

Article

Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids

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

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

† These authors contributed equally to this work.

Abstract: *Caryopteris x Clandonensis*, also known as bluebeard, is an ornamental plant containing a large variety of terpenes and terpene-like compounds. Four different cultivars were subjected to a principal component analysis to elucidate variations in terpenoid-biosynthesis and consequently, two representative cultivars were sequenced on a genomic level. Functional annotation of genes as well as comparative genome analysis on long read datasets enabled the identification of cultivar-specific terpene synthase and cytochrome p450 enzyme sequences. This enables new insights, especially since terpenoids in research and industry are gaining increasing interest due to their importance in areas such as food preservation, fragrances, or as active ingredients in pharmaceutical formulations. According to BUSCO assessments, the presented genomes have an average size of 355 Mb and about 96.8% completeness. An average of 52,090 genes could be annotated as putative proteins, whereas about 42 were associated with terpene synthases and about 1340 with cytochrome p450 enzymes.

Keywords: reference genome; terpene synthases; *Caryopteris x clandonensis*; plant volatiles; long read sequencing; TPS subfamilies



Citation: Ritz, M.; Ahmad, N.; Brueck, T.; Mehlmer, N. Comparative Genome-Wide Analysis of Two *Caryopteris x Clandonensis* Cultivars: Insights on the Biosynthesis of Volatile Terpenoids. *Plants* **2023**, *12*, 632. <https://doi.org/10.3390/plants12030632>

Academic Editors: Noe Fernandez-Pozo and M. Gonzalo Claros

Received: 21 December 2022

Revised: 25 January 2023

Accepted: 27 January 2023

Published: 1 February 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

Throughout the last decades, terpenes and terpenoids became more and more important in industrial applications. In the food industry terpenes are used, e.g., as flavoring compounds [1] or preservatives [2]. Due to its plant origin, the acceptance as a food additive is higher compared to chemical synthesis. In a pharmaceutical context the research and use of essential oils—with terpenes as their main components—range from anti-inflammatory [3], and immunomodulatory [4] to antiviral [5] and further indications [6–11]. The anti-cancer drug Taxol consists of a diterpenoid backbone [12] and is employed in different cancer treatments [13]. The success of this terpenoid surely is one of the reasons to further research terpenoids for pharmaceutical applications. Along with these applications, this class of molecules can be found throughout most organisms. Flowering plants show a vast diversity of terpenoids, which is a unique characteristic of the class *Angiospermae* [14]. In plants, they are used as a defense mechanism against biotic (e.g., herbivores or pests) and abiotic influences (e.g., radiation or climate stress) [15]. An example of a defense mechanism against biotic stress is the insect repellent activity of volatiles, such as p-menthane-3,8-diol from *Corymbia citriodora* [16,17]. This compound shows activity against the yellow fever mosquito *Aedes aegypti* [18]. *Caryopteris x clandonensis* essential oils also harbor a biological activity against these insects [19]. However, for these plants, the active agent is not yet identified. Additionally, terpenoids function as attractors for pollinators or as a possibility for energy storage [14].

The extensive diversity of natural terpenes derives from the conserved evolution of terpene synthases (TPS) and terpene-modifying enzymes, such as cytochrome p450

enzymes [20,21]. Terpenes are divided into different classes defined by their backbone. The basis is two building blocks, isopentenylpyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP) which are synthesized in plants via the mevalonate pathway. IPP is the activated form of an isoprene unit consisting of five C-atoms (C₅), also called hemiterpene. These are connected to larger units forming monoterpenes (C₁₀), sesquiterpenes (C₁₅), diterpenes (C₂₀) and higher terpene structures [22]. Further steps of increasing terpenoid diversity involve the promiscuity of TPS as well as the subsequent modification by cytochrome p450 enzymes, which may encompass hydroxylation, carboxylation, acetylation or peroxide linkage. Examples include the biosynthesis of p-menthane-3,8-diol [17], gibberellin [23], taxol [24], and artemisinin [25], respectively. This results in a vast pool of natural compounds which account for a multitude of possible applications [14,26].

In general, plant TPS are divided into eight subfamilies which are grouped into classes I, II and III. This separation is based on functional assessment, sequence likelihood and architecture of genes. Class I is comprised of copalyl diphosphate synthases (TPS-c), *ent*-kaurene synthases (TPS-e), other diterpene synthases (TPS-f) and lycopod specific (TPS-h). TPS-d is only included in class II, which is specific for *Gymnosperms*. Lastly, class III consists of TPS-a, cyclic monoterpene synthases and hemi-TPS (TPS-b) and acyclic mono-TPS (TPS-g), which are *Angiosperm* specific [27].

With the advent of state-of-the-art bioinformatic technologies, deciphering the molecular mechanisms involved in the formation of terpenoids has become significantly easier [28]. Furthermore, the possibility to produce terpenes recombinantly by means of biotechnological production systems, rather than chemical synthesis, makes it an ecological and cost-effective technology for the increasing demand for terpenes in industrial applications, despite open challenges [29].

The combination of cutting-edge bioinformatics and next-generation sequencing technologies provided by Pacific Biosciences, Oxford Nanopore and Illumina allows for the rapid generation of draft genomes as well as the annotation of valid gene models. In this context, long-read sequencing technologies will be highlighted, as they exhibit no amplification biases. Consequently, these technologies provide a reliable basis for de novo whole-genome assemblies. Openly accessible bioinformatic tools enable cost-efficient assemblies, annotations and secondary downstream analyses for a broad range of scientists, and are publicly available via www.github.com (accessed on 11 December 2022) [30]. Two of these are the Quality Assessment Tool for Genome Assemblies (QUAST) [31] and Benchmarking Universal Single-Copy Orthologs (BUSCO). The latter is employed to assess the completeness of the obtained genome assemblies. Here, conserved and species-specific gene sequences are curated in databases and detected via a match-making algorithm to check for the gene set completeness of the evaluated taxonomic group [32]. An investigated genome is classified as complete if respective single-copy orthologs are present in the assembly.

In this work, we present the genomes of two *Caryopteris x clandonensis* cultivars (Dark Knight and Pink Perfection) from the *Lamiaceae* family in high quality employing long-read sequencing. These plants display a wide range of different metabolic pathways in regard to terpenoid biosynthesis, as also seen in other plants of the order *Lamiales*, e.g., in *Jasminum sambac* [33]. To elucidate variations between these multivariate datasets a principal component analysis (PCA) was conducted. Based on evident differences in volatile compound composition the two cultivars, Dark Knight and Pink Perfection were compared on a genomic level. This submission will be the 12th whole genome sequence within *Lamiaceae*, consisting of about 4788 further species, making it a source for gene sequences and further experimental basis in plant and natural product focused biosynthesis research.

2. Results and Discussion

2.1. PCA Analysis of Volatile Compounds

Differences between the volatile compounds of four cultivars were investigated using a GC-MS Headspace analysis. Ten main volatile components visible between the cultivars

were selected, predominantly monoterpenoids and sesquiterpenoids, which are listed in Table 1. It has already been shown that there is a variety of monoterpene synthases that are able to catalyze ionization and isomerization starting from geranyl diphosphate [34]. Furthermore, the analysis of the cultivars revealed that a switch between pinene and limonene-derived compounds took place, which was sparsely synthesized in the other plants. In Table 1, these compounds are marked with an asterisk, one (*) represents limonene-related terpenoids, and two (**) represents pinene-related terpenoids. This especially is visible in the C4-C6 shift compared to the limonene backbone as seen in pinene (C4 to C6, see Figure S1). Similar substances could be identified as investigated previously for this plant species [19].

Table 1. Ten main volatile compounds of four *Caryopteris x clandonensis* cultivars, visible between the cultivars were selected and are hierarchically listed (top: higher concentration, bottom: lower concentration). GC-MS Headspace was performed and an identification with a NIST/EPA/NIH MS library version 2.0 was conducted. * represents limonene-related terpenoids. ** represents pinene-related terpenoids.

Dark Knight	Good as Gold	Hint of Gold	Pink Perfection
α -pinene **	D-limonene *	D-limonene *	D-limonene *
trans-pinocarveol **	Cubebol	Cubebol	cis-p-mentha-1(7),8-dien-2-ol *
Pinocarvone **	Carvone *	trans-carveol *	trans-p-mentha-2,8-dien-1-ol *
Caryophyllene oxide	trans-carveol *	Carvone *	Caryophyllene oxide
β -pinene **	cis-p-mentha-1(7),8-dien-2-ol *	Caryophyllene oxide	trans-carveol *
(E,E)- α -farnesene	Caryophyllene oxide	trans-p-mentha-1(7),8-dien-2-ol *	cis-p-mentha-2,8-dien-1-ol *
α -campholenal	α -copaene	cis-p-mentha-1(7),8-dien-2-ol *	Carvone
α -copaene	β -pinene **	cis-p-mentha-2,8-dien-1-ol *	α -pinene **
Caryophyllene	cis-p-mentha-2,8-dien-1-ol *	α -copaene	β -pinene **
D-limonene *	trans-p-mentha-2,8-dien-1-ol *	trans-p-mentha-2,8-dien-1-ol *	Caryophyllene

As the plants are cultivars from *Caryopteris x clandonensis* a common base profile (e.g., caryophyllene, perillyl alcohol, sabinene, farnesene or campholenal) of volatiles was expected, see Table S1, as has been shown for other plants and their cultivars [35,36]. In this study, distinct differences between Dark Knight, Good as Gold, Hint of Gold, as well as Pink Perfection, can be shown.

