Article

Transcriptome Landscape Variation in the Genus *Thymus*

Aboozar Soorni [1],*, Tahereh Borna [2], Ali Alemardan [2], Manohar Chakrabarti [3], Arthur G. Hunt [3] and Aureliano Bombarely [4,5,*

[1] Department of Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan 84156-83111, Iran
[2] Department of Horticultural Science, College of Agriculture and Natural Resources, University of Tehran, Karaj 31587-11167, Iran
[3] Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546-0312, USA
[4] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24060, USA
[5] Department of Bioscience, University of Milano, 20133 Milan, Italy
* Correspondence: soorni@cc.iut.ac.ir (A.S.); aureliano.bombarely@unimi.it (A.B.)

check for updates

Abstract: Among the *Lamiaceae* family, the genus *Thymus* is an economically important genera due to its medicinal and aromatic properties. Most *Thymus* molecular research has focused on the determining the phylogenetic relationships between different species, but no published work has focused on the evolution of the transcriptome across the genus to elucidate genes involved in terpenoid biosynthesis. Hence, in this study, the transcriptomes of five different *Thymus* species were generated and analyzed to mine putative genes involved in thymol and carvacrol biosynthesis. High-throughput sequencing produced ~43 million high-quality reads per sample, which were assembled de novo using several tools, then further subjected to a quality evaluation. The best assembly for each species was used as queries to search within the UniProt, KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Clusters of Orthologous Groups) and TF (Transcription Factors) databases. Mining the transcriptomes resulted in the identification of 592 single-copy orthogroups used for phylogenetic analysis. The data showed strongly support a close genetic relationship between *Thymus vulgaris* and *Thymus daenensis*. Additionally, this study dates the speciation events between 1.5–2.1 and 9–10.2 MYA according to different methodologies. Our study provides a global overview of genes related to the terpenoid pathway in *Thymus*, and can help establish an understanding of the relationship that exists among *Thymus* species.

Keywords: *Thymus*; terpenoid biosynthesis; transcriptome; phylotranscriptome

1. Introduction

Thymus is one of the most consequential genera belonging to the *Lamiaceae* family, and contains more than 336 species of aromatic perennial herbaceous plants with valuable medicinal properties [1,2]. Most species originated in the Mediterranean region but have since have been distributed worldwide [3]. Several members of the genus are also cultivated as culinary herbs or for ornamentation [4].

Many studies have investigated the chemical composition and biological properties of essential oils in different species of *Thymus* [5,6]. The six major constituents showing variations among *Thymus* species include thymol, carvacrol, γ-terpinene, geraniol, linalool, and p-cymene, [5–7]. In regard to previous research, the most important factors that strongly affect the quantities and chemical compositions of essential oils are species, environmental parameters, growth regions, cultivation practices, and plant growth stages [5–7]. Thymol has been identified as the major constituent in *Thymus*

vulgaris [7,8] *Thymus daenensis* [9,10], and *Thymus lancifolius* [10], while carvacrol is known to be major component in *Thymus pubescens* [11,12] and *Thymus persicus* [13].

Since the introduction of next generation sequencing technology (NGS), there has been a major transformation due to the possibility to extracting abundant genetic information from biological systems through the genome and transcriptome sequence data of any species. NGS has made it possible, in a single sequencing process, to generate millions of genomic DNA or transcriptomic fragments in record time and at ever-lowering costs. NGS has catalyzed a number of important breakthroughs in fields such as agriculture or evolutionary science [14,15]. This tool provides a powerful method of developing molecular markers, studying evolutionary processes, and discovering natural biosynthetic pathways in plants, especially in species with unknown genomes. De novo transcriptome assembly is a critically important approach to exploring gene sequences in organisms without genome sequences [16], and recent progress in RNA-sequencing techniques has provided a fast, reliable, and affordable framework for transcriptomic studies [17].

Since the molecular biology underlying the terpenoid pathway in the genus *Thymus* has not been sufficiently investigated yet, RNA-Seq technology was applied to five different species to explore (a) the molecular basis and potential genes involved in terpenoid biosynthesis, with a focus on thymol and carvacrol biosynthesis and transcription factor (TF) families; and (b) comparative phylogenetic analysis based on single-copy orthogroups, to elucidate the origin of these species. This study will supply basic genetic resources for understanding the biology of the genus *Thymus* and reinforce further research endeavors.

2. Material and Methods

2.1. Ethics Statement

All Thymus seeds used in this study were collected from natural fields. No specific permits were required according the Iranian laws, and no endangered, vulnerable, threatened, or protected species were involved.

2.2. Plant Material

Seeds of *Thymus* accessions were collected from different locations in Iran during August 2015. The information of the studied samples is shown in Table 1. Sowing seeds was carried out in plastic pots with a perlite–coco peat mixture (1:1 *v/v*), then grown in a greenhouse under controlled temperature conditions for 45 days at the greenhouse facilities of Virginia Tech, beginning in April 2016.

Table 1. Mean (three replicate measurements per plant sample) 2C-value and genome size of *Thymus* accessions.

Species	Location (Province)	1C-Value ± SD (pg)	Haploid Size (Mbp)	Estimated Polyploidy
Thymus pubescens	East Azerbaijan	0.63 ± 0.02	616.14	2n = 2x
T. pubescens	West Azerbaijan	1.11 ± 0.01	1085.84	2n = 4x
Thymus migricus	West Azerbaijan	0.61 ± 0.01	596.72	2n = 2x
Thymus daenensis	Lorestan	0.56 ± 0.01	547.81	2n = 2x
T. daenensis	Fereydunshahr	0.55 ± 0.27	538.03	2n = 2x
T. daenensis	Arak	0.54 ± 0.28	528.25	2n = 2x
Thymus persicus	West Azerbaijan	1.08 ± 0.01	1056.49	2n = 2x
Thymus kotschyanus	Qazvin	1.09 ± 0.05	1066.28	2n = 2x
T. kotschyanus	Qazvin	1.09 ± 0.04	1066.28	2n = 2x
Thymus lancifolius	Isfahan	0.56 ± 0.01	547.81	2n = 2x
T. lancifolius	Kordestan	1.11 ± 0.04	1085.84	2n = 4x
Thymus vulgaris	Cultivated	0.82 ± 0.01	802.15	2n = 2n
Thymus fedtschenkoi	Semnan	0.61 ± 0.01	596.72	2n = 2x
T. fedtschenkoi	Yazd	1.17 ± 0.02	1144.54	2n = 4x
Thymus transcaspicus	Khorasan	1.09 ± 0.02	1066.28	2n = 2x
T. transcaspicus	Yazd	1.07 ± 0.02	1046.71	2n = 2x
Thymus fallax	Yazd	1.14 ± 0.02	1115.19	2n = 2x
Thymus carmanicus	Kerman	1.08 ± 0.02	1056.49	2n = 2x

2.3. Genome Size Measures

Fifty mg of fresh leaf tissue was chopped in 1 mL of DeLaat's buffer [18] for 20–60 s with the *Capsicum annum* lab strain as an internal standard. Suspensions of nuclei were filtered through 20–50 μm Celltrics filters and centrifuged for 5 min at 200 RPM, then reduced to 200 μl by carefully removing the upper solution. Next, 400 μl of staining solution (propidium iodide (PI)) was added to each tube, mixed well by gently shaking the tubes by hand, then incubated at room temperature in darkness for 20–30 min. Flow cytometry measurement was carried out with a Partec PAS flow cytometer (Partec, http://www.partec.de/, Münster, Germany), then nuclear DNA content was calculated using the following formula [19]:

$$\text{Sample 2C DNA (pg) content} = ((\text{Sample G1 peak mean})/(\text{Standard G1 peak mean})) \times \text{Standard 2C DNA content (pg DNA).}$$

2.4. RNA Extraction, Library Preparation and Sequencing

Based on the genome size pattern, five different species that could cover the whole genome size distribution were chosen for the transcriptome sequencing. Total RNA was extracted from young green leaves for each accession obtained from three different plants using the RNEasy Plant Mini kit (Qiagen, Hilden, Germany), following the manufacturer's instructions for the library preparation and sequencing. RNA quality and quantity were verified with a Nanodrop spectrophotometer and Agilent 2100 Bioanalyzer.

A fraction of 2 μg of total RNA was used as the starting material for library synthesis using the SMART approach [20]. First, the poly-A mRNA in the total RNA was pulled down using poly-T oligo-attached magnetic beads. After purification, the mRNA was fragmented at 95 °C for 2 min along with RT primer and first strand buffer. The fragmented RNA was used for the synthesis of the first-strand cDNA by adding DTT, dNTPs, RNase inhibitor, and SMARTScribe. Reactions were incubated for 2 h at 42 °C. This was followed by second-strand cDNA synthesis using a template-switching oligonucleotide and an additional 1 μL of SMARTScribe. Subsequently, the cDNAs were purified using two rounds of AMPure bead purification. The samples were amplified using PCR to create the final library. The libraries were confirmed by electrophoresis on a 1% TAE agarose gel and the Agilent BioAnalyzer 2100 after purification by the AMPure bead. The read sizes were estimated to range between 250–300 bp. One library was prepared per accession. In the last step, equal amounts of quantified libraries were pooled to be sequenced on the Illumina HiSeq 4000 system at the Duke Center for Genomics and Computational Biology with a length of 2 × 150 bp (pair ends).

2.5. Data Processing, Transcriptome Assembly and Annotation

Prior to analysis of data, raw reads were processed to filter out low quality nucleotides (<Q30), contaminant sequence, adapters, and short reads (<50 bp). Pre-processing quality was performed by using the Trimmomatic v0.32 program [21] with ILLUMINACLIP:2:30:10, LEADING:5, TRAILING:5, SLIDINGWINDOW:4:30, and MINLEN:50 parameters. Then, all RNA-Seq data were deposited in the NCBI SRA database under the project PRJNA481444.

Only clean reads were used in downstream analysis. De novo transcriptome assemblies were constructed and evaluated with the four following tools: (1) Trinity v2.4.0, using a k-mer size of 25 [22]. (2) Velvet v1.2.10 [23] followed by Oases v0.2.09 [24] with a series of k-mer sizes from 21 to 61 (step size 2). The parameters used with Velvet were "-shortPaired and -read_trkg" and Oases "-min_trans_lgth 200" to set the minimum length for a transcript in the output file to 200 bp. (3) BinPacker v1.0 [25]. (4) Bridger [26]. Similar to Trinity, BinPacker and Bridger programs were used with a k-mer size of 25, because this k-mer size works well for most organisms.

