**, *9*, e0144. [[CrossRef](#)]
39. Xie, Y.; Ye, S.; Wang, Y.; Xu, L.; Zhu, X.; Yang, J.; Feng, H.; Yu, R.; Karanja, B.; Gong, Y.; et al. Transcriptome-based gene profiling provides novel insights into the characteristics of radish root response to Cr stress with next-generation sequencing. *Front. Plant Sci.* **2015**, *6*, 202. [[CrossRef](#)]
40. Manzano, A.; Carnero-Diaz, E.; Herranz, R.; Medina, F.J. Recent transcriptomic studies to elucidate the plant adaptive response to spaceflight and to simulated space environments. *iScience* **2022**, *25*, 104687. [[CrossRef](#)]
41. Howlader, J.; Robin, A.H.K.; Natarajan, S.; Biswas, M.K.; Sumi, K.R.; Song, C.Y.; Park, J.-I.; Nou, I.-S. Transcriptome Analysis by RNA-Seq Reveals Genes Related to Plant Height in Two Sets of Parent-hybrid Combinations in Easter lily (*Lilium longiflorum*). *Sci. Rep.* **2020**, *10*, 9082. [[CrossRef](#)] [[PubMed](#)]
42. Kang, D.D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, *7*, e7359. [[CrossRef](#)]
43. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [[CrossRef](#)] [[PubMed](#)]
44. Chaudhary, S.; Khokhar, W.; Jabre, I.; Reddy, A.S.N.; Byrne, L.J.; Wilson, C.M.; Syed, N.H. Alternative splicing and protein diversity: Plants versus animals. *Front. Plant Sci.* **2019**, *10*, 708. [[CrossRef](#)]
45. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2015**, *44*, D279–D285. [[CrossRef](#)] [[PubMed](#)]

46. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309. [[CrossRef](#)]
47. Curie, C.; Cassin, G.; Couch, D.; Divol, F.; Higuchi, K.; Le Jean, M.; Misson, J.; Schikora, A.; Czernic, P.; Mari, S. Metal movement within the plant: Contribution of nicotianamine and yellow stripe 1-like transporters. *Ann. Bot.* **2009**, *103*, 1–11. [[CrossRef](#)]
48. Ishimaru, Y.; Masuda, H.; Bashir, K.; Inoue, H.; Tsukamoto, T.; Takahashi, M.; Nakanishi, H.; Aoki, N.; Hirose, T.; Ohsugi, R.; et al. Rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. *Plant J.* **2010**, *62*, 379–390. [[CrossRef](#)]
49. Li, Z.; Li, W.; Guo, M.; Liu, S.; Liu, L.; Yu, Y.; Mo, B.; Chen, X.; Gao, L. Origin, evolution and diversification of plant ARGONAUTE proteins. *Plant J.* **2022**, *109*, 1086–1097. [[CrossRef](#)]
50. Zhang, Z.; Liu, X.; Li, R.; Yuan, L.; Dai, Y.; Wang, X. Identification and functional analysis of a protein disulfide isomerase (AtPDI1) in *Arabidopsis thaliana*. *Front. Plant Sci.* **2018**, *9*, 913. [[CrossRef](#)]
51. Finkelstein, R. Abscisic Acid Synthesis and Response. *Arab. Book* **2013**, *11*, e0166. [[CrossRef](#)] [[PubMed](#)]
52. Liao, W.; Zhao, S.; Zhang, M.; Dong, K.; Chen, Y.; Fu, C.; Yu, L. Transcriptome assembly and systematic identification of novel cytochrome P450s in *taxus chinensis*. *Front. Plant Sci.* **2017**, *8*, 1468. [[CrossRef](#)]
53. Degtyarenko, K.N. Structural domains of P450-containing monooxygenase systems. *Protein Eng. Des. Sel.* **1995**, *8*, 737–747. [[CrossRef](#)] [[PubMed](#)]
54. Vasav, A.P.; Barvkar, V.T. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters. *BMC Genom.* **2019**, *20*, 116. [[CrossRef](#)] [[PubMed](#)]