To further investigate the variations in the compound profile found during the analysis, a principal component analysis (PCA) was performed (Figure 1). Good as Gold and Hint of Gold express high morphological and metabolic similarity (see Figure S2 and Table S1). This is also evident in Figure 1, as both cultivars are located close to each other. On the other hand, Dark Knight and Pink Perfection showed the highest deviation in volatile compound composition. Moreover, the switch between C1 and C6 as mentioned above results in an intriguing product spectrum. These data underline the variations between the cultivars and demonstrate a need for further investigations into the molecular makeup of underlying TPS and cytochrome p450 enzymes, which are key for generating the molecular diversity of plant-based terpenoid structures in plants [20]. Therefore, due to their distinct differences revealed in the PCA, the two cultivars, Dark Knight and Pink Perfection, were sequenced to elucidate genomic differences and identify unique and yet unknown genes.

2.2. Genome Sequencing and Quality Assessment

In Table 2, the sequencing metrics of the respective Sequel IIe runs are depicted. Details regarding sequencing quality reports can be found in Figure S3. Total bases were nearly twofold higher in Dark Knight than in Pink Perfection, the same as obtained HiFi reads and yield. However, the HiFi read length, read quality and number of passes are comparable in both sequencing runs. Deviations in sequencing parameters are closely related to utilized libraries and input DNA quality. As the read quality is well above Q20 both runs were subjected to further analyses.

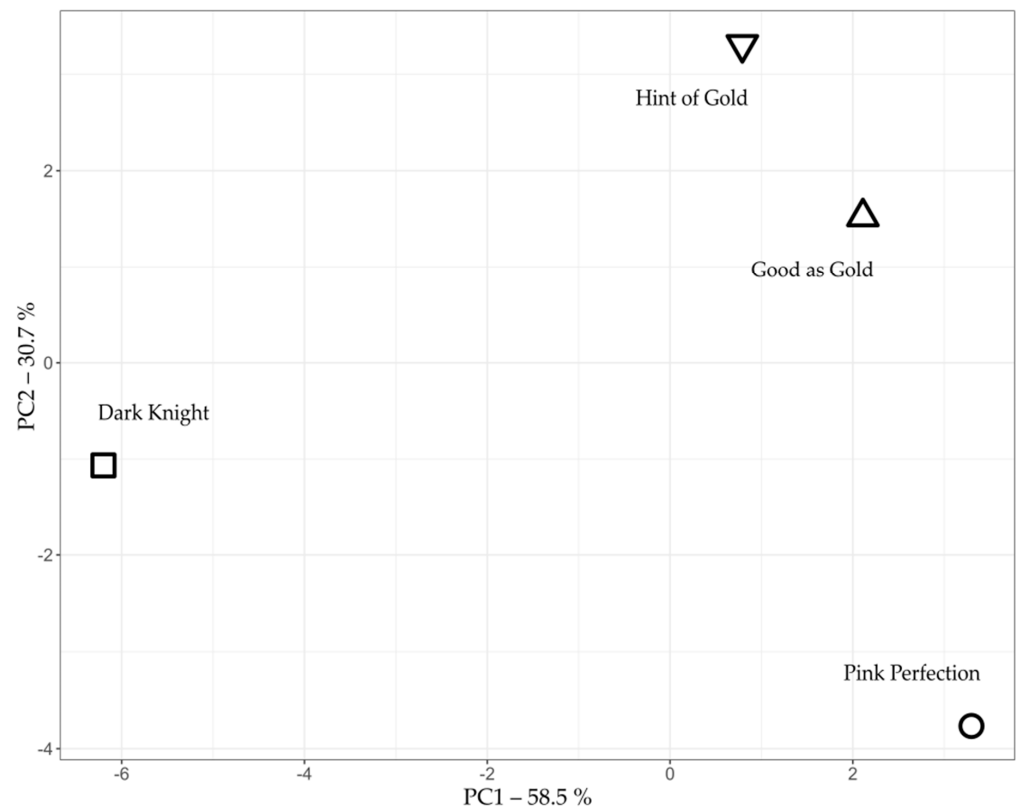


Figure 1. A principal component analysis of four different *Caryopteris x clandonensis* cultivars, Dark Knight, Good as Gold, Hint of Gold and Pink Perfection regarding the area of their volatile compounds analyzed by GC-MS Headspace.

Table 2. Sequencing parameters of the PacBio Sequel IIe runs of *Caryopteris x clandonensis* cultivars Dark Knight and Pink Perfection.

Analysis Metric	Dark Knight	Pink Perfection
Total Bases (Gb)	444.13	229.43
HiFi Reads	1,823,939	843,632
HiFi Yield (Gb)	27.28	12.92
HiFi Read Length (mean, bp)	14,954	15,312
HiFi Read Quality (median)	Q35	Q34
HiFi Number of Passes (mean)	12	13

In this study, both genomes of Dark Knight and Pink Perfection were assembled using the IPA assembler with a consecutive duplicate purging and phasing step. A QAST analysis was conducted to assess assembly contiguity (see Table 3).

Table 3. Genome contiguity assessment based on statistics generated by using QAST.

Assembly	Dark Knight	Pink Perfection
# contigs	1183	782
Largest contig	29,672,976	31,977,049
Total length	366,625,098	344,117,456
Estimated reference length	300,000,000	300,000,000
GC (%)	31.50	31.77
N50	8,177,750	7,086,741
L50	13	14
# N's per 100 kbp	0.41	0.44

The number of assembled contigs diverged in both candidates (see Table 3). However, respective L50 values were small (13 for Dark Knight and 14 for Pink Perfection) compared to obtained N50 (8.2 Mb and 7.1 Mb respectively), which assures gene integrity with only low or no fragmentation. The total contig length of complete genomes corresponds to their size, which is comparable (3.44 to 3.66×10^8 bp), and the same as seen for GC content (31.5% and 31.77%). Furthermore, genome size was calculated using a k-mer-based analysis, with a k-mer size of 20. Results support the haploid genome size of ~355 Mb and estimated a diploid genome, see Figures S4 and S5. Based on the calculated genome size the coverage of Dark Knight and Pink Perfection resembles 74 and 38, respectively.

To assess the genome completeness and reliability of both genome sequences, a Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis was performed (see Figure 2). Both genomes were compared to the kingdom *Viridiplantae* and the clades *Embryophyta* and *Eudicotidae*, respectively. The selection of these lineages was based on the increasing grade of affiliation and the different accompanying BUSCO gene sets (in former order). For closer clades, more concise sequences are necessary in order to be identified as complete. In our case, even more affiliated clades show less deviation of completeness than expected in comparison to *Viridiplantae*. As the genomes were compared to different BUSCO datasets, the obtained results were depicted after normalization in Figure 2 to enable a concise comparison. Assessed genome completeness from the closest related clade (*Eudicotidae*) was 96.6% for Dark Knight and 96.8% for Pink Perfection, which were also compared to reference genomes of *Salvia splendens* (92.1%) [37] and *Sesamum indicum* (95.1%) [38]. The latter were only compared with the *Viridiplantae* database with BUSCO v2.0.1 and v3.0, whereas our data were analyzed by BUSCO v5.3.2. This may have caused the difference between 425 and 1440 BUSCO datasets, as frequent updates of the gene sets are necessary to improve BUSCO analysis [39]. The reference genomes were chosen due to the high prevalence in BLAST searches [40,41] using *Caryopteris x clandonensis* sequences. *S. splendens* appears to harbor mostly complete and duplicated BUSCOs, whereas *S. indicum* shows comparable results to the new genomes of Pink Perfection and Dark Knight with a majority of complete and single-copy BUSCOs. To interpret BUSCO results, it is necessary to understand duplicated BUSCOs and their nature, as these can be of biological or technical origin. In eukaryotic genomes, divergences in haplotypes often lead assemblers to form duplicates of high heterozygosity regions, resulting in contiguity issues and obstacles in further evaluation steps, such as gene annotation [42,43]. To circumvent these issues, tools such as “purge_dups” are utilized to remove duplicate regions (haplotigs) from the assembly to assure genome contiguity [42]. A consecutive polishing of obtained contigs and haplotigs using phasing results in increased genome quality. Of the newly assembled genomes only 0.24–0.69%/0.71–2.67% are fragmented or missing, respectively. The absence of some BUSCO genes may be due to a loss of true genes or these may be existing as true gene duplications [43].

2.3. Evaluation of Structural Differences between Genome Assemblies

To concisely compare genomes, the collinear gene order also known as synteny or syntenic blocks needs to be assessed [44]. It plays an important role in visualizing matches between organisms [45].