All de novo assemblies were performed with default settings. To evaluate the quality of transcriptome assemblies (assembly completeness), the number of full-length or nearly full-length transcripts was counted. In this case, a sequence homology search was run on all de novo assemblies

with BLASTX and the Swiss-Prot/UniProt database [27]. Then, the BLASTX hits were filtered from annotated transcriptomes with an e-value $> 1 \times 10^{-20}$. Next, the length coverage percentage for the top matching database entries was examined using analyze_blastPlus_topHit_coverage.pl script from the Trinity package [22]. The detonate score was calculated by DETONATE's RSEM-EVAL [27] as an additional criterion to evaluate the quality of different assemblies.

TransDecoder and Trinotate software suites were used for functional annotation of each assembly following the method outlined at http://trinotate.github.io/. TransDecoder v2.0.1 (available at http://transdecoder.sourceforge.net/) was used to predict open reading frames (ORFs) of at least 100 amino acids in length. To further maximize sensitivity for capturing ORFs, all ORFs were scanned for homology to known proteins (including a BLASTP search against UniProtKB/Swiss-Prot, and a search of the peptides for protein domains using Pfam). After extraction of ORFs from the assembly, the Trinotate v2.0.2 (available at http://trinotate.github.io) was preferred for assigning functions, as the package uses multiple databases for annotation. The Trinotate annotation pipeline comprises several software packages such as BLASTX, BLASTP, Pfam search, SignalP, and RNAmmer. Initially, transcripts were searched against the Swiss-Prot protein database by running BLASTX, allowing one hit and with the output in tabular format. Then, the expected protein translations obtained using TransDecoder were searched against the Swiss-Prot/UniProt databases. Functional domains were identified using HMMER v.3.1b2 tools and the Pfam domain database [28,29]. Potential signal peptides were predicted using SignalP v4.1 [30]. TMHMM v.2.0c was used for prediction of transmembrane helices in proteins [31] and RNAmmer v.1.2 was applied to predict ribosomal RNA [28]. All results were deposited into the Trinotate-provided SQLite database, and a spreadsheet summary report was generated. WEGO software [32] was used for Gene ontology (GO) functional classification using the Trinotate-integrated UniProtKB GO annotations. Furthermore, a COG screening for phylogenetic classification was performed using predicted proteins and the EggNog database [33], with an e-value threshold of 1×10^{-5}. TFs were identified and classified using iTAK (http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi), and finally unigenes involved in the monoterpenoid biosynthesis pathways were identified based on the annotation results, and were assessed using BLAST against the NCBI and UniProt databases.

2.6. Identification of Orthologous Genes and Comparative Phylogenetic Analysis

As an initial scenario, OrthoFinder was used [34] with the default parameters to identify orthogroups among predicted protein sequences obtained from Transdecoder v2.0.1 for the five assemblies. We also broadened the sampling by including *Salvia splendens* [35] and *Mentha spicata* (Bioproject ID PRJNA359989) as outgroups. Orthogroups with only single-copy genes were retained for further analysis. The nucleotide sequences for each group were collected into a FASTA file and were multiply aligned using MUSCLE v3.8.31 [36]. Output alignment files were further applied to filter out the poorly aligned regions by the trimAl v1.4 [37] with the parameter "-gt 0.8 -st 0.001". For each alignment sequence, RAxML v8.2.11 [38] was used to reconstruct the maximum likelihood (ML) tree under the GTRGAMMA model with 100 bootstrap replicates. All gene trees were fed into ASTRAL-II [39] to infer the species phylogeny under a multi-species coalescent model. In addition to ML analysis, a Bayesian methodology using BEAST v.2.5.2 [40] was applied as an alternative method to infer the taxa relationship and estimate divergence time based on all single-copy genes. To generate input files for BEAST, the BEAUti interface was used in an HKY model with four γ categories, a strict molecular clock, and a calibrated Yule process of speciation with two calibration points: Mentha-Thymus at 50 MYA (M = 3.88, L = 0.25), and Salvia-Thymus at 30.6 MYA (M = 3.443, L = 0.15). The calibration dates were obtained from TimeTree (http://timetree.org). The chain length was set to 10 million. The trees were then interpreted by the program TreeAnnotator v1.6.1 prior to visualization in the program FigTree v1.4.0 [41].

2.7. Positive Selection Analysis

In the positive selection analysis, the corresponding codon-based nucleotide alignments (single-copy genes) for five *Thymus* species were obtained using PAL2NAL [42] with the help of amino acid alignments. Next, a gene tree was constructed by RAxML [38] using a GTR + γ model for each alignment. In order to do tests for positive selection, the branch-site model was implemented in the CodeML program [43] to estimate the dN/dS (non-synonymous vs synonymous) substitution rates (ω value) for each single-copy orthologous sequence individually. The CodeML analysis used an unrooted tree obtained from relative alignment. We executed CodeML four times to stipulate four separate wild species (*T. daenensis*, *T. lancifolius*, *T. persicus*, and *T. pubescens)* as the foreground lineages. Genes were deemed under positive selection if they yielded statistically significant results for the likelihood ratio test (LRT, *p*-value < 0.01) comparison between the null and alternative models. For each positively selected gene (PSG), functional information was deduced based on its ortholog in *Arabidopsis thaliana*. GO enrichment analyses of PSGs were conducted using the web-based agriGO [44] with the TAIR database, the singular enrichment analysis (SEA) method, 0.01 as the significance level, and 5 as the minimum number of mapping entries.

2.8. Ks Analysis for Whole Genome Duplication (WGD)

A common approach was used based on pairwise divergence between paralogs and orthologs at synonymous sites (Ks) to check for possible whole genome duplication (WGD) events in the genus *Thymus*. The FASTKs pipeline with default settings [45] was used to construct Ks plots. In the FastKs pipeline, translated transcriptomes obtained using TransDecoder were blasted against themselves and each other (a cutoff of 1×10^{-40}) to identify putative paralogs and orthologs. Pairs longer than 100 amino acids were aligned using MUSCLE v3.8.31 [36], then the corresponding codon-based nucleotide alignments were obtained using PAL2NAL [42] with the help of amino acid alignments. The Ks value was calculated for each pair using CodeML in PAML v4.8 [43], with the paired sequence settings (yn00) [46] and the F3x4 model [47]. Finally, mclust v5.0.2 [48] was implemented in R to estimate the normal mixture models for Ks values.

3. Results and Discussion

3.1. Genome Size Variation across the Genus Thymus

Prior to the transcriptome sequencing of the different *Thymus* species, the genome size of the different accessions of the collection was estimated by flow cytometry. The flow cytometry in 18 accessions belonging to 11 species of the genus *Thymus* resulted in a maximum difference of 2.102-fold (Table 1). The smallest haploid genome size was estimated to be 528.25 Mb (1C-value = 0.54 pg) for *T. daenensis*, and the highest value was 1144.54 Mb (1C-value = 1.17 pg) for *Thymus fedtschenkoi*. The haploid genome size for *T. vulgaris* and *Thymus migricus* (West Azerbaijan) were measured to be 802.15 and 596.72 Mbp (1C-value: 0.82 and 0.61 pg), respectively, which is similar to the results from previous reports [49,50]. The haploid genome in two populations of *Thymus kotschyanus*, collected from the same geographical region (Qazvin), was 1066.28 Mbp.

The 1C-value for two populations of *T. lancifolius* varied from 0.56 (Isfahan) to 1.11 (Kordestan), a 1.98-fold difference that shows intraspecific variation in genome size (from 547.81 to 1085.84 Mbp). Two species of *T. fedtschenkoi* and *T. pubescens* showed different genome sizes between populations. Within-species genome size variation has been found in samples from geographically separated populations [51], which probably reflects polyploidy in species populations. *T. persicus* is an endemic species that is only found in a restricted region of Iran [52]. The 1C DNA contents and genome size of this species located in West Azerbaijan measured 1.08 pg and 1056.49 Mbp, respectively. In the three accessions of *T. daenensis*, no variation was observed among the 1C-values of populations (0.56, 0.55, and 0.54 pg, referring to Lorestan, Fereydunshahr, and Arak accessions, respectively). The 1C-value given here is consistent with a previous report that describing the DNA content of similar accessions [50],

except in the case of the Lorestan population. The changes in genome size may not only be restricted to species divergence but also be associated with various environmental conditions and developmental stages affecting different populations or individual plants [53].

3.2. RNA Sequencing and De Novo Transcriptome Assembly

To obtain *Thymus* transcriptome data, paired-end RNA-Seq libraries were constructed for *T. daenensis* (Tdae, from Lorestan), *T. vulgaris* (Tvul), *T. lancifolius* (Tlan, from Kordestan), *T. persicus* (Tper, from West Azerbaijan), and *T. pubescens* (Tpub, from West Azerbaijan). The five libraries described above were represented by 60.98, 64.99, 65.85, 61.81, and 62.20 million reads on Illumina HiSeq 4000, respectively. To facilitate and improve quality and efficiency of assembly, a fraction of reads (28.59% on average, ranging from 22.59% to 32.41%) were removed by filtering adapter sequences and low-quality and short reads (shorter than 50 base pairs). After filtering, a total of 43.38, 43.92, 47.10, 47.62, and 43.38 million high-quality clean reads from the cDNA libraries of Tdae, Tvul, Tlan, Tper, and Tpub were obtained, respectively.

Clean reads obtained from samples were independently assembled de novo using Trinity, Bridger, BinPacker, and Velvet/Oases as described in Section 2, and the resultant assemblies were analyzed to evaluate quality. There are different criteria used to evaluate de novo transcriptome assembly quality, each of which has its advantages and disadvantages. As the first assembly quality metrics, assemblies were evaluated through the number of predicted transcripts and assembly size with different k-mer lengths. Among Trinity and BinPacker with k-mer 25 and Bridger with two k-mers (25 and 27 bp), Trinity produced the highest number of transcripts in all samples (ranging from 85,084 to 123,698) whilst BinPacker and Bridger with k-mer 27 perched on the opposite side. With Velvet/Oases, the total number of transcripts dropped off with an increment in k-mer size, similar to previous studies [54,55], but increased slightly between k-mer 51 and k-mer 57. BinPacker produced a larger assembly size for all samples (ranging from 68.45 to 98.79 Mb) compared to the other programs. Results for the assembly size in Velvet/Oases showed a reversed trend as k-mer size increased. K-mer size plays an important role in increasing or decreasing the number of predicted transcripts and assembly size [55,56]. In theory, large k-mer sizes lead to a decrease in the total number of transcripts, because high k-mer values have a tendency to produce a more contiguous assembly consisting of high-coverage transcripts [56]. This may be due to the capturing of only highly represented reads, or the better assembly of highly expressed transcripts. In contrast, low k-mer values produce a highly fragmented assembly due to sequencing errors and lack of overlap. Short k-mers can lead to the generation of multiple paths in the assembly graph. Therefore, the graph structure can be ambiguous, and the represented isoforms can be challenging to resolve. Because of ambiguous regions, a single path is almost impossible to find in the assembly [22,56]. The results in the present study show that assembly size and the total number of contigs depend on k-mer length, which is highly linked to how the assembler handles the ambiguous paths.

