forests

MDPI

Article

# Transcriptome Analysis of Bael (*Aegle marmelos* (L.) Corr.) a Member of Family Rutaceae

**Prashant Kaushik [1] and Shashi Kumar [2,*]**

[1] Independent Researcher, 46022 Valencia, Spain; prashantumri@gmail.com
[2] International Center for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India
* Correspondence: skrhode@icgeb.res.in; Tel.: +91-11-26741358

Received: 13 June 2018; Accepted: 10 July 2018; Published: 26 July 2018; Corrected: 3 July 2024


**Abstract:** *Aegle marmelos* (L.) Corr. is a medicinally and horticulturally important tree member of the family Rutaceae. It is native to India, where it is also known as Bael. Despite its importance, the genomic resources of this plant are scarce. This study presented the first-ever report of expressed transcripts in the leaves of *Aegle marmelos*. A total of 133,616 contigs were assembled to 46,335 unigenes with minimum and maximum lengths of 201 bp and 14,853 bp, respectively. There were 7002 transcription factors and 94,479 simple sequence repeat (SSR) markers. The *A. marmelos* transcripts were also annotated based on information from other members of Rutaceae; namely *Citrus clementina* and *Citrus sinensis*. A total of 482 transcripts were annotated as cytochrome p450s (CYPs), and 314 transcripts were annotated as glucosyltransferases (GTs). In the *A. marmelos* leaves, the monoterpenoid biosynthesis pathway was predominant. This study provides an important genomic resource along with useful information about *A. marmelos*.

**Keywords:** *Aegle marmelos* (L.) Corr.; transcripts; transcriptome assembly; simple sequence repeats; transcription factors; cytochrome p450; glycotransferases; metabolic pathway

## 1. Introduction

*Aegle marmelos* (L.) Corr. (2 n = 18) or Bael is an underexploited member of family Rutaceae. Believed to be native to the Indian subcontinent, it is well distributed throughout the tropical and subtropical belts of southeast Asia [1,2]. Botanically, *A. marmelos* is a deciduous tree stretching up to 10 m in height that flowers during the months of May–June [3,4]. It is also commonly grown as a horticultural plant in India, and its fruits are processed as juice or candies, as well as eaten fresh. During the past few decades, a spike in its cultivation as a horticulture plant has been attributed to its medicinal properties, along with a hardy nature that allows it to be cultivated on marginal lands with acidic or alkaline soils [5,6].

The traditional medicine system of Ayurveda in India routinely uses every part of *A. marmelos* as a therapy for medical conditions [7,8]. The leaves are most easily accessible, and are therefore more regularly used for the treatments than any other plant part. *A. maremelos* leaves are used to treat jaundice and help in wound-healing when applied as a paste on a wound surface [9]. Moreover, *A. marmelos* leaf extracts have been proved to be a better cure for gastrointestinal and hematopoietic damage than its fruits [10]. The leaf extract of *A. marmelos* is used as a medication against a number of chronic diseases such as diabetes, pancreatic cancer, and arthritis [11–14]. All of these medicinal properties of *A. marmelos* leaves are attributed to various phytochemicals present in the leaves such as aeglin, rutin, γ-sitosterole, β-sitosterol, eugenol, marmesinin, glycoside, skimmianine, etc. Broadly, these phytochemicals can be divided into three main classes: alkaloids, phenylpropanoids, and terpenoids [15,16]. However, there is no genomic data-based information about the pathways

of these important metabolic compounds that are present in the *A. marmelos* leaf. The information regarding the biosynthetic pathways and the encoding enzymes present in the *A. marmelos* leaf will be highly useful for the functional genomics in *A. marmelos* via transgenics and metabolic engineering approaches. Furthermore, *A. marmelos* leaf extract is used for the green synthesis of gold and silver nanoparticles [17,18].

The sum total of all of the transcripts captured in the cell of an individual organism is called its transcriptome [19]. There are two ways to capture the expressed transcripts: either by microarray, which is limited to predefined sequences, or by performing RNA-Seq using second-generation sequencing technologies [20]. This kind of sequencing has revolutionized the understanding of non-model organisms, and has evolved as one of the first choices of methods to apply to gene discovery and the expression profiling of non-model organisms [21,22]. The availability of well-defined computational tools, along with a well-applied methodology, has further demonstrated the effectiveness of de novo transcriptome assemblies in organisms even without a reference genome [23,24].

Genomic resources in A. marmelos are scarce compared with other members of Rutaceae, such as *Citrus sinensis* (Sweet Orange) and *Citrus clementina* (Clementine), both having well-annotated genomes [25,26]. Moreover, the unavailability of molecular markers based on the genomic information has further decelerated the molecular breeding efforts in *A. marmelos*. Earlier, a diversity study was carried out using only 12 random amplification of polymorphic DNA (RAPDs) [1]. This limitation can be overcome by developing an appropriate resource of genomic information-based molecular markers using a next-generation sequencing (NGS)-based approach such as transcriptomics [20,22]. To the best of our knowledge, this is the first detailed report on the transcriptome of this medically important plant. Moreover, only six expressed sequence tag (ESTs) are available in the National Center for Biotechnology Information (NCBI) database (accessed on 25 May 2018) [27]. An investigation into the leaf transcriptome of *A. marmelos* can help answer key questions regarding various aspects related to genes and their gene function, via the pathways involved in the metabolic compound formation. Therefore, we used RNA sequencing followed by the de novo transcriptome assembly of *A. marmelos* leaves to identify the transcription factors, simple sequence repeats (SSRs), and transcripts related to important metabolic pathways in the leaves of *A. marmelos*. Also, the information regarding cytochrome P450s (CYPs) and glucosyltransferases (GTs) extant in the leaf of *A. marmelos* was also accomplished.

## 2. Materials and Methods

### 2.1. RNA Isolation and Sequencing

Young and tender leaves from three mature and healthy plants of *A. marmelos* variety "Kaghzi" (~five years old) were collected from the Government Garden Nursery (coordinates at 29°58′06.9″ N 76°52′50.8″ E) in Haryana, India. The sampled leaf tissues were stored in RNAlater (Life Technologies, Carlsbad, CA, USA) till further use. RNA was extracted with a TRIZOL reagent (Life Technologies Corporation, Carlsbad, CA, USA) based RNA extraction protocol for plant leaves [28,29]. The quality of the extracted RNA was checked on a 1% formaldehyde denaturing agarose gel, and further quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Montchanin, DE, USA). A pooled sample of RNA from three selected plants was used for a single cDNA library preparation. The library was prepared with a TruSeq RNA Library Prep Kit v2 from Illumina® (Illumina Inc., San Diego, CA, USA), and the library quantification was done using a Qubit Fluorometer (Qubit™ dsDNA HS Assay Kit, Life Technologies