Investigating the synteny between cultivar genomes shows their close relation. Here, factors such as low contiguity and fragmentation have an effect on the analysis and lead to high error rates [46]. In our case, previously performed evaluations assured high contiguity and low fragmentation. Mauve was used to perform a multiple sequence alignment and applied to generate synteny blocks (Figure S6) [47]. Connections between these blocks reveal the high similarity within both genomes. This is typical for plant breeding, as specific traits are inherited from previous generations leading to inversions, duplications, or truncations in gene sets [48]. Furthermore, marker synteny can be used for phylogenetic analyses of cultivar evolution [49]. Thus, the plant samples seem to be closely related to the species *Caryopteris x clandonensis*.

2.4. Gene Models and Functional Annotation

Gene models were computed using the presented genome assembly and a long-read IsoSeq database as hints via AUGUSTUS [50–53]. As a training set *Solanum lycopersicum* was chosen due to its ancestral relation to *Lamiaceae*. For the cultivars, a total of 52,865 (Dark Knight), and 51,315 (Pink Perfection) genes were predicted and resemble putative proteins. The Cluster of Orthologous Groups (COG) and Gene Ontology (GO) terms were evaluated for all cultivars. It is to mention that only ~81% of the predicted genes were annotated using COG and GO databases. Out of these ~30% are poorly characterized (Figure 3E) and only a fraction (30%) of those can be assigned with GO terms. In regard to the complete genomes, nearly 20% of the proposed gene models remain without an assigned function. Figure 3 shows the COG counts for the following categories: (3B) cellular processes and signaling (3C) information storage and processing (3D) metabolism and (3E) poorly characterized. Figure 3A combines all the aforementioned categories. The obtained results emphasize a strong similarity in the compared cultivars. Further data in regard to the exact amount of COG per category can be found in Tables S2 and S3. This finding is a further indicator of the completeness of the presented genomes, as different cultivars have a similar set of genes, only varying in small nucleotide polymorphisms or other structural variants, which distinguish them [54,55].

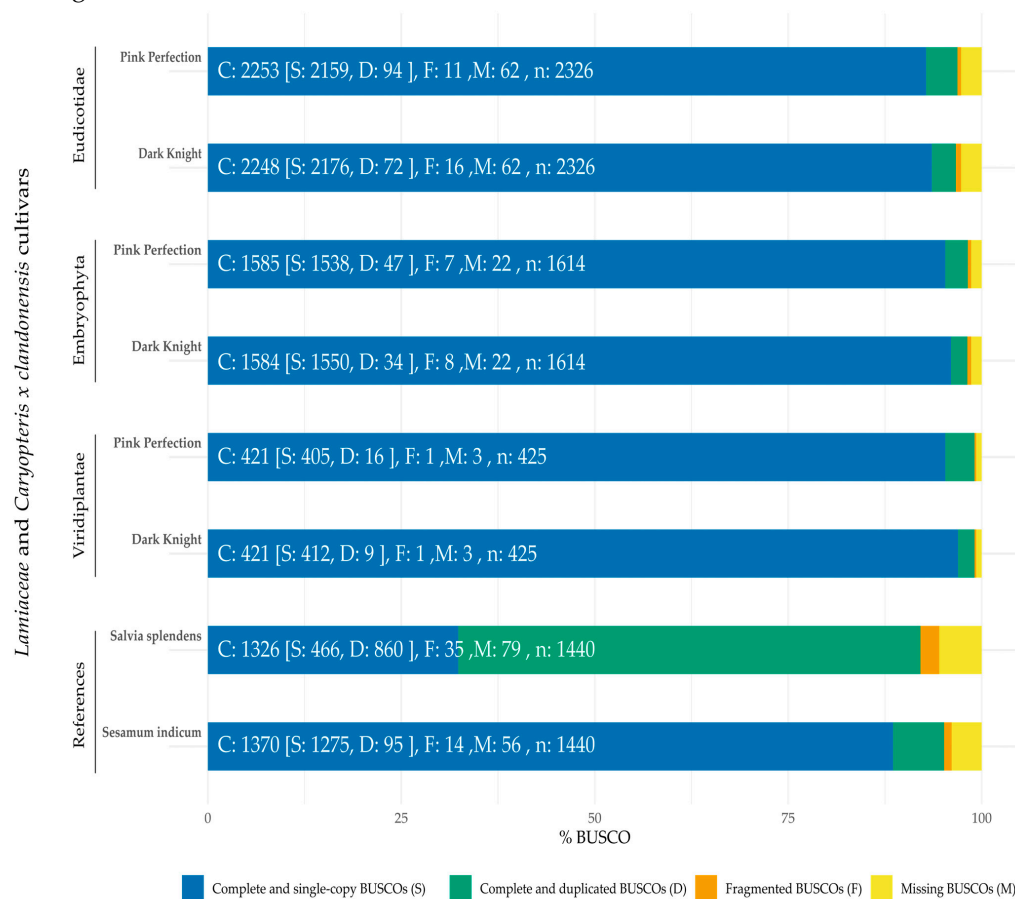


Figure 2. Comparison of BUSCO completeness of different cultivars of *Caryopteris x clandonensis* as well as *Salvia splendens* [37] and *Sesamum indicum* [38]. As the genomes were compared to other Benchmarking Universal Single-Copy Orthologs (BUSCO) datasets a normalization was performed to enable a comparison in genome completeness. Pink Perfection and Dark Knight were compared to the BUSCO datasets of *Viridiplantae*, *Embryophyta* and *Eudicotidae*, whereas *S. splendens* and *S. indicum* were compared to *Viridiplantae* only. Reference genomes were obtained from [37,38].

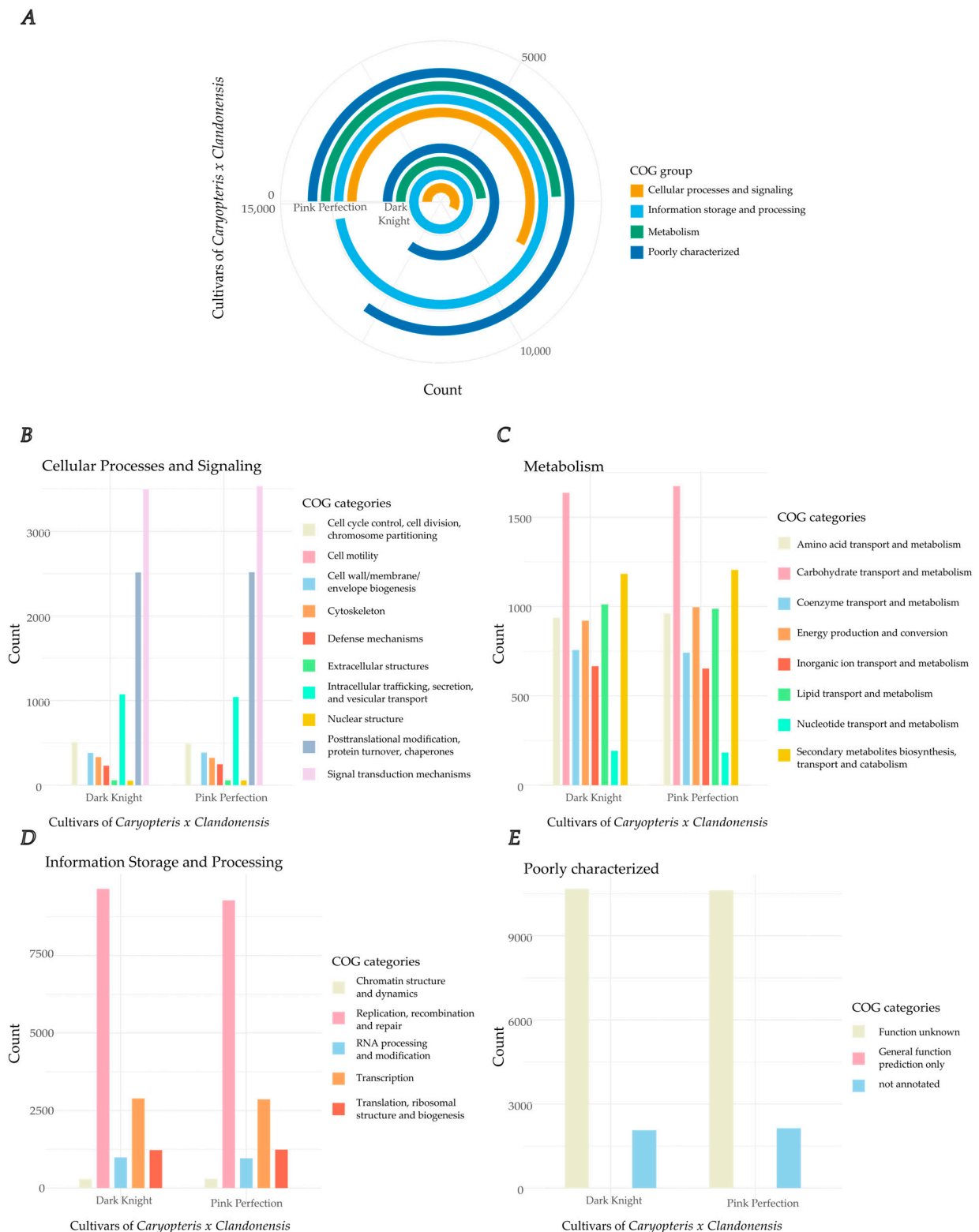


Figure 3. Annotation of gene sets for Cluster of Orthologous Groups (COG) for both cultivars, Dark Knight and Pink Perfection. (A) COG of two different cultivars of *Caryopteris x clandonensis*, Pink Perfection (outer ring) and Dark Knight (inner ring). Groups are divided in cellular processes and signaling, information storage and processing, metabolism, and a category for poorly characterized gene sets. (B) COG of cellular processes and signaling associated genes, total counts. (C) COG of metabolism-associated genes, total counts. (D) COG of information storage and processing associated genes, total counts. (E) COG of poorly characterized genes, total counts.