Bridger and BinPacker yielded N50 values ranging from 819 to 1553 bp in Tdae, 732 to 1603 bp in Tvul, 767 to 1634 bp in Tlan, 999 to 1665 bp in Tper, and 818 to 1574 bp in Tpub. BinPacker produced the longest N50 value and had the highest average contig length, while Trinity performed the worst for the N50 length and average contig length in all samples. Bridger had slightly better N50 values in comparison to Trinity, similar to the results from previous research [56]. In Velvet/Oases, the N50 and average contig length slightly increased with a simultaneous increase of k-mer size to k = 29 or k = 31 before deteriorating. However, a longer N50 does not necessarily produce a better assembly, as it may indicate a high level of chimerism [55,57], and previous research indicates that the largest N50 can artificially infer a higher quality of assembly [58,59]. A summary of statistics for all assemblies in all species is shown in Tables S1 and S2.

We further evaluated the different assemblies by estimating the number of recovered full-length transcripts when the Swiss-Prot/UniProt protein database was compared with Blastx. In terms of percentage of full-length transcripts, BinPacker generated more full-length transcripts (ranging from

5403 to 6008), with more than 80% coverage for Tdae and Tvul, respectively, while Velvet/Oases was the worst (Table S3). As in previous studies, our results indicate the Trinity assembler is able to recover more full-length transcripts than Velvet/Oases [21,56,60]. In this case, Bridger assembled more full-length transcripts than Trinity, and was the closest assembler to BinPacker [25,55].

According to the detonate score, all assemblies were very similar except Velvet/Oases (which ranged from −3,716,349,411.54 to −3,787,528,395.51). The assemblies with the highest to lowest RSEM-EVAL scores were as follows: Trinity > Bridger > BinPacker > Velvet/Oases. Assemblies produced by Trinity and Velvet/Oases scored well with the highest (>96%) and lowest (<70%) percentage of RMBT (reads mapped back to transcriptome), respectively. RMBT results showed that BinPacker, Bridger, and Trinity performed similarly (ranging from 92% to 96%).

Based on the evaluation metrics performed, comprehensive comparison showed that BinPacker achieved the best results with a longer assembly size and N50, whereas the total number of contigs, RMBT percentage, and detonate score did not show a significant difference in comparison with Bridger (the assembler that produced the lowest number of contigs) and Trinity (the assembler that produced the highest RMBT percentage and detonate score). Hence, the BinPacker assembly was selected as the best for conducting downstream analyses.

3.3. Database Searches and Gene Ontology

We used the BinPacker assemblies for each species as queries and ran BLASTX to search against the UniProt database with no e-value cut-off, but an e-value of 1×10^{-5} as the threshold to report the results. A total of 39,968 (percentage of all unigenes in a species: 60.02%), 54,555 (61.33%), 53,922 (56.79%), 41,372 (61.58%), and 44,347 (65.39) unigenes of Tdae, Tvul, Tlan, Tper. and Tpub, respectively, were annotated. Overall, Tpub had a higher percentage of unigenes annotated compared to the other species. The number of predicted proteins ranged from 13,571 to 11,549. Tvul and Tlan had the highest numbers of predicted proteins, possibly because a greater size of sequencing data can provide a better assembly. Furthermore, predicted protein lists with significant sequence similarity were further used to compare the number of shared proteins among species (Figure 1). Five assemblies shared 6158 proteins. The highest number of protein sequence homology (21.5–24%) was shared between Tlan and Tvul (9447 proteins), while Tlan and Tpub shared the lowest number of proteins (8596 proteins), followed closely by Tdae and Tper (8597 proteins).
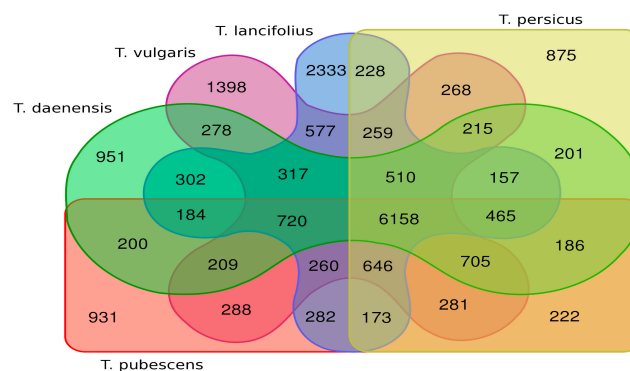
Figure 1. Venn diagram showing the number of overlapping proteins between the five *Thymus* transcriptome assemblies.

To categorize the biological functions of transcripts, the contigs having sequence homology with UniProt annotations were given GO (gene ontology) assignments under biological processes, cellular components, and molecular function domains. The distribution of functional categories for all five transcriptomes was visualized using the Web Gene Ontology Annotation Plotting (WEGO) application [32]. The derived graph (Figure 2) indicates that enriched gene ontology term annotations were broadly distributed across the three main domains, and the percentages of sequences mapped to

a given sub-ontology were highly similar for all species. Based on the obtained results, most of the GO terms were significantly overrepresented in cellular component domain. Within this category, cell class, followed by cell part and intracellular classes, were predominant. In the biological process domain, the two most common categories were cellular process and metabolic process. As far as molecular function, the most abundant classes were binding and catalytic activity.
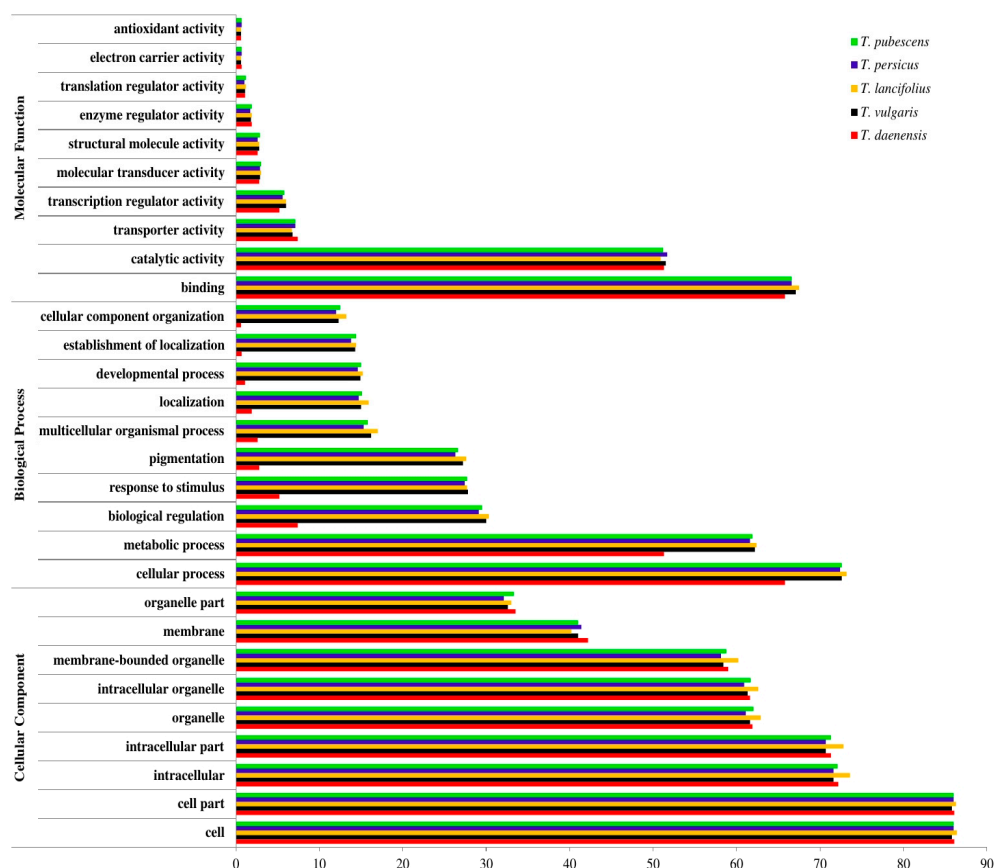

Figure 2. Gene ontology (GO) classification map derived from *Thymus* transcriptomes. The results are summarized in three GO categories: biological process, molecular function, and cellular component. The x-axis indicates the percentage of sequences in each category.

Overall, the GO term abundance results revealed that the main GO classifications were metabolism and fundamental biological regulation, and a majority of unigenes involved in terpenoid biosynthesis were found in the categories of metabolic processes. These results are similar to other terpene-producing plant species such as *Salvia miltiorrhiza* [61] and *Melaleuca alternifolia* [62].

According to functional annotations based on the COG databases, unigenes were clustered into 24 COG categories (Figure 3). The largest category was function unknown, followed by posttranslational modification, transcription, and signal transduction mechanisms. Nuclear structures and extracellular structures were the smallest groups. The number of assigned proteins to secondary metabolite biosynthesis ranged from 350 to 409. The presence of genes with unknown functions in these species suggests that the regulation of biosynthesis of secondary metabolites may involve novel factors that require further study.
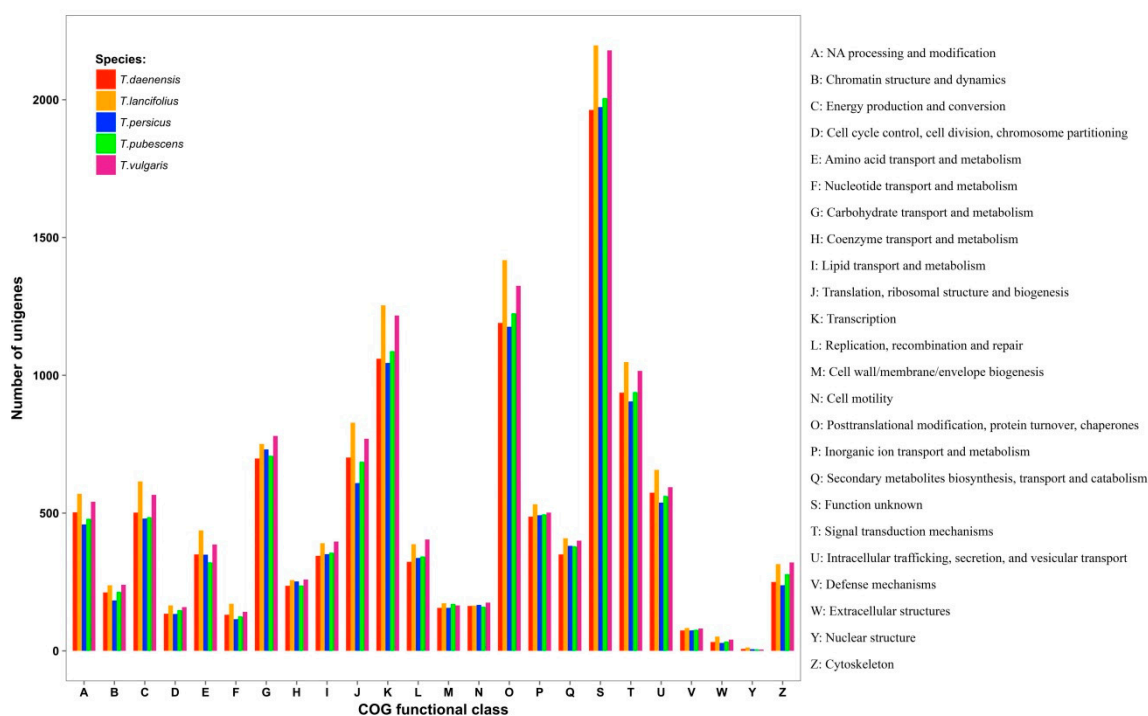
Figure 3. COG classification map. The abscissa represents 24 COG group names, while the vertical axis represents the number of unigenes annotated into each group.