55. Wegrzyn, G.; Schachner, M.; Gabbiani, G.; Minerdi, D.; Savoie, S.; Sabbatini, P. Role of Cytochrome P450 Enzyme in Plant Microorganisms & Communication: A Focus on Grapevine. *Int. J. Mol. Sci.* **2023**, *24*, 4695. [[CrossRef](#)]
56. Bathe, U.; Tissier, A. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* **2019**, *161*, 149–162. [[CrossRef](#)]
57. Hansen, C.C.; Nelson, D.R.; Møller, B.L.; Werck-Reichhart, D. Plant cytochrome P450 plasticity and evolution. *Mol. Plant* **2021**, *14*, 1244–1265. [[CrossRef](#)]
58. Haudenschild, C.; Schalk, M.; Karp, F.; Croteau, R. Functional Expression of Regiospecific Cytochrome P450 Limonene Hydroxylases from Mint (*Mentha* spp.) in *Escherichia coli* and *Saccharomyces cerevisiae*. *Arch. Biochem. Biophys.* **2000**, *379*, 127–136. [[CrossRef](#)]
59. Lupien, S.; Karp, F.; Wildung, M.; Croteau, R. Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. *Arch. Biochem. Biophys.* **1999**, *368*, 181–192. [[CrossRef](#)]
60. Chen, X.; Zhang, C.; Too, H.P. Multienzyme Biosynthesis of Dihydroartemisinic Acid. *Molecules* **2017**, *22*, 1422. [[CrossRef](#)]
61. Wu, Y.; Hillwig, M.L.; Wang, Q.; Peters, R.J. Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Lett.* **2011**, *585*, 3446. [[CrossRef](#)]
62. Sawai, S.; Saito, K. Triterpenoid Biosynthesis and Engineering in Plants. *Front. Plant Sci.* **2011**, *585*, 3446–3451. [[CrossRef](#)] [[PubMed](#)]
63. Chen, Z.; Qi, X.; Yu, X.; Zheng, Y.; Liu, Z.; Fang, H.; Li, L.; Bai, Y.; Liang, C.; Li, W. Genome-Wide Analysis of Terpene Synthase Gene Family in *Mentha longifolia* and Catalytic Activity Analysis of a Single Terpene Synthase. *Genes* **2021**, *12*, 518. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, W.; Liu, Y.; Yan, J.; Cao, S.; Bai, F.; Yang, Y.; Huang, S.; Yao, L.; Anzai, Y.; Kato, F.; et al. New reactions and products resulting from alternative interactions between the P450 enzyme and redox partners. *J. Am. Chem. Soc.* **2014**, *136*, 3640–3646. [[CrossRef](#)] [[PubMed](#)]
65. Hernandez-Ortega, A.; Vinaixa, M.; Zebec, Z.; Takano, E.; Scrutton, N.S. A Toolbox for Diverse Oxyfunctionalisation of Monoterpenes OPEN. *Sci. Rep.* **2018**, *8*, 14396. [[CrossRef](#)]
66. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
67. Hoff, K.J.; Lomsadze, A.; Borodovsky, M.; Stanke, M. Whole-Genome Annotation with BRAKER. *Methods Mol. Biol.* **2019**, *1962*, 65. [[CrossRef](#)]
68. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **2016**, *32*, 767–769. [[CrossRef](#)]
69. Brůna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **2021**, *3*, lqaa108. [[CrossRef](#)]
70. Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform.* **2006**, *7*, 62. [[CrossRef](#)] [[PubMed](#)]
71. Girgis, H.Z. Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinform.* **2015**, *16*, 227. [[CrossRef](#)]
72. Grüning, B.; Yusuf, D.; Houwaart, T.; Anika; Miladi, M.; Gu, Q.; Batut, B.; Soranzo, N.; Gamaleldin, H.; Von Kuster, G.; et al. *Bgruening/Galaxytools: September Release 2019*; Zenodo: Geneva, Switzerland, 2018. [[CrossRef](#)]

73. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 24 March 2023).
74. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10. [[CrossRef](#)]
75. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
76. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
77. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; Von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)] [[PubMed](#)]
78. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)] [[PubMed](#)]
79. Afgan, E.; Baker, D.; Batut, B.; Van Den Beek, M.; Bouvier, D.; Ech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B.A.; et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [[CrossRef](#)]
80. Batut, B.; Freeberg, M.; Heydarian, M.; Erxleben, A.; Videm, P.; Blank, C.; Doyle, M.; Soranzo, N.; van Heusden, P.; Delisle, L. Reference-Based RNA-Seq Data Analysis (Galaxy Training Materials). Available online: <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/ref-based/tutorial.html#citing-this-tutorial> (accessed on 25 March 2023).
81. Batut, B.; Hiltmann, S.; Bagnacani, A.; Baker, D.; Bhardwaj, V.; Blank, C.; Bretaudeau, A.; Brillet-Guéguen, L.; Čech, M.; Chilton, J.; et al. Community-Driven Data Analysis Training for Biology. *Cell Syst.* **2018**, *6*, 752–758.e1. [[CrossRef](#)] [[PubMed](#)]
82. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
83. Kweon, O.; Kim, S.J.; Kim, J.H.; Nho, S.W.; Bae, D.; Chon, J.; Hart, M.; Baek, D.H.; Kim, Y.C.; Wang, W.; et al. CYPminer: An automated cytochrome P450 identification, classification, and data analysis tool for genome data sets across kingdoms. *BMC Bioinform.* **2020**, *21*, 160. [[CrossRef](#)]
84. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [[CrossRef](#)] [[PubMed](#)]
85. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
86. Zhu, A.; Ibrahim, J.G.; Love, M.I. Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* **2019**, *35*, 2084–2092. [[CrossRef](#)] [[PubMed](#)]
87. Bioinformatics Training at the Harvard Chan Bioinformatics Core. Available online: <https://hbctraining.github.io/main/> (accessed on 19 April 2023).
88. Pantiora, P.; Furlan, V.; Matiadis, D.; Mavroidi, B.; Perperopoulou, F.; Papageorgiou, A.C.; Sagnou, M.; Bren, U.; Pelecanou, M.; Labrou, N.E. Monocarbonyl Curcumin Analogues as Potent Inhibitors against Human Glutathione Transferase P1-1. *Antioxidants* **2023**, *12*, 63. [[CrossRef](#)] [[PubMed](#)]
89. Kores, K.; Kolenc, Z.; Furlan, V.; Bren, U. Inverse Molecular Docking Elucidating the Anticarcinogenic Potential of the Hop Natural Product Xanthohumol and Its Metabolites. *Foods* **2022**, *11*, 1253. [[CrossRef](#)]
90. Wen, W.; Yu, R. Artemisinin biosynthesis and its regulatory enzymes: Progress and perspective. *Pharmacogn. Rev.* **2011**, *5*, 189. [[CrossRef](#)]
91. Sun, W.; Xu, Z.; Song, C.; Chen, S. Herbgenomics: Decipher molecular genetics of medicinal plants. *Innovation* **2022**, *3*, 100322. [[CrossRef](#)]
92. Alami, M.M.; Ouyang, Z.; Zhang, Y.; Shu, S.; Yang, G.; Mei, Z.; Wang, X. The Current Developments in Medicinal Plant Genomics Enabled the Diversification of Secondary Metabolites' Biosynthesis. *Int. J. Mol. Sci.* **2022**, *23*, 15932. [[CrossRef](#)]
93. Cheng, Q.Q.; Ouyang, Y.; Tang, Z.Y.; Lao, C.C.; Zhang, Y.Y.; Cheng, C.S.; Zhou, H. Review on the Development and Applications of Medicinal Plant Genomes. *Front. Plant Sci.* **2021**, *12*, 2981. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.