A closer look into the different groups reveals characteristic functions in the cultivars. Most genes identified and functionally annotated are associated with replication, recombination and repair, which make up about 20.5% of total annotated genes (Figure 3D) followed by signal transduction mechanism (~8%) (Figure 3A). Plants are exposed to endogenous and exogenous stresses such as chemicals or UV-radiation which can significantly alter DNA, thus there is high importance for repair mechanisms [56]. High redundancy of those ensures the safe replication of DNA with almost no errors [57].

In Figure 3C, proteins related to the COG category secondary metabolites biosynthesis, transport and catabolism, rank in second place within metabolism (2.8%). This category harbors TPS and cytochrome p450 enzymes. However, proteins associated with carbohydrate transport and metabolism are most abundant in this group as they are important for general metabolism and backbone synthesis.

Compared to about 29,458 with COG functionally annotated genes, 11,118 unique GO terms were assigned to 14,280 different genes (27% of total gene models). COG terms are ancestrally conserved regions, GO terminology in contrast proposes functional annotation of each hypothetical gene. A gene-set enrichment analysis was conducted with GO terms as a source for gene sets [58]. The following figures show the GO term clustering regarding the three main categories in plants: biological process (Figure 4), molecular function (Figure 5) and cellular components (Figure 6). For all three an analysis was conducted based on GO terms identified in Pink Perfection. Detailed data for Dark Knight and Pink Perfection can be found in Tables S4 and S5. The GO analysis was visualized using REVIGO [59]. Respective cluster position within the semantic space is irrelevant, as similar semantic terms are located in vicinity of each other in the plot [58].

In Figure 4, GO terms related to biological processes are depicted with their respective prevalence (dot size). In addition, some clusters with similar functions were grouped by circles into the main function of these GO terms, as can be seen, e.g., with “translation” in the bottom right corner. Incorporated into this cluster are the terms: protein modification process, DNA metabolic process, nucleobase-containing compound metabolic process, and protein metabolic process. The cluster organelle organization includes cytoskeleton organization, cytoplasm organization, and mitochondrion organization. Clustered with transport: ion transport, protein transport. The last cluster response to stress contains the GO terms response to a biotic stimulus, response to an abiotic stimulus, response to an external stimulus, and response to an endogenous stimulus. GO terms without clustering but still strongly prevalent in the PCA are biological, metabolic and biosynthetic processes.

For the GO analysis of the category molecular function, only one larger cluster was formed, which is nucleic acid binding. It incorporates the functions of DNA binding, RNA binding, and nucleotide binding. The two main components in this category are molecular function and catalytic activity.

GO analysis in the category of cellular components yielded as the main results, intracellular anatomical structure and cellular components, as well as genes related to the cytoplasm. However, no semantic clustering was feasible based on the annotated GO terms.

2.5. Identification of Terpenoid Biosynthesis Enzymes

InterProScan predicts distinct protein domains and classifies them into families [60,61]. The seed files PF01397, PF03936 and IPR036965 are associated with TPS activity. In the annotated protein database, these seeds were used as homology motifs. For Dark Knight 43 and Pink Perfection 41 TPS were identified. The seed file IPR001128 is related to cytochrome p450 enzymes. Here, we were able to identify Dark Knight and Pink Perfection 1316 and 1363 sequences. Compared to other plants these findings are comparable, both for TPS and cytochrome p450 enzymes [62–65].

To investigate the similarity and the affiliation into TPS subfamilies regarding identified TPS, a phylogenetic tree was constructed. Analysis was based on multiple sequence alignment by Clustal Omega using default parameters (see Figure 7). To differentiate between TPS families, 55 selected sequences of representative plant species were utilized

as anchor sequences along with putative TPS from Dark Knight and Pink Perfection; the root was *Physcomitrella patens* (adapted from [66]). The multicolored clades belong to the different TPS subfamilies and are used as references, for a more detailed overview see the supplemental *lamiaceae* reference. Concise numbers of TPS subfamily distribution in both cultivars are shown in Table 4. The most prominent subfamilies are TPS-a (green), TPS-b (black) and TPS-c (purple), which is in line with the distribution in *Eudicots*, *Angiosperms* and land plants [67]. The subfamilies TPS-d and TPS-h are not present in the investigated cultivars. These findings are supported by the literature, as TPS-d clusters are derived from *Gymnosperm* species [63,68] and TPS-h are specific for *Selaginella moellendorffii* [67].

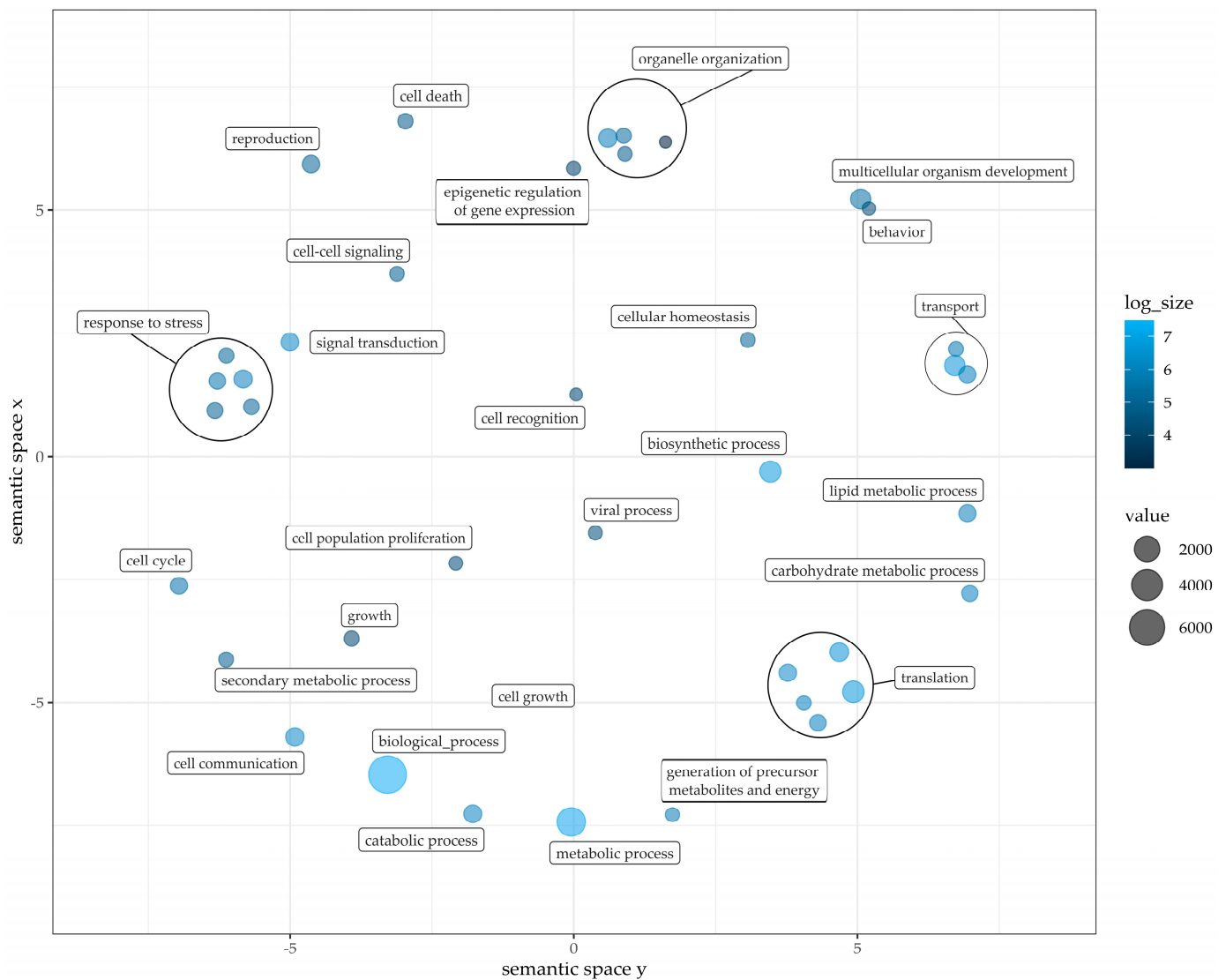


Figure 4. Gene Ontology term classification within biological processes of Pink Perfection. Clustered with response to stress: response to biotic stimulus, response to abiotic stimulus, response to external stimulus, response to endogenous stimulus. Clustered with translation: protein modification process, DNA metabolic process, nucleobase-containing compound metabolic process, protein metabolic process. Clustered with organelle organization: cytoskeleton organization, cytoplasm organization, mitochondrion organization. Clustered with transport: ion transport, protein transport. Figure was drafted employing REVIGO [59] and customized with R. Value and log size represents the counted GO terms across annotated gene models.