Based on the KEGG assignment, the majority of predicted protein-coding genes were distributed under the "metabolism" classification (Figure S1). The genes in this category were further classified into 12 subcategories, of which "global and overview maps" were the most prevalent, followed by "carbohydrate metabolism", while the metabolism of terpenoids and polyketides categories contained a few gene members. This may be because the associated metabolites could be specific in each species. These results agree with transcriptome studies on other medicinal plants (e.g., *Cinnamomum camphora* [63] and *Murraya koenigii* [64]). It was also found that the number of genes in the secondary metabolic pathways related to diterpenoid biosynthesis (ko00904; ranging from 129 to 146 for Tvul and Tper) and terpenoid backbone biosynthesis (ko00900; ranging from 69 to 77 for Tlan, and Tper) were higher than monoterpenoid biosynthesis (ko00902; ranging from 22 to 31 for Tlan and Tvul) and sesquiterpenoid and triterpenoid biosynthesis (ko00909; ranging from 24 to 48 for Tdae and Tpub).

3.4. Identification of Transcription Factors

Terpenes are one of the most diverse and the largest known group of secondary plant metabolites [65,66], and are among the most important compounds in thyme [67,68]. Studies in Arabidopsis and transgenic roses have shown the importance of bHLH and MYB TFs in controlling terpenoid production [69,70]. Homologs of bHLH and MYC2 play roles in the regulation of the biosynthesis of sesquiterpenes in *Artemisia annua* [71], *Solanum lycopersicum* [72], and *Arabidopsis thaliana* [69]. TSAR1 and TSAR2 from the bHLH family activate all genes of the precursor mevalonate pathway [73]. MtTSAR and CrBIS TFs activate distinct terpenoid pathways in their species of origin [74]. Members of the WRKY family such as GaWRKY1 in *Gossypium arboreum* [75], SlWRKY73 in *Solanum lycopersicum* [71], OsWRKY76 in rice [76], and GrWRKY7 in *Gentiana rigescens* [77] have been shown to regulate sesquiterpene and terpene biosynthesis. In kiwifruit, the action of NAC TFs regulate monoterpene production [78]. The AP2/ERF family (e.g., CitERF71) has also been shown to regulate sesquiterpenes [79].

To identify TFs, *Thymus* datasets were annotated against all plant TFs. Unigenes encoding TFs were classified into 53 different TF families (Table S4). In our datasets, bHLH, NAC, MYB-related, ARF, WRKY, ERF, C2H2, C3H, FAR1, and B3 families constituted a large proportion of transcripts encoding

transcription factors in all samples, suggesting their potential function in regulating terpenoids biosynthesis (Figure 4). However, evaluation of expression changes and co-expression pattern analysis of these transcription factors are needed to further indicate the possible positive regulators of terpene metabolism and provide information for the possible genetic regulation network of terpenes in *Thymus*.
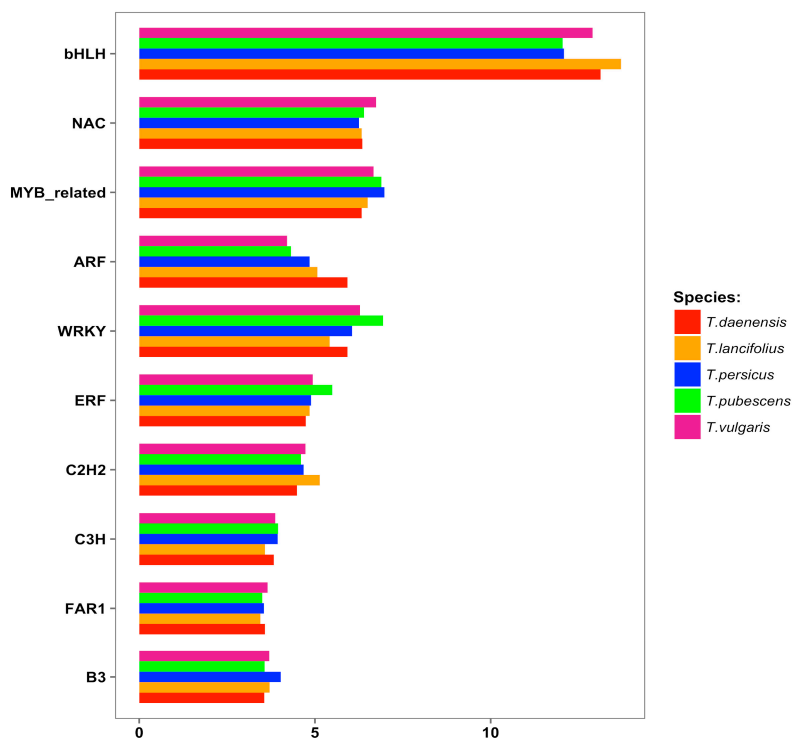
Figure 4. Distribution of transcripts in different transcription factor (TF) families showing over-representation in bHLH.

3.5. Thymol Biosynthetic Pathway

Triannotated transcriptome contigs were analyzed and mined for putative genes involved in thymol biosynthesis. Thymol formation is proposed to be through p-cymene and γ-terpinene intermediates [80,81]. Terpenes are derived from two five-carbon precursors, DMAPP (Dimethylallyl diphosphate) and IPP (Isopentenyl diphosphate), generated through the MVA (Mevalonate) and MEP (2-C-methylerythritol-4-phosphate) pathways, respectively. IPP and DMAPP are subsequently converted to GPP (Geranyl pyrophosphate), GGPP (Geranyl geranyl pyrophosphate) and FPP (Farnesyl pyrophosphate). GGDS occurs in both homomeric and heteromeric forms. The homomeric form is composed of a large subunit (LSU) and the heteromer form consists of an LSY and a small subunit (SSU). Monoterpenes are produced from GPP. Genes associated with terminal steps leading to thymol biosynthesis are GGDS, γ-terpinene synthase (GTP) and cytochrome P 450 71 family (Figure S2).

Based on the integrated annotation process, 81, 91, 80, 97, and 94 transcripts were annotated as 11 genes involved in thymol and other monoterpene biosyntheses (Table S5) in Tdae, Tvul, Tlan, Tper, and Tpub, respectively. For *DXS* (1-deoxy-D-xylulose 6-phosphate (DOXP) synthase), *HDS* (4-hydroxy-3-methyl-but-2-enyl diphosphate (HMB-PP) synthase), *HDR* (4-hydroxy-3-methyl-but-2-enyl diphosphate (HMB-PP) reductase), and *TPS*, there were high contig numbers (9 to 25 transcripts). *DXR* (1-deoxy-D-xylulose 6-phosphate reductoisomerase) and *IDI* (Isopentenyl diphosphate isomerase) had moderate numbers of contigs (3–10), while *MCT* (2-C-methyl-D-erythritol cytidyltransferase), *CMK* (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase), *MDS* (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase), and *GDS* (Geranyl diphosphate synthase) all had fewer than four contigs. It is reasonable to expect that the differences in contig number

reflects differences in transcript abundance. Given this, the MEP pathway would seem to be the major contributor of monoterpene thymol production, given the higher numbers of contigs associated with this pathway.

Due to the high gene expression level of monoterpene synthase genes in leaves, a query-based search showed the presence of complete coding regions for most of these genes (Table S6). However, IDI, GDS, and some terpene synthases (TPSs) were found as partial sequences in some species. TPSs are highly diversified throughout the plant kingdom and many exhibit tissue-specific expression. Analysis of the several plant genomes indicates TPS genes are a mid-size family, with the gene number ranging from 18 to 152, although not all of them have activity [82]. To date, a number of TPSs including TPS1–TPS9 have been identified and characterized in *T. Vulgaris*, *T. caespititius*, and *T. albicans* [80,83–86]. Nine to 25 *Thymus* transcripts were identified as putative TPS genes in different species grouped into four TPS classes, including TPS2, TPS3, TPS4, and TPS5. Among the identified TPS genes, the maximum number of unigenes was identified for TPS5, followed by TPS3. TPS5 displays a responsibility for the generation of volatile monoterpenes because of a similar trend with the emission of monoterpenes [87,88]. Through the homology search in the current assembled transcriptomes, the full-length sequences were identified for TPS5 in all species, TPS2 in only Tdae, TPS3 in Tvul, Tlan, and Tper, and finally TPS4 in all species except Tper. Global alignment revealed a >89% similarity between TPS5 and the sequence reported in NCBI (KC461937.1) belonging to *T. vulgaris* with length of 1795 bp. The sequences of unigene candidates encoding terpenoid biosynthesis enzymes are listed in Additional File 1. All four putative TPS sequences displayed high similarity to known plant TPS genes based on a comparison at the amino acid level. The amino acid comparison showed the highest similarity (~93%) between TPS2 and TPS3, followed by TPS4 and TPS5 (~89%). In contrast with this, the sequence of TPS4 displayed low levels of similarity (max. 63%) to TPS2.