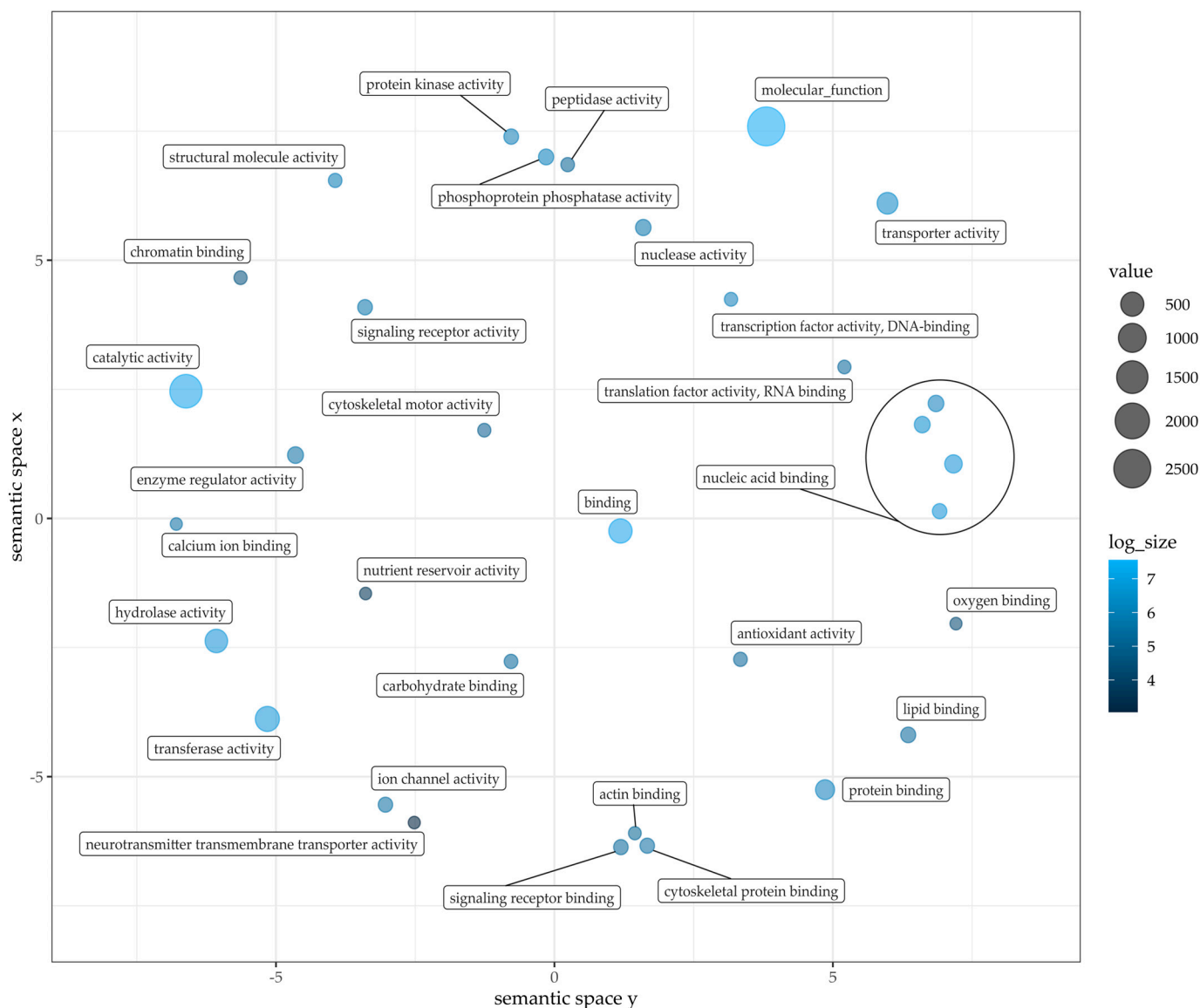


Figure 5. Gene Ontology term classification within molecular functions of Pink Perfection, clustered with nucleic acid binding: DNA binding, RNA binding, Nucleotide binding. Figure was drafted employing REVIGO [59] and customized with R. Value and log size represents the counted GO terms across annotated gene models.

Table 4. Terpene synthase (TPS) subfamilies and their distribution in the *Caryopteris x clandonensis* cultivars Dark Knight and Pink Perfection. TPS-a, -b and -c show the highest prevalence in both cultivars.

TPS Subfamily	Dark Knight	Pink Perfection
a (green)	16	14
b (black)	7	7
c (purple)	10	10
d (blue)	-	-
e (turquoise)	2	2
f (petrol)	5	5
g (red)	3	3
h (pink)	-	-

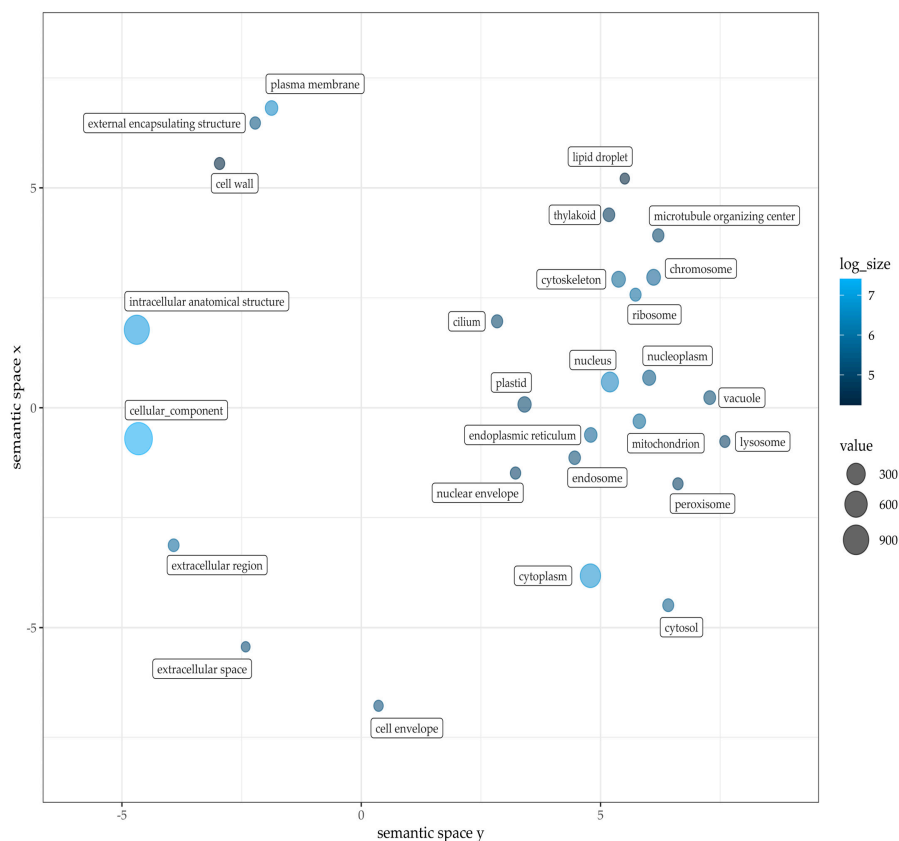


Figure 6. Gene Ontology term classification within cellular components of Pink Perfection. Figure was drafted employing REVIGO [59] and customized with R. Value and log size represent the counted GO terms across annotated gene models.

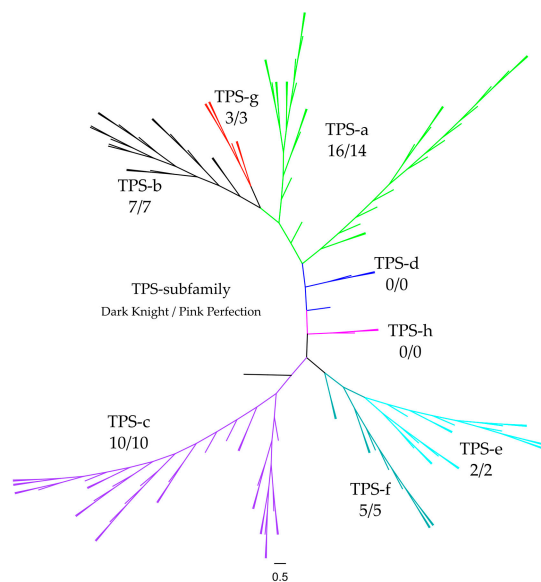


Figure 7. Phylogenetic tree of putative terpene synthases (TPS) within *Caryopteris x clandonensis* cultivars Dark Knight (DK) and Pink Perfection (PP). TPS-a (green), TPS-b (black), TPS-c (purple), TPS-d (blue), TPS-e (turquoise), TPS-f (petrol), TPS-g (red), TPS-h (pink). For phylogenetic tree construction, TPS a-h of selected plant species were included to assure correct classification of identified TPS. Numbers below the respective TPS subfamily indicate the count of predicted TPS in the genomes of the cultivars.