TPSs are divided into three subclasses based on their functional roles and product formation, namely, monoterpene synthases (Mono-TPSs), diterpene synthases (Di-TPSs), and sesquiterpene synthases (Sesqui-TPSs). Based on sequence homology, TPSs are classified into seven families, from TPSa to TPSg [66,82]. However, most monoterpene synthase genes, especially in *T. vulgaris*, belong to the Mono-TPSb subfamily [85,89]. Here, we used the TERZYME program [90] to classify TPSs. As anticipated, all putative TPSs identified in the *Thymus* transcriptomes were assigned to the Mono-TPSb subfamily, supporting the existing view that *Thymus* TPSs belong to a distinct clade. All the discovered TPSs were investigated for the presence of terpene synthaseconserved motifs. The deduced amino acids of all TPSs contained the DDXXD motif, which is a highly conserved motif in TPSs and plays a role in the complexation of the diphosphate group after the ionization of substrates [82,91]. To obtain further information, phylogenetic analysis was performed to distinguish the evolutionary relationships of terpene synthases. Phylogenetic analysis placed TPS2 and TPS3 in a clade (Figure 5), while TPS5 was determined in a different clade. In the tree, it was observed that all TPSs identified in this study clustered into the TPS-b subfamily, which consisted mainly of Mono-TPSs. Phylogenetic analysis shows that TPSb is closely related to TPSg, which is congruent with previous phylogenetic studies of the TPS gene family [81,91]

TPSs along with CYPs organize the core components of terpene biosynthetic pathways [92]. It has been reported that the terminal gene involved in thymol biosynthesis is a regio-, stereo-, -specific Cytochrome P 450, specifically a member of the CYP71D family [93]. One previous study suggests that TPS genes are predominantly found in combination with CYP71 clan genes in both eudicots and monocots, and that among the TPS/CYP pairings there is a significant correlation between the TPSb and CYP71D subfamilies [92]. In the *Thymus* transciptome data, at least 45 putative CYP450s transcripts were identified belonging to five families including CYP71D, CYP72A, CYP736A, CYP749A, and CYP82d, with the majority of CYP716B2 family members. Thus, from this data we can conclude that the CYP genes may be involved in thymol and carvacrol biosynthesis.
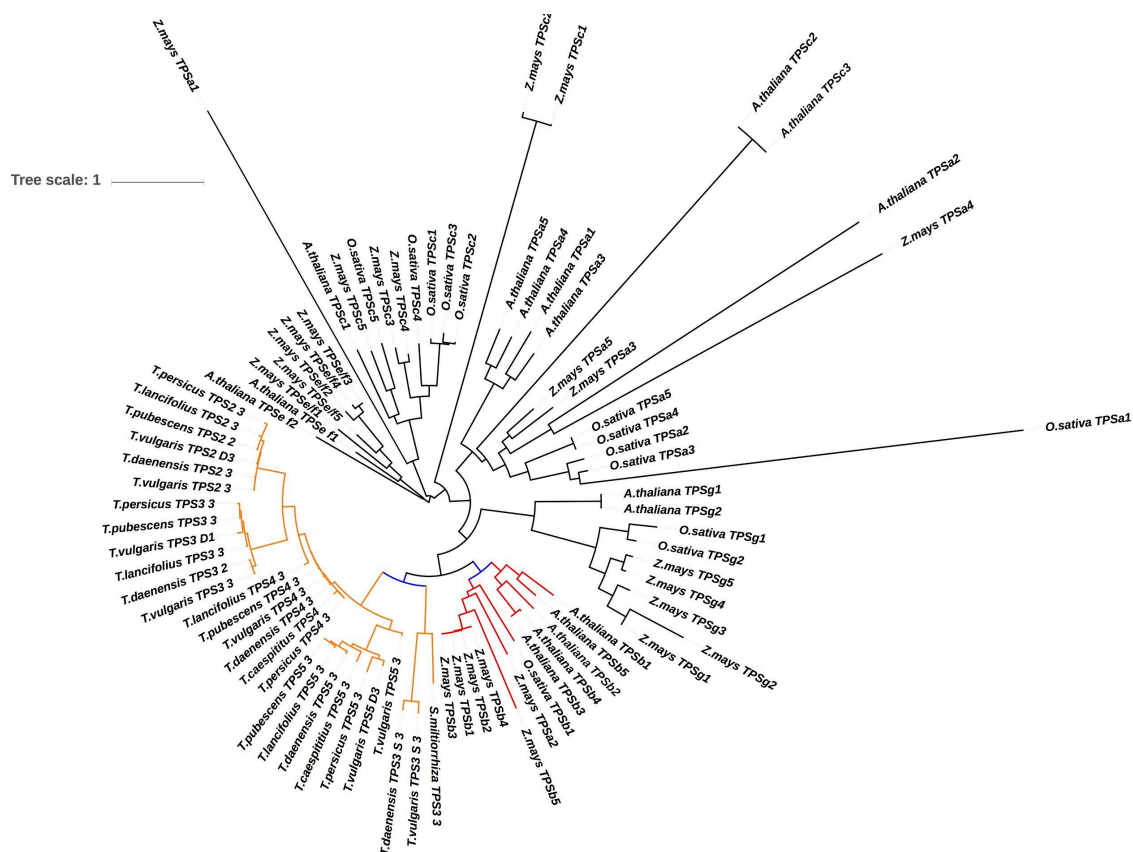
Figure 5. Maximum likelihood (ML) phylogeny of putative TPS genes in *Thymus* species compared to other species of TPS subfamilies.

3.6. Orthogroups Identified, Phylogenetic Analysis and Finding Genes under Positive Selection

Our understanding of the phylogenetic relationships of Lamiaceae has improved in recent years, with new phylogenetic studies placing Lamiaceae within a large clade called "core Lamiales" [94] that comprises more than 7000 species assigned to 236 genera [2]. However, the relationships among most genera and species have remained unresolved. In the present study, a large-scale phylogenetic reconstruction of *Thymus* species, which has been investigated so far mainly using small-scale molecular data [95], was carried out using 592 single-copy genes as follows.

An orthogroup is the set of genes that have evolved from a single gene of the last common ancestor of all the species via speciation [96]. To identify the orthogroups for the five *Thymus* species, with *S. splendens* and *M. spicata* as outgroups, protein sequences obtained from assembled and annotated transcriptome data were used with OrthoFinder [34]. The total number of orthogroups with all species present was 592. These were used for a subsequent phylogenetic analysis to define evolutionary relationships between these species.

Relationships found in our ML tree for the most part agree with the Bayesian tree (Figure 6). These phylogenetic trees strongly support a close genetic relationship between *T. daenensis* and *T. vulgaris* as sisters, as well as between this clade and a clade composed of *T. pubescens* and *T. lancifolius*.

Interestingly, the close relationship between *T. daenensis* and *T. vulgaris* is strongly supported by morphological and oil content similarity. *T. daenensis* is characterized by a high percentage of thymol, and *T. vulgaris* displays a mixture of thymol and g-terpinene as major constituents [97].

Divergence time estimation resulting from a Bayesian analysis suggests that *Thymus* originated during the late Pleistocene (2.24 MYA), then diverged and dispersed. This estimate falls into the range of ages estimated by Drew et al. (2012) [98]. Phylogenetic analysis of 121 accessions of the tribe

Mentheae based on cpDNA and nrDNA suggest that *Thymus* originated during the Mid-Pliocene (3.06 MYA, between 2.40 and 3.72 MYA) [98].

In a ML framework, divergence dates generally resulted in an increase in the estimated ages of younger nodes. The ML tree indicated that *Thymus* originated during the Tortonian (10.2 MYA). This difference is due to the methodology used for time calibration [99], molecular substitution [100], and clock model choice [101].
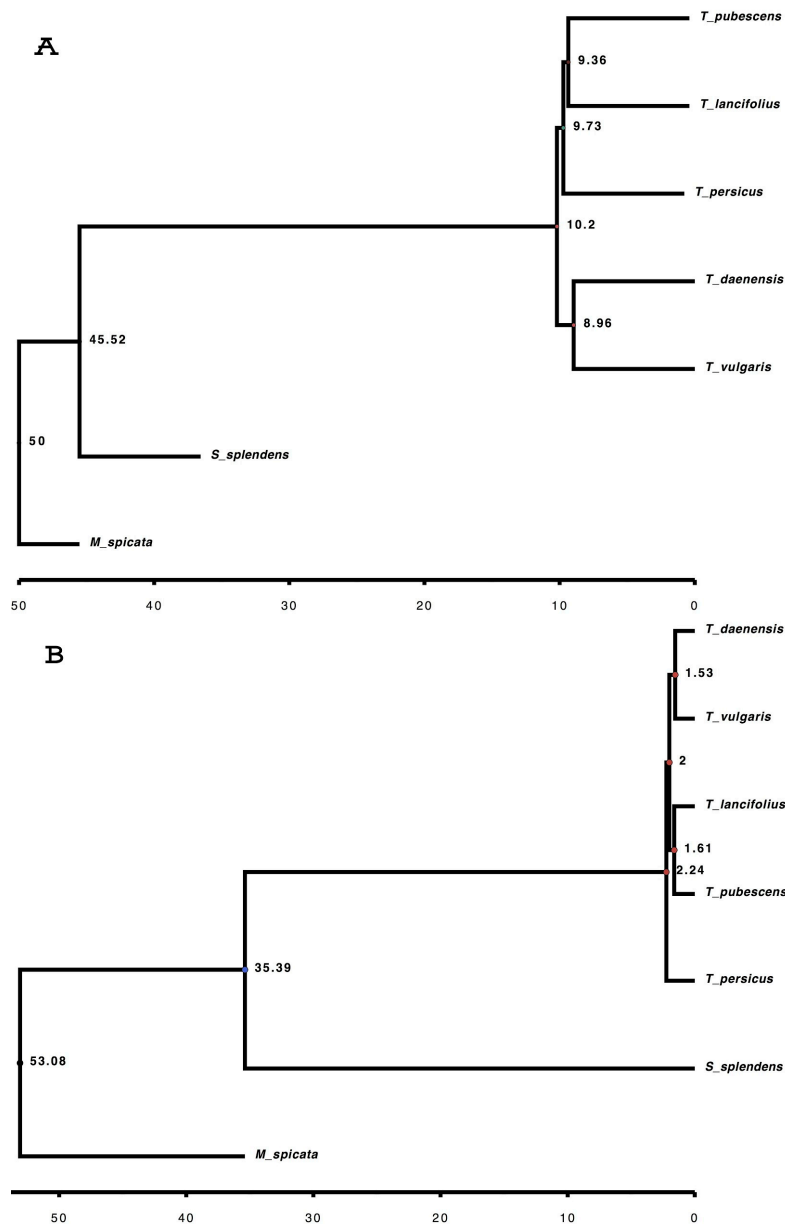
Figure 6. Phylogenetic relationships among five *Thymus* species inferred from 592 single-copy orthogroups. (**A**) Maximum likelihood tree. (**B**) Bayesian tree.

Orthogroups shared between *Thymus* species ranged from 18,412 between Tdae and Tper, and 21,071 between Tlan and Tvul. There were also 727 single-copy orthogroups in our species comparison that were retained for positive selection analysis. Using the branch-site model, 55, 49, 29, and 54 genes were identified as being under positive selection in Tdae, Tlan, Tper, and Tpub, respectively. Positively selected genes were further used to identify functional categories according to their assigned gene ontologies. The distribution of classification of PSGs showed that in all species (Figure S3), biological

process (GO:0008150) had the most hits. The second most abundant GO category was related to metabolic processes (GO: 0008152).