3. Materials and Methods

3.1. Plant Material

Four cultivars of *Caryopteris x clandonensis* were acquired from a local nursery (Foerster Pflanzen GmbH, Bietigheim-Bissingen, Germany) and grown to maturity in the open in a warm, moderate climate zone. After maturity, healthy leaves and blossoms were sampled and snap frozen in liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$ until preparation for transcriptome and genome sequencing. Fresh mature leaves were used for GC-MS headspace analysis of volatile compounds.

3.2. GC-MS Analysis of Volatile Compounds

Fresh mature leaves were weighed in GC headspace vials and analyzed using a Trace GC-MS Ultra system with DSQII (Thermo Scientific, Waltham, MA, USA). Vials were incubated for 30 min at $100\text{ }^{\circ}\text{C}$ and a TriPlus autosampler was used to inject $500\text{ }\mu\text{L}$ of the sample in split mode onto a SGE BPX5 column (30 m, I.D. 0.25 mm , film $0.25\text{ }\mu\text{m}$); an injector temperature of $280\text{ }^{\circ}\text{C}$ was used. The initial oven temperature was kept at $50\text{ }^{\circ}\text{C}$ for 2.5 min. The temperature was increased with a ramp rate of $10\text{ }^{\circ}\text{C}/\text{min}$ to $320\text{ }^{\circ}\text{C}$ with a final hold for 5 min. Helium was used as a carrier gas with a flow rate of $1.2\text{ mL}/\text{min}$ and a split ratio of 8. The MS chromatograms and spectra were recorded at 70 eV (EI). Masses were detected between $50\text{ }m/z$ and $650\text{ }m/z$ in the positive mode [69]. Samples were measured in biological triplicates and the area average was used to compare peaks. Compounds were identified by spectral comparison with a NIST/EPA/NIH MS library version 2.0. To provide insight into the differentiation between plant samples a PCA was conducted.

3.3. High Molecular Weight DNA Extraction and Library Preparation

High molecular weight genomic DNA (HMW gDNA) suitable for long-read sequencing was achieved using a plant-optimized CTAB—PCI extraction method based on different protocols [70–72]; 1 g of frozen, unthawed plant leaves were ground using a CryoMill (Retsch, Haan, Germany; three cycles, 6 min precool at 5 Hz , disruption 2:30 min 25 Hz , cooling between cycles 0:30 min at 5 Hz). A CTAB extraction buffer (2% CTAB, 100 mM Tris pH 8.0, 20 mM EDTA, 1.4 M NaCl) was supplemented with 2% PVP prior to usage and solved at $60\text{ }^{\circ}\text{C}$. The unthawed fine powder was mixed with 5 ml buffer and incubated with $200\text{ }\mu\text{L}$ Proteinase K (Qiagen, Venlo, The Netherlands) for 30 min at $50\text{ }^{\circ}\text{C}$ and occasionally inverted. At room temperature, 1 mg RNase A (Thermo Scientific, Waltham, MA, USA) was added and incubated for 10 min. The mixture was washed twice, saving and reusing the aqueous upper phase, with one volume PCI (25:24:1) and three times with chloroform ($10,000\times g$, 5 min, $10\text{ }^{\circ}\text{C}$). To pellet the HMW gDNA, 30% PEG was added to the aqueous phase (1:4), inverted, incubated for 30 min on ice and spun for 30 min at $12,000\times g$, $10\text{ }^{\circ}\text{C}$. The resulting shallow and colorless pellet was washed three times with 70% ethanol ($5000\times g$, 5 min, $10\text{ }^{\circ}\text{C}$) and consequently, air dried at $40\text{ }^{\circ}\text{C}$ and resuspended with $100\text{ }\mu\text{L}$ elution buffer (Qiagen, Venlo, The Netherlands). Quality and size of the gDNA were assessed using a Qubit dsDNA HS Kit (Thermo Scientific, Waltham, MA, USA), a Nanodrop photometer (Implen, Munich, Germany) and a Femto Pulse system (Agilent, Santa Clara, CA, USA), respectively. If variations in DNA concentration between Qubit and Nanodrop were $> 50\%$ an AMPure PB bead clean up or an electrophoretic clean up using a BluePippin system (Sage Science, Beverly, MA, USA) was performed; $5\text{ }\mu\text{g}$ HMW gDNA were sheared in a gTube (Covaris, Woburn, MA, USA; $1700\times g$) and used for whole genome library preparation using SMRTbell prep kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. Size selection of the resulting library was performed using AMPure PB beads. Libraries were stored at $-20\text{ }^{\circ}\text{C}$. Prior sequencing, primer and polymerase were bound using a Sequel II Binding Kit 3.2 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations.

3.4. Genome Sequencing and Assembly

Sequencing was performed on a Sequel IIE (Pacific Biosciences, Menlo Park, CA, USA) with two hours pre-extension, two hours adaptive loading (target $p1 + p2 = 0.95$) to an on-plate concentration of 85 pM, and 30 h movie time. The initial de novo genome assembly was performed using SMRT Link (v11.0.0+, Pacific Biosciences, Menlo Park, CA, USA) which uses Improved Phased Assembly (IPA) [73]. After polishing, the contigs were divided into primary and haplotype-associated contigs using `purge_dups` [74].

The assembled sequences can be found within the National Center for Biotechnology Information (NCBI). BioSample accession number: Dark Knight SAMN32308289, Pink Perfection SAMN32308290.

3.5. RNA Long Read IsoSeq

To increase the quality of the genome assembly, long-read transcripts were sequenced to add more depth and accuracy to the proposed gene models. For RNA extraction, frozen, unthawed leaves were ground using a CryoMill and an RNeasy Plant Mini Kit (Qiagen, Venlo, Niederlande). A Turbo DNA free Kit (Invitrogen) was used to further clean the RNA. The high-quality RNA was used to perform an IsoSeq library prep using SMRTbell prep kit 3.0 and Sequel II Binding Kit 3.2. (Pacific Biosciences, Menlo Park, CA, USA).

3.6. Bioinformatic and Statistical Analysis

Gene models were prepared through AUGUSTUS [50–53] using genomic data and long-read transcriptomic data as hints. Quality and completeness of the genome were estimated with QUAST (v5.2.0) [31] and BUSCO (v5.3.2) [39,43,75,76]. NCBI BLAST (v2.12.0+) [40,41] and InterProScan (v5.54-87) [60,61] were computed on a local computational unit. This analysis provided an annotation that was the basis for the determination of distinct protein families, in this case, terpene synthases and cytochrome p450 enzymes. EggNOG Mapper (v2.1.5) was used to determine COG and GO terms. Statistical analysis and figures were conducted using R (v4.2.1, revigo [59] and cateGORizer [77]). Synteny analysis was performed using Mauve [47] (v2.4.0) and Geneious Prime (Geneious). For k-mer analysis jellyfish (v2.3.0) [78] was used (k-mer size: 20). GenomeScope [79,80] was used for the visualization of k-mer frequencies. The following analyses were conducted using galaxy project [81]: BUSCO, QUAST, EggNOG, Jellyfish, and GenomeScope. If not further specified default parameters were used for analysis.

3.7. Identification of TPS and Cytochrome p450 Enzymes

Genes associated with these protein classes were found using InterProScan and the domain seed files IPR036965 (TPS activity) and IPR01128 (cytochrome p450 enzymes). The phylogenetic tree was constructed using a global alignment with Blosum62. As a genetic distance model, Jukes–Cantor was chosen along with Neighbor-Joining as the Tree building method. The outlier was *Physcomitrella patens*, XP_024380398. Software used: Geneious Prime (Geneious).

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12030632/s1>, Figure S1: Chemical structure of D-limonene backbone and difference to C6–C4 shift in α -pinene, Figure S2: Cultivars of *Caryopteris x clandonensis* used in this manuscript, Figure S3: PacBio sequencing quality reports of different *Caryopteris x clandonensis* cultivars, Figure S4: GenomeScope profile of k-mer analysis of Dark Knight, Figure S4: GenomeScope profile of k-mer analysis of Pink Perfection, Figure S6: Synteny evaluation between the *Caryopteris x clandonensis* cultivars, Table S1: GC-MS Headspace data of TOP30 identified compounds via NIST database Table S2: Data Pink Perfection COG, Table S3: Data Dark Knight COG, Table S4: Data Pink Perfection GO cluster, Table S5: Data Dark Knight GO cluster, Supplemental Lamiaceae Reference: Phylogenetic tree references in FASTA format.

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., N.A.; resources, T.B.; data curation, M.R. and N.A.; writing—original draft preparation, M.R. and N.A.; 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 available in a publicly accessible repository. The data presented in this study are openly available in National Center for Biotechnology Information (NCBI). BioSample accession number: Dark Knight SAMN32308289, Pink Perfection SAMN32308290.

Acknowledgments: M.R., N.A., N.M. and T.B. gratefully acknowledge the support of colleagues at the Werner Siemens Chair for Synthetic Biotechnology during conducting experiments and writing this manuscript. Furthermore, all authors want to thank Foerstner Pflanzen GmbH for providing plant materials.

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

- Caputi, L. Use of terpenoids as natural flavouring compounds in food industry. *Recent Pat. Food Nutr. Agric.* **2011**, *3*, 9–16. [[CrossRef](#)] [[PubMed](#)]
- Masyita, A.; Sari, R.M.; Astuti, A.D.; Yasir, B.; Rumata, N.R.; Emran, T.B.; Nainu, F.; Simal-Gandara, J. Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food Chem. X* **2022**, *13*, 100217. [[CrossRef](#)] [[PubMed](#)]
- da Silva, G.L.; Luft, C.; Lunardelli, A.; Amaral, R.H.; Melo, D.A.D.S.; Donadio, M.V.; Nunes, F.B.; DE Azambuja, M.S.; Santana, J.C.; Moraes, C.M.; et al. Antioxidant, analgesic and anti-inflammatory effects of lavender essential oil. *An. Acad. Bras. Cienc.* **2015**, *87*, 1397–1408. [[CrossRef](#)] [[PubMed](#)]
- Mediratta, P.; Sharma, K.; Singh, S. Evaluation of immunomodulatory potential of *Ocimum sanctum* seed oil and its possible mechanism of action. *J. Ethnopharmacol.* **2002**, *80*, 15–20. [[CrossRef](#)] [[PubMed](#)]
- da Silva, J.K.R.; Figueiredo, P.L.B.; Byler, K.G.; Setzer, W.N. Essential oils as antiviral agents, potential of essential oils to treat SARS-CoV-2 infection: An in-silico investigation. *Int. J. Mol. Sci.* **2020**, *21*, 3426. [[CrossRef](#)] [[PubMed](#)]
- Abdollahi, M.; Karimpour, H.; Monsef-Esfehani, H.R. Antinociceptive effects of *Teucrium polium* L. total extract and essential oil in mouse writhing test. *Pharmacol. Res.* **2003**, *48*, 31–35. [[CrossRef](#)] [[PubMed](#)]
- Đorđević, S.; Petrović, S.; Dobrić, S.; Milenković, M.; Vučićević, D.; Žižić, S.; Kukić, J. Antimicrobial, anti-inflammatory, anti-ulcer and antioxidant activities of *Carlina acanthifolia* root essential oil. *J. Ethnopharmacol.* **2007**, *109*, 458–463. [[CrossRef](#)]
- Cowen, D.; Wolf, A.; Paige, B.H. Toxoplasmic encephalomyelitis. *Arch. Neurol. Psychiatry* **1942**, *48*, 689–739. [[CrossRef](#)]
- Jantan, I.; Ping, W.O.; Visuvalingam, S.D.; Ahmad, N.W. Larvicidal activity of the essential oils and methanol extracts of Malaysian plants on *Aedes aegypti*. *Pharm. Biol.* **2008**, *41*, 234–236. [[CrossRef](#)]
- Cox-Georgian, D.; Ramadoss, N.; Dona, C.; Basu, C. Therapeutic and medicinal uses of terpenes. *Med. Plants Farm Pharm.* **2019**, *67*, 333–359. [[CrossRef](#)]
- Sicora, O. The ethanolic stem extract of *Caryopteris x Clandonensis* Posseses antiproliferative potential by blocking breast cancer cells in mitosis. *Farmacologia* **2019**, *67*, 1077–1082. [[CrossRef](#)]
- Wani, M.C.; Taylor, H.L.; Wall, M.E.; Coggon, P.; Mcphail, A.T. Plant antitumor agents. VI. The isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **1971**, *93*, 2325–2327. [[CrossRef](#)]
- Weaver, B.A. How Taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **2014**, *25*, 2677–2681. [[CrossRef](#)]
- Pichersky, E.; Raguso, R.A. Why do plants produce so many terpenoid compounds? *New Phytol.* **2018**, *220*, 692–702. [[CrossRef](#)]
- Holopainen, J.K.; Himanen, S.J.; Yuan, J.S.; Chen, F.; Stewart, C.N. *Ecological Functions of Terpenoids in Changing Climates*; Ramawat, K., Mérillon, J.M., Eds.; Natural Products; Springer: Berlin/Heidelberg, Germany, 2013. [[CrossRef](#)]
- Drapeau, J.; Rossano, M.; Touraud, D.; Obermayr, U.; Geier, M.; Rose, A.; Kunz, W. Green synthesis of para-Menthane-3,8-diol from *Eucalyptus citriodora*: Application for repellent products. *Comptes Rendus Chim.* **2011**, *14*, 629–635. [[CrossRef](#)]
- Lee, S.Y.; Kim, S.H.; Hong, C.Y.; Park, S.Y.; Choi, I.G. Biotransformation of (–)- α -pinene and geraniol to α -terpineol and p-menthane-3,8-diol by the white rot fungus, *Polyporus brumalis*. *J. Microbiol.* **2017**, *53*, 462–467. [[CrossRef](#)]

18. Drapeau, J.; Verdier, M.; Touraud, D.; Kröckel, U.; Geier, M.; Rose, A.; Kunz, W. Effective insect repellent formulation in both surfactantless and classical microemulsions with a long-lasting protection for human beings. *Chem. Biodivers.* **2009**, *6*, 934–947. [[CrossRef](#)]
19. Blythe, E.; Tabanca, N.; Demirci, B.; Bernier, U.; Agramonte, N.; Ali, A.; Khan, I. 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.
20. Bathe, U.; Tissier, A. Cytochrome P450 enzymes: A driving force of plant diterpene diversity. *Phytochemistry* **2019**, *161*, 149–162. [[CrossRef](#)]
21. Hernandez-Ortega, A.; Vinaixa, M.; Zebec, Z.; Takano, E.; Scrutton, N.S. A toolbox for diverse Oxyfunctionalisation of monoterpenes. *Sci. Rep.* **2018**, *8*, 1–8. [[CrossRef](#)]
22. Mabou, F.D.; Belinda, I.; Yossa, N. TERPENES: Structural classification and biological activities. *IOSR J. Pharm. Biol. Sci. e-ISSN* **2021**, *16*, 2319–7676.
23. Nett, R.S.; Montanares, M.; Marcassa, A.; Lu, X.; Nagel, R.; Charles, T.C.; Hedden, P.; Rojas, M.C.; Peters, R.J. Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. *Nat. Chem. Biol.* **2016**, *13*, 69–74. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. Wen, W.; Yu, R. Artemisinin biosynthesis and its regulatory enzymes: Progress and perspective. *Pharmacogn. Rev.* **2011**, *5*, 189–194. [[CrossRef](#)]
26. Gershenzon, J.; Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **2007**, *3*, 408–414. [[CrossRef](#)]
27. Zhang, X.; Niu, M.; da Silva, J.A.T.; Zhang, Y.; Yuan, Y.; Jia, Y.; Xiao, Y.; Li, Y.; Fang, L.; Zeng, S.; et al. Identification and functional characterization of three new terpene synthase genes involved in chemical defense and abiotic stresses in *Santalum album*. *BMC Plant Biol.* **2019**, *19*, 115. [[CrossRef](#)]
28. Sharma, V.; Sarkar, I.N. Bioinformatics opportunities for identification and study of medicinal plants. *Brief. Bioinform.* **2013**, *14*, 238–250. [[CrossRef](#)]
29. Helfrich, E.J.N.; Lin, G.-M.; Voigt, C.A.; Clardy, J. Bacterial terpene biosynthesis: Challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **2019**, *15*, 2889–2906. [[CrossRef](#)]
30. Amarsinghe, S.L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M.E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **2020**, *21*, 1–16. [[CrossRef](#)]
31. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
32. 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](#)]
33. Chen, G.; Mostafa, S.; Lu, Z.; Du, R.; Cui, J.; Wang, Y.; Liao, Q.; Lu, J.; Mao, X.; Chang, B.; et al. The Jasmine (*Jasminum sambac*) genome provides insight into the biosynthesis of flower fragrances and Jasmonates. *Genom. Proteom. Bioinform.* **2022**, in press. [[CrossRef](#)]
34. Degenhardt, J.; Köllner, T.G.; Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **2009**, *70*, 1621–1637. [[CrossRef](#)]
35. Zhu, X.; Li, Q.; Li, J.; Luo, J.; Chen, W.; Li, X. Comparative study of volatile compounds in the fruit of two banana cultivars at different ripening stages. *Molecules* **2018**, *23*, 2456. [[CrossRef](#)]
36. Cramer, A.-C.J.; Mattinson, D.S.; Fellman, J.K.; Baik, B.-K. Analysis of volatile compounds from various types of barley cultivars. *J. Agric. Food Chem.* **2005**, *53*, 7526–7531. [[CrossRef](#)]
37. Dong, A.-X.; Xin, H.-B.; Li, Z.-J.; Liu, H.; Sun, Y.-Q.; Nie, S.; Zhao, Z.-N.; Cui, R.-F.; Zhang, R.-G.; Yun, Q.-Z.; et al. High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. *Gigascience* **2018**, *7*, giy068. [[CrossRef](#)]
38. Li, C.; Li, X.; Liu, H.; Wang, X.; Li, W.; Chen, M.-S.; Niu, L.-J. Chromatin architectures are associated with response to dark treatment in the oil crop *Sesamum indicum*, based on a high-quality genome assembly. *Plant Cell Physiol.* **2020**, *61*, 978–987. [[CrossRef](#)]
39. Manni, M.; Berkeley, M.R.; Seppely, M.; A Simão, F.; Zdobnov, E.M. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [[CrossRef](#)]
40. 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](#)]
41. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
42. Guan, D.; A McCarthy, S.; Wood, J.; Howe, K.; Wang, Y.; Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **2020**, *36*, 2896–2898. [[CrossRef](#)] [[PubMed](#)]
43. Manni, M.; Berkeley, M.R.; Seppely, M.; Zdobnov, E.M. BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* **2021**, *1*, e323. [[CrossRef](#)] [[PubMed](#)]