The unigenes subjected to positive selection were further annotated in the Swiss-Prot database (Table S7). Our search for gene matches from this study with other research revealed remarkable similarities. A protein involved in the pathway protein ubiquitination, namely F-box protein, was found to be common among all species. Indeed, it has been indicted that some F-box proteins are under selection and show a strong tendency of positive selection [102,103]. F-box genes show various roles in plants, including diverse abiotic and biotic responses and flowering and growth regulation [104]. Another group of positively selected genes, PPR (pentatricopeptide repeat), which has been found in all species, performs essential roles in plant embryogenesis and chloroplast RNA metabolism [105]. In many cases, PPR proteins are coded by restorer of fertility (Rf) genes that suppress cytoplasmic male sterility (CMS). The interaction of CMS genes and Rf genes leads to gynodioecy, which appears to be a stable sexual system in *Thymus* [106]. These Rf-PPR genes show a high dN/dS ratio in different species [107,108]. Our results also revealed that some genes belonging to the bHLH, MYB, and MADS-box TF families might also be positively selected and have roles in the diversification of these species.

3.7. Whole Genome Duplication Analysis

To identify WGD, first, the synonymous mutation rate (Ks) was estimated for each paralogous pair, then gene pairs showing abnormally high (>3) or low (<0.02) Ks values were discarded to make plots. As shown in Figure 7, the Ks values of paralogous genes showed an L-shaped distribution in all five *Thymus* transcriptomes, and no significant second peak was observed in the Ks-distributions. Overall, our result revealed that the small-scale gene duplications were dominant over long-term period (Ks < 3) and most of the paralogous genes were probably newly duplicated, which agrees well with the concept that most duplications are young. This small-scale duplication may also be associated with the genes related to secondary metabolites, and perhaps can explain the widespread monoterpene diversification [109,110]
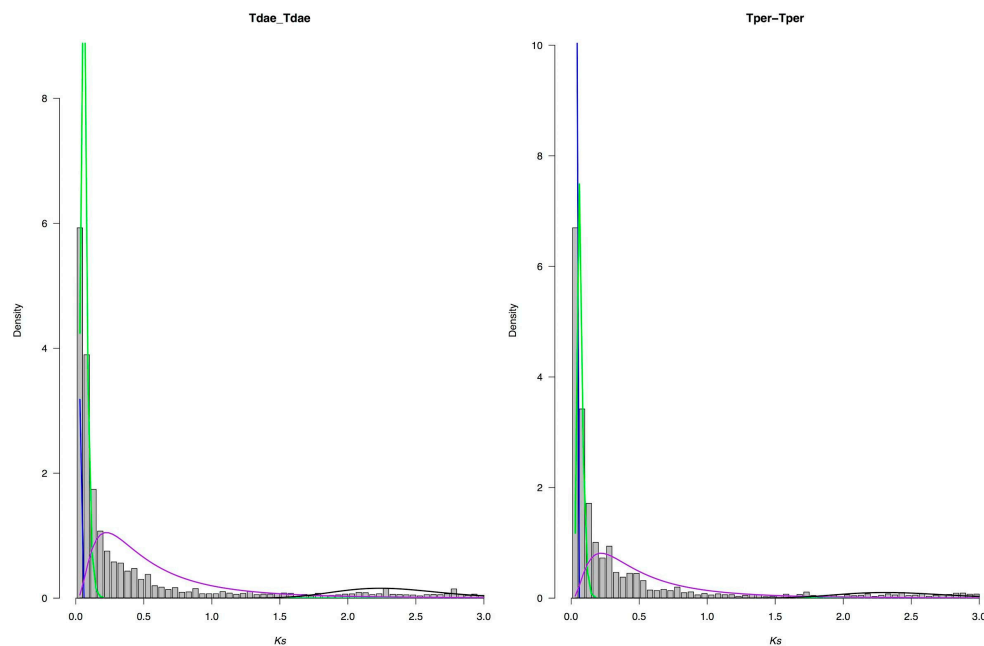
Figure 7. Ks distribution plots for *T. daenensis* and *T. persicus.* The components are (1) background gene duplications (green), (2) putative whole-genome duplications (purple), (3) alleles or sequencing errors resulting in high identity transcripts (blue), and (4) potential whole-genome duplications with low support (black).

4. Conclusions

In this study, we presented annotated transcriptomes for five different *Thymus* species using mRNA sequencing (RNA-seq), providing valuable resources for studying molecular mechanisms of terpenoid biosynthesis, the development of molecular markers, and evaluating phylogenetic relationships among species. This also led to the identification of putative genes involved in thymol biosynthesis. Most notably, three unigenes (germacrene D synthase, γ-terpinene synthase, and a member of the cytochrome P450 71 family) were identified as probable terminal steps of thymol biosynthesis. GO, COG, and KEGG enrichment analysis revealed that numerous contigs encode enzymes involved in secondary metabolic pathways. The functional characterization identified key TPS genes that were clustered into the TPS-b subfamily. The phylogenetic tree based on single-copy orthogroups showed a close genetic relationship between *T. daenensis* and *T. vulgaris*. Between 50 and 60 genes showed evidence of positive selection in the *Thymus* lineages. The present analysis based on L-shaped Ks distribution indicated that WGD is probably absent in *Thymus* species, or that WGD has probably not contributed substantially to duplicated genes in these species, and that polyploidy does not appear to be responsible for the large genome size in *T. lancifolius*, *T. persicus*, or *T. pubescens*. The study also provides worthy resources for a bioengineering study of terpenoids in *Thymus* species with pharmacologically increased metabolite content.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/10/8/620/s1, Figure S1: KEGG classification of unigenes for *Thymus* transcriptomes, Figure S2: Schematic representation of thymol and carvacrol biosynthetic pathway, Figure S3: GO enrichment of PSGs related to ecological adaptation in *Thymus* species, Table S1: Assembly statistics results of *Thymus* species, Table S2: Velvet/Oases assembly statistics results, Table S3: The number of full-length transcripts for three assemblers and five species, Table S4: Number of transcripts for each transcription factor genes families of *Thymus* species, Table S5: The contigs number of 11 genes involved in thymol and other monoterpene biosynthesis identified in *Thymus* transcriptomes, Table S6: Lengths of genes involved in thymol and other monoterpene biosynthesis identified in *Thymus* transcriptomes, Table S7: The unigenes subjected to positive selection.

Author Contributions: A.S. and A.B. designed the experiment. A.S., T.B. and A.A. collected the samples. A.S., A.G.H., and M.C. performed RNA extraction and prepared the libraries. A.S. carried out bioinformatics and comparative analyses. A.S. and A.B. wrote the manuscript. All authors reviewed and approved the final manuscript for publication.

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

References

1. World Checklist of Selected Plant Families. Available online: http://apps.kew.org/wcsp/incfamilies.do (accessed on 5 August 2018).
2. Harley, M.R.; Atkins, S.; Budantsev, L.A.; Cantino, P.D.; Conn, B.J.; Grayer, R.; Harley, M.M.; de Kok, R.; Krestovskaja, T.; Morales, R.; et al. *The Families and Genera of Vascular Plants*; Kubitzki, K., Kadereit, J.W., Eds.; Springer: Berlin, Germany, 2004; Volume 6, pp. 167–275.
3. Morales, R. The history, botany and taxonomy of the genus Thymus. In *Stahl-Biskup*; Saez, F., Ed.; Taylor & Francis: London, UK, 2002; pp. 1–43.
4. Lajayer, H.M.; Zakizadeh, H.; Hamidoghli, Y.; Bigluei, M.H.; Chamani, E. Ornamental potential and freezing tolerance of six Thymus spp. species as ground-covering plants in the landscape. *Zemdirb. Agric.* **2018**, *105*, 79–88. [CrossRef]
5. Leal, F.; Taghouti, M.; Nunes, F.; Silva, A.; Coelho, A.C.; Matos, M. Thymus Plants: A Review—Micropropagation, Molecular and Antifungal Activity. In *Active Ingredients from Aromatic and Medicinal Plants*; IntechOpen: Rijeka, Croatia, 2017.
6. Hazzit, M.; Baaliouamer, A.; Verissimo, A.R.; Faleiro, M.L.; Miguel, M.G. Chemical composition and biological activities of Algerian Thymus oils. *Food Chem.* **2009**, *116*, 714–721. [CrossRef]
7. Basch, E.; Ulbricht, C.; Hammerness, P.; Bevins, A.; Sollars, D. Thyme (*Thymus vulgaris* L.), thymol. *J. Herbal Pharma* **2004**, *4*, 49–67. [CrossRef]

8.	Thompson, J.D.; Chalchat, J.-C.; Michet, A.; Linhart, Y.B.; Ehlers, B. Qualitative and Quantitative Variation in Monoterpene Co-Occurrence and Composition in the Essential Oil of Thymus vulgaris Chemotypes. *J. Chem. Ecol.* **2003**, *29*, 859–880. [CrossRef] [PubMed]

9.	Karami-Osboo, R.; Miri, R.; Asadollahi, M.; Jassbi, A.R. Comparison Between Head-Space SPME and Hydrodistillation-GC-MS of the Volatiles of Thymus daenensis. *J. Essent. Oil Bear. Plants* **2015**, *18*, 925–930. [CrossRef]

10.	Sajjadi, S.E.; Khatamsaz, M. Composition of the Essential Oil of *Thymus daenensis* Celak. ssp. *lancifolius* (Celak.) Jalas. *J. Essent. Oil Res.* **2003**, *15*, 34–35. [CrossRef]

11.	Rasooli, I.; Mirmostafa, S.A. Antibacterial properties of *Thymus pubescens* and *Thymus serpyllum* essential oils. *Fitoterapia* **2002**, *73*, 244–250. [CrossRef]

12.	Hashemi-Moghaddam, H.; Mohammadhosseini, M.; Basiri, M. Optimization of Microwave Assisted Hydrodistillationon Chemical Compositions of the Essential Oils from the Aerial Parts of *Thymus pubescens* and Comparison with Conventional Hydrodistllation. *TEoP* **2015**, *18*, 884–893. [CrossRef]

13.	Rasooli, I.; Mirmostafa, S.A. Bacterial susceptibility to and chemical composition of essential oils from *Thymus kotschyanus* and *Thymus persicus*. *J. Agric. Food Chem.* **2003**, *51*, 2200–2205. [CrossRef]

14.	Thottathil, G.P.; Jayasekaran, K.; Othman, A.S. Sequencing Crop Genomes: A Gateway to Improve Tropical Agriculture. *Trop Life Sci. Res.* **2016**, *27*, 93–114.

15.	Pérez-de-Castro, A.M.; Vilanova, S.; Cañizares, J.; Pascual, L.; Blanca, J.M.; Díez, M.J.; Prohens, J.; Picó, B. Application of genomic tools in plant breeding. *Curr. Genom.* **2012**, *13*, 179–195. [CrossRef] [PubMed]

16.	Garg, R.; Jain, M. RNA-Seq for Transcriptome Analysis in Non-model Plants. *Adv. Struct. Saf. Stud.* **2013**, *1069*, 43–58. [CrossRef]

17.	He, B.; Zhao, S.; Chen, Y.; Cao, Q.; Wei, C.; Cheng, X.; Zhang, Y. Optimal assembly strategies of transcriptome related to ploidies of eukaryotic organisms. *BMC Genom.* **2015**, *16*, 65. [CrossRef] [PubMed]

18.	De Laat, A.M.M.; Blaas, J. Flow-cytometric characterization and sorting of plant chromosomes. *Theor. Appl. Genet.* **1984**, *67*, 463–467. [CrossRef] [PubMed]

19.	Dolezel, L.; Dolezel, J.; Bartos, J.; Voglmayr, H.; Greilhuber, J. Nuclear DNA content and genome size of trout and human. *Cytom. Part A* **2003**, *128*, 127–128. [CrossRef]

20.	Hunt, A.G. A rapid, simple and inexpensive method for the preparation of strand-specific RNA-Seq libraries. *Methods Mol. Biol.* **2015**, 195–207. [CrossRef]

21.	Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

22.	Grabherr, M.G. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol* **2011**, *29*, 644–652. [CrossRef]

23.	Zerbino, D.R.; Birney, E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [CrossRef]

24.	Schulz, M.H.; Zerbino, D.R.; Vingron, M.; Birney, E. Oases: Robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **2012**, *28*, 1086–1092. [CrossRef]

25.	Liu, J.; Li, G.; Chang, Z.; Yu, T.; Liu, B.; McMullen, R.; Chen, P.; Huang, X. BinPacker: Packing-Based *De Novo* Transcriptome Assembly from RNA-seq Data. *PLoS Comput. Biol.* **2016**, *12*, 1–15. [CrossRef]

26.	Chang, Z.; Li, G.; Liu, J.; Zhang, Y.; Ashby, C.; Liu, D.; Cramer, C.L.; Huang, X. Bridger: A new framework for *de novo* transcriptome assembly using RNA-seq data. *Genome Biol* **2015**, *16*, 1–10. [CrossRef] [PubMed]

27.	Li, B.; Fillmore, N.; Bai, Y.; Collins, M.; Thomson, J.; Stewart, R.; Dewey, C.N. Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Boil.* **2014**, *15*, 663. [CrossRef]

28.	Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, 29–37. [CrossRef] [PubMed]

29.	Punta, M.; Coggill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2011**, *40*, 290–301. [CrossRef] [PubMed]

30.	Petersen, T.N.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [CrossRef] [PubMed]

31.	Krogh, A.; Larsson, È.; Von Heijne, G.; Sonnhammer, E.L.L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [CrossRef] [PubMed]

32. Ye, J.; Fang, L.; Zheng, H.; Zhang, Y.; Chen, J.; Zhang, Z.; Wang, J.; Li, S.; Li, R.; Bolund, L.; et al. WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.* **2006**, *34*, 293–297. [CrossRef] [PubMed]

33. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggnog 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2016**, *44*, D286–D293. [CrossRef]

34. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 1–14. [CrossRef]

35. Ge, X.X.; Chen, H.W.; Wang, H.L.; Shi, A.P.; Liu, K.F. De novo assembly and annotation of Salvia splendens transcriptome using the illumina platform. *PLoS ONE* **2014**, *9*, e87693. [CrossRef] [PubMed]

36. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]

37. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef] [PubMed]

38. Stamatakis, A.; Ludwig, T.; Meier, H. RAxML: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **2005**, *21*, 456–463. [CrossRef] [PubMed]

39. Mirarab, S.; Warnow, T. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **2015**, *31*. [CrossRef] [PubMed]

40. Rambaut, A.; Drummond, A.J. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003537.

41. Rambaut, A.; Drummond, A.J. FigTree v1.4.0 in Computer Program and Documentation Distributed by the Author. 2012. Available online: http://tree.bio.ed.ac.uk/software/figtree/ (accessed on 5 August 2018).

42. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, 609–612. [CrossRef] [PubMed]

43. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]

44. Du, Z.; Zhou, X.; Ling, Y.; Zhang, Z.; Su, Z. agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **2010**, *38*, 64–70. [CrossRef]

45. McKain, M.R.; Tang, H.; McNeal, J.R.; Ayyampalayam, S.; Davis, J.I.; dePamphilis, C.W.; Givnish, T.J.; Pires, J.C.; Stevenson, D.W.; Leebens-Mack, J.H. A Phylogenomic Assessment of Ancient Polyploidy and Genome Evolution across the Poales. *Genome Biol. Evol.* **2016**, *8*, 1150–1164. [CrossRef]

46. Yang, Z.; Niele, R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* **2000**, *17*, 32–43. [CrossRef] [PubMed]

47. Goldman, N.; Yang, Z. A Codon-based Model of Nucleotide Substitution for Protein-coding DNA Sequences. *Mol. Bid. Evol.* **1980**, *11*, 725–736. [CrossRef]

48. Fraley, C.; Raftery, A.E.; Murphy, T.B.; Scrucca, L. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. *Tech. Rep.* **2012**, *597*, 1–50.

49. Marie, D.; Brown, S.C. A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol. Cell* **1993**, *78*, 41–51. [CrossRef]

50. Mahdavi, S.; Karimzadeh, G. Karyological and Nuclear DNA Content Variation in Some Iranian Endemic *Thymus Species* (Lamiaceae). *J. Agric. Sci. Technol* **2014**, *12*, 447–458.

51. Šmarda, P.; Bureš, P.; Horová, L.; Rotreklová, O. Intrapopulation Genome Size Dynamics in Festuca pallens. *Ann. Bot.* **2008**, *102*, 599–607. [CrossRef]

52. Jamzad, Z. New species and new plant records of lamiaceae from Iran. *Iran J. Bot.* **2009**, *15*, 51–56.

53. Ohri, D. Genome Size Variation and Plant Systematics. *Ann. Bot.* **1998**, *82*, 75–83. [CrossRef]

54. Moreton, J.; Dunham, S.P.; Emes, R.D. A consensus approach to vertebrate de novo transcriptome assembly from RNA-seq data: Assembly of the duck (Anas platyrhynchos) transcriptome. *Front. Genet.* **2014**, *5*, 1–6. [CrossRef]

55. Wang, S.; Gribskov, M. Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis. *Bioinformatics* **2016**, *33*, 327–333. [CrossRef]

56. Rana, S.B.; Zadlock, F.J.; Zhang, Z.; Murphy, W.R.; Bentivegna, C.S. Comparison of de Novo transcriptome assemblers and *k-mer* strategies using the killifish, *Fundulus heteroclitus*. *PLoS ONE* **2016**, *11*, e0153104. [CrossRef] [PubMed]

57. Smith-Unna, R.D.; Boursnell, C.; Patro, R.; Hibberd, J.M.; Kelly, S.; Street, D.; Brook, S.; Road, S.P. TransRate: Reference free quality assessment of de novo transcriptome assemblies. *Genome Res.* **2016**, *26*, 021626. [CrossRef] [PubMed]

58. Zeng, V.; Villanueva, K.E.; Ewen-Campen, B.S.; Alwes, F.; Browne, W.E.; Extavour, C.G. De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean *Parhyale hawaiensis*. *BMC Genom.* **2011**, *12*, 581. [CrossRef] [PubMed]

59. O'Neil, S.T.; Emrich, S.J. Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC Genom.* **2013**, *14*, 1–12. [CrossRef] [PubMed]

60. Chopra, R.; Burow, G.; Farmer, A.; Mudge, J.; Simpson, C.E.; Burow, M.D. Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-Seq data. *PLoS ONE* **2014**, *9*, 1–16. [CrossRef] [PubMed]

61. Zhou, W.; Huang, Q.; Wu, X.; Zhou, Z.; Ding, M.; Shi, M.; Huang, F.; Li, S.; Wang, Y.; Kai, G. Comprehensive transcriptome profiling of *Salvia miltiorrhiza* for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef] [PubMed]

62. Bustos-Segura, C.; Padovan, A.; Kainer, D.; Foley, W.J.; Külheim, C.; Bustos-Segura, C. Transcriptome analysis of terpene chemotypes of Melaleuca alternifolia across different tissues. *Plant Cell Environ.* **2017**, *40*, 2406–2425. [CrossRef] [PubMed]

63. Chen, C.; Zheng, Y.; Zhong, Y.; Wu, Y.; Li, Z.; Xu, L.A.; Xu, M. Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora*. *BMC Genom.* **2018**, *19*, 1–15. [CrossRef]

64. Meena, S.; Rajeev Kumar, S.; Dwivedi, V.; Kumar Singh, A.; Chanotiya, C.S.; Akhtar, M.Q.; Kumar, K.; Kumar Shasany, A.; Nagegowda, D.A. Transcriptomic insight into terpenoid and carbazole alkaloid biosynthesis, and functional characterization of two terpene synthases in curry tree (*Murraya koenigii*). *Sci. Rep.* **2017**, *7*, 1–14. [CrossRef]

65. Ashour, M.; Wink, M.; Gershenzon, J. Biochemistry of Terpenoids: Monoterpenes, Sesquiterpenes and Diterpenes. *Annu. Plant Rev.* **2010**, *40*, 258–303. [CrossRef]

66. Singh, B.; Sharma, R.A. Plant terpenes: Defense responses, phylogenetic analysis, regulation and clinical applications. *Biotech* **2015**, *5*, 129–151. [CrossRef] [PubMed]

67. Abousaber, M.; Khanavi, M.; Khoshchehreh, M.; Hadjiakhoondi, A.; Shams Ardekani, M.R. Composition of the Essential Oils of *Thymus deanensis* Celak Var. *deanensis* from Different Regions of Iran. *J. Med. Plants* **2012**, *11*, 34–39.

68. Tohidi, B.; Rahimmalek, M.; Arzani, A. Essential oil composition, total phenolic, flavonoid contents, and antioxidant activity of *Thymus* species collected from different regions of Iran. *Food Chem.* **2017**, *220*, 153–161. [CrossRef] [PubMed]

69. Hong, G.J.; Xue, X.Y.; Mao, Y.B.; Wang, L.J.; Chen, X.Y. *Arabidopsis* MYC2 Interacts with DELLA Proteins in Regulating Sesquiterpene Synthase Gene Expression. *Plant Cell* **2012**, *24*, 2635–2648. [CrossRef] [PubMed]

70. Zvi, M.M. B.; Shklarman, E.; Masci, T.; Kalev, H.; Debener, T.; Shafir, S.; Ovadis, M.; Vainstein, A. *PAP1* transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers. *New Phytol.* **2012**, *195*, 335–345. [CrossRef] [PubMed]

71. Ji, Y.; Xiao, J.; Shen, Y.; Ma, D.; Li, Z.; Pu, G.; Li, X.; Huang, L.; Liu, B.; Ye, H.; et al. Cloning and characterization of AabHLH1, a bHLH transcription factor that positively regulates artemisinin biosynthesis in *Artemisia annua*. *Plant Cell Physiol.* **2014**, *55*, 1592–1604. [CrossRef] [PubMed]

72. Spyropoulou, E.A.; Haring, M.A.; Schuurink, R.C. RNA sequencing on *Solanum lycopersicum* trichomes identifies transcription factors that activate terpene synthase promoters. *BMC Genom.* **2014**, *15*, 1–16. [CrossRef]

73. Mertens, J.; Pollier, J.; Vanden Bossche, R.; Lopez-Vidriero, I.; Franco-Zorrilla, J.M.; Goossens, A. The bHLH Transcription Factors TSAR1 and TSAR2 Regulate Triterpene Saponin Biosynthesis in *Medicago truncatula*. *Plant Physiol.* **2016**, *170*, 194–210. [CrossRef]

74. Mertens, J.; Van Moerkercke, A.; Bossche, R.V.; Pollier, J.; Goossens, A. Clade IVa basic helix-loop-helix transcription factors form part of a conserved jasmonate signaling circuit for the regulation of bioactive plant terpenoid biosynthesis. *Plant Cell Physiol.* **2016**, *57*, 2564–2575. [CrossRef]

75. Xu, Y.H. Characterization of GaWRKY1, a Cotton Transcription Factor That Regulates the Sesquiterpene Synthase Gene (+)–Cadinene Synthase-A. *Plant Physiol.* **2004**, *135*, 507–515. [CrossRef]

76. Yokotani, N.; Sato, Y.; Tanabe, S.; Chujo, T.; Shimizu, T.; Okada, K.; Yamane, H.; Shimono, M.; Sugano, S.; Takatsuji, H.; et al. WRKY76 is a rice transcriptional repressor playing opposite roles in blast disease resistance and cold stress tolerance. *J. Exp. Bot.* **2013**, *64*, 5085–5097. [CrossRef] [PubMed]

77. Zhang, X.; Allan, A.C.; Li, C.; Wang, Y.; Yao, Q. De Novo assembly and characterization of the transcriptome of the Chinese medicinal herb, Gentiana rigescens. *Int. J. Mol. Sci.* **2015**, *16*, 11550–11573. [CrossRef] [PubMed]

78. Nieuwenhuizen, N.J.; Chen, X.; Wang, M.Y.; Matich, A.J.; Perez, R.L.; Allan, A.C.; Green, S.A.; Atkinson, R.G. Natural Variation in Monoterpene Synthesis in Kiwifruit: Transcriptional Regulation of Terpene Synthases by NAC and ETHYLENE-INSENSITIVE3-Like Transcription Factors. *Plant Physiol.* **2015**, *167*, 1243–1258. [CrossRef] [PubMed]

79. Yu, Z.X.; Li, J.X.; Yang, C.Q.; Hu, W.L.; Wang, L.J.; Chen, X.Y. The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol. Plant* **2012**, *5*, 353–365. [CrossRef] [PubMed]

80. Rudolph, K.; Parthier, C.; Egerer-Sieber, C.; Geiger, D.; Muller, Y.A.; Kreis, W.; Müller-Uri, F. Expression, crystallization and structure elucidation of γ-terpinene synthase from Thymus vulgaris. *Acta Crystallogr. Sect. F Struct. Boil. Commun.* **2016**, *72*, 16–23. [CrossRef]

81. Trindade, H.; Pedro, L.G.; Figueiredo, A.C.; Barroso, J.G. Chemotypes and terpene synthase genes in Thymus genus: State of the art. *Ind. Crop Prod.* **2018**, *124*, 530–547. [CrossRef]

82. 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] [PubMed]

83. Mendes, M.D.D.S.; Barroso, J.G.; Oliveira, M.M.; Trindade, H. Identification and characterization of a second isogene encoding γ-terpinene synthase in Thymus caespititius. *J. Plant Physiol.* **2014**, *171*, 1017–1027. [CrossRef]

84. Lima, A.S.; Schimmel, J.; Lukas, B.; Novak, J.; Barroso, J.; Figueiredo, A.C.; Pedro, L.; Degenhardt, J.; Trindade, H. Genomic characterization, molecular cloning and expression analysis of two terpene synthases *from Thymus caespititius* (Lamiaceae). *Planta* **2013**, *238*, 191–204. [CrossRef]

85. Julia, A.; Christina, N. The Molecular Regulation of Terpene Polymorphism in *Thymus Vulgaris*. Ph.D. Thesis, Friedrich-Schiller-Universität, Jena, Germany, 2014.

86. Filipe, A.; Cardoso, J.C.; Miguel, G.; Anjos, L.; Trindade, H.; Figueiredo, A.C.; Barroso, J.; Power, D.M.; Marques, N.T. Molecular cloning and functional characterization of a monoterpene synthase isolated from the aromatic wild shrub Thymus albicans. *J. Plant Physiol.* **2017**, *218*, 35–44. [CrossRef]

87. He, S.; Wang, X.; Yang, S.; Dong, Y.; Zhao, Q.; Yang, J.; Cong, K.; Zhang, J.; Zhang, G.; Wang, Y.; et al. De novo Transcriptome Characterization of *Rhodomyrtus tomentosa* Leaves and Identification of Genes Involved in α/β-Pinene and β-Caryophyllene Biosynthesis. *Front Plant Sci.* **2018**, *9*, 1231. [CrossRef] [PubMed]

88. Yue, Y.; Yu, R.; Fan, Y. Transcriptome profiling provides new insights into the formation of floral scent in Hedychium coronarium. *BMC Genom.* **2015**, *16*, 470. [CrossRef] [PubMed]

89. Zapata, F.; Fine, P.V.A. Diversification of the monoterpene synthase gene family (TPSb) in Protium, a highly diverse genus of tropical trees. *Mol. Phylo Genet. Evol.* **2013**, *68*, 432–442. [CrossRef] [PubMed]

90. Priya, P.; Yadav, A.; Chand, J.; Yadav, G. Terzyme: A tool for identification and analysis of the plant terpenome. *Plant Methods* **2018**, *10*, 4. [CrossRef] [PubMed]

91. Degenhardt, J.; Kollner, T.G.; Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **2009**, *70*, 1621–1637. [CrossRef] [PubMed]

92. Boutanaev, A.M.; Moses, T.; Zi, J.; Nelson, D.R.; Mugford, S.T.; Peters, R.J.; Osbourn, A. Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 81–88. [CrossRef] [PubMed]

93. Krause, S. Biosynthesis of Oxygenated Monoterpenes in Thyme, Melaleuca, and Eucalyptus Species. Ph.D. Dissertation, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany, 2016.

94. Schäferhoff, B.; Fleischmann, A.; Fischer, E.; Albach, D.C.; Borsch, T.; Heubl, G.; Müller, K.F. Towards resolving Lamiales relationships: Insights from rapidly evolving chloroplast sequences. *BMC Evol. Biol.* **2010**, *10*, 352. [CrossRef] [PubMed]

95. Federici, S.; Galimberti, A.; Bartolucci, F.; Bruni, I.; De mattia, F.; Cortis, P. DNA barcoding to analyse taxonomically complex groups in plants: The case of *Thymus* (Lamiaceae). *Bot. J. Linn. Soc.* **2013**, *171*, 687–699. [CrossRef]

96. Trachana, K.; Larsson, T.A.; Powell, S.; Chen, W.H.; Doerks, T.; Muller, J.; Bork, P. Orthology prediction methods: A quality assessment using curated protein families. *Bioessays* **2011**, *33*, 769–780. [CrossRef]

97. Rustaiee, AR.; Yavari, A.; Nazeri, V.; Shokrpour, M.; Sefidkon, F.; Rasouli, M. Genetic diversity and chemical polymorphism of some *Thymus* species. *Chem. Biodivers.* **2013**, *10*, 1088–1098. [CrossRef]

98. Drew, B.T.; Sytsma, K.J. Phylogenetics, biogeography, and staminal evolution in the tribe Mentheae (Lamiaceae). *Am. J. Bot.* **2012**, *99*, 933–953. [CrossRef]

99. Parham, J.F.; Donoghue, P.C.J.; Bell, C.J.; Calway, T.D.; Head, J.J.; Holroyd, P.A. Best practices for justifying fossil calibrations. *Syst. Biol.* **2012**, *61*, 346–359. [CrossRef] [PubMed]

100. Buckley, T.R.; Simon, C.; Chambers, K. Exploring among-site rate variation models in a maximum likelihood framework sing empirical data: Effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Syst. Biol.* **2001**, *50*, 67–86. [CrossRef] [PubMed]

101. Ho, S.Y.W.; Duchene, S. Molecular clock methods for estimating evolutionary rates and timescales. *Mol. Ecol.* **2014**, *23*, 5947–5965. [CrossRef] [PubMed]

102. Thomas, J.H. Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res.* **2006**, *16*, 1017–1030. [CrossRef] [PubMed]

103. Navarro-Quezada, A.; Schumann, N.; Quint, M. Plant F-Box protein evolution is determined by lineage-specific timing of major gene family expansion waves. *PLoS ONE* **2013**, *8*, e68672. [CrossRef] [PubMed]

104. Yang, X.; Kalluri, U.C.; Jawdy, S.; Gunter, L.E.; Yin, T.; Tschaplinski, T.J.; Weston, D.J.; Ranjan, P.; Tuskan, G.A. The F-Box Gene Family Is Expanded in Herbaceous Annual Plants Relative to Woody Perennial Plants1[W][OA]. *Plant Physiol.* **2008**, *148*, 1189–1200. [CrossRef]

105. Cushing, D.A.; Forsthoefel, N.R.; Gestaut, D.R.; Vernon, D.M. Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* **2005**, *221*, 424–436. [CrossRef]

106. Manicacci, D.; Atlan, A.; Rossello, J.A.E.; Couvet, D. Gynodioecy and reproductive trait variation in three *Thymus* species (Lamiaceae). *Int. J. Plant Sci.* **1998**, *159*, 948–957. [CrossRef]

107. Geddy, R.; Brown, G.G. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genom.* **2007**, *8*, 130. [CrossRef]

108. Fujii, S.; Bond, C.S.; Small, I.D. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1723–1728. [CrossRef] [PubMed]

109. Blanc, G.; Wolfe, K.H. Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. *Plant Cell* **2004**, *16*, 1667–1678. [CrossRef] [PubMed]

110. Mint Evolutionary Genomics Consortium. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* **2018**, *11*, 1084–1096. [CrossRef] [PubMed]