44. Tang, H.; Lyons, E.; Pedersen, B.; Schnable, J.C.; Paterson, A.H.; Freeling, M. Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinform.* **2011**, *12*, 102. [[CrossRef](#)] [[PubMed](#)]
45. Lee, J.; Hong, W.-Y.; Cho, M.; Sim, M.; Lee, D.; Ko, Y.; Kim, J. Synteny Portal: A web-based application portal for synteny block analysis. *Nucleic Acids Res.* **2016**, *44*, W35–W40. [[CrossRef](#)] [[PubMed](#)]
46. Liu, D.; Hunt, M.; Tsai, I.J. Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinform.* **2018**, *19*, 1–13. [[CrossRef](#)]
47. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)]
48. Arús, P.; Toshiya, Y.; Elisabeth, D.; Abbott, A.G. Synteny in the Rosaceae. In *Plant Breeding Reviews*; John and Wiley and Sons: Hoboken, NJ, USA, 2010; pp. 175–211.
49. Devos, K.M.; Moore, G.; Gale, M.D. Conservation of marker synteny during evolution. *Euphytica* **1995**, *85*, 67–372. [[CrossRef](#)]
50. Hoff, K.J.; Lomsadze, A.; Borodovsky, M.; Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **2019**, *1962*, 65–95. [[CrossRef](#)]
51. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS: Table 1. *Bioinformatics* **2016**, *32*, 767–769. [[CrossRef](#)]
52. Brúna, 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](#)]
53. 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](#)]
54. Yang, Z.; Ge, X.; Yang, Z.; Qin, W.; Sun, G.; Wang, Z.; Li, Z.; Liu, J.; Wu, J.; Wang, Y.; et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* **2019**, *10*, 2989. [[CrossRef](#)]
55. Liu, S.; An, Y.; Tong, W.; Qin, X.; Samarina, L.; Guo, R.; Xia, X.; Wei, C. Characterization of genome-wide genetic variations between two varieties of tea plant (*Camellia sinensis*) and development of InDel markers for genetic research. *BMC Genom.* **2019**, *20*, 1–16. [[CrossRef](#)]
56. Chatterjee, N.; Walker, G.C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **2017**, *58*, 235–263. [[CrossRef](#)] [[PubMed](#)]
57. Raina, A.; Sahu, P.K.; Laskar, R.A.; Rajora, N.; Sao, R.; Khan, S.; Ganai, R.A. Mechanisms of genome maintenance in plants: Playing it safe with breaks and bumps. *Front. Genet.* **2021**, *12*, 675686. [[CrossRef](#)]
58. Lim, C.; Pratama, M.Y.; Rivera, C.; Silvestro, M.; Tsao, P.S.; Maegdefessel, L.; Gallagher, K.A.; Maldonado, T.; Ramkhelawon, B. Linking single nucleotide polymorphisms to signaling blueprints in abdominal aortic aneurysms. *Sci. Rep.* **2022**, *12*, 20990. [[CrossRef](#)]
59. Supek, F.; Bošnjak, M.; Škunca, N.; Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [[CrossRef](#)]
60. Blum, M.; Chang, H.-Y.; Chuguransky, S.; Grego, T.; Kandasamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **2021**, *49*, D344–D354. [[CrossRef](#)]
61. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
62. Butler, J.B.; Freeman, J.; Potts, B.M.; Vaillancourt, R.; Grattapaglia, D.; Silva-Junior, O.B.; Simmons, B.; Healey, A.L.; Schmutz, J.; Barry, K.; et al. Annotation of the *Corymbia terpena* synthase gene family shows broad conservation but dynamic evolution of physical clusters relative to *Eucalyptus*. *Heredity* **2018**, *121*, 87–104. [[CrossRef](#)]
63. Warren, R.L.; Keeling, C.I.; Yuen, M.M.S.; Raymond, A.; Taylor, G.A.; Vandervalk, B.P.; Mohamadi, H.; Paulino, D.; Chiu, R.; Jackman, S.D.; et al. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant J.* **2015**, *83*, 189–212. [[CrossRef](#)] [[PubMed](#)]
64. Jia, K.-H.; Liu, H.; Zhang, R.-G.; Xu, J.; Zhou, S.-S.; Jiao, S.-Q.; Yan, X.-M.; Tian, X.-C.; Shi, T.-L.; Luo, H.; et al. Chromosome-scale assembly and evolution of the tetraploid *Salvia splendens* (*Lamiaceae*) genome. *Hortic. Res.* **2021**, *8*, 1–15. [[CrossRef](#)] [[PubMed](#)]
65. Chen, Z.; Vining, K.J.; Qi, X.; Yu, X.; Zheng, Y.; Liu, Z.; Fang, H.; Li, L.; Bai, Y.; Liang, C.; et al. 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](#)]
66. Hamilton, J.P.; Godden, G.T.; Lanier, E.; Bhat, W.W.; Kinser, T.J.; Vaillancourt, B.; Wang, H.; Wood, J.C.; Jiang, J.; Soltis, P.S.; et al. Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing *Lamiaceae* species, *Callicarpa americana*. *Gigascience* **2020**, *9*, gaa093. [[CrossRef](#)] [[PubMed](#)]
67. 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](#)]
68. Shalev, T.J.; Yuen, M.M.S.; Gesell, A.; Yuen, A.; Russell, J.H.; Bohlmann, J. An annotated transcriptome of highly inbred *Thuja plicata* (*Cupressaceae*) and its utility for gene discovery of terpenoid biosynthesis and conifer defense. *Tree Genet. Genomes* **2018**, *14*, 35. [[CrossRef](#)]
69. Ringel, M.; Reinbold, M.; Hirte, M.; Haack, M.; Huber, C.; Eisenreich, W.; Masri, M.A.; Schenk, G.; Guddat, L.W.; Loll, B.; et al. Towards a sustainable generation of pseudopterisin-type bioactives. *Green Chem.* **2020**, *22*, 6033–6046. [[CrossRef](#)]

70. Inglis, P.W.; Pappas, M.; Resende, L.V.; Grattapaglia, D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS ONE* **2018**, *13*, e0206085. [[CrossRef](#)]
71. Healey, A.; Furtado, A.; Cooper, T.; Henry, R.J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **2014**, *10*, 21. [[CrossRef](#)]
72. Rogers, S.O.; Bendich, A.J. Extraction of total cellular DNA from plants, algae and fungi. *Plant Mol. Biol. Man.* **1994**, *2*, 183–190. [[CrossRef](#)]
73. GitHub—PacificBiosciences/pbipa: Improved Phased Assembler. Available online: <https://github.com/PacificBiosciences/pbipa> (accessed on 11 December 2022).
74. GitHub—dfguan/purge_dups: Haplotypic Duplication Identification Tool. Available online: https://github.com/dfguan/purge_dups (accessed on 11 December 2022).
75. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; A Simão, F.; Zdobnov, E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **2019**, *47*, D807–D811. [[CrossRef](#)] [[PubMed](#)]
76. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **2019**, *1962*, 227–245. [[CrossRef](#)] [[PubMed](#)]
77. Hu, Z.-L.; Bao, J.; Reecy, J. CateGORizer: A web-based program to batch analyze gene ontology classification categories. *Online J. Bioinform.* **2008**, *9*, 108–112. Available online: <http://www.animalgenome.org/bioinfo/tools/catego/> (accessed on 11 December 2022).
78. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770. [[CrossRef](#)]
79. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [[CrossRef](#)]
80. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **2020**, *11*, 1432. [[CrossRef](#)]
81. Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Čech, 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](#)]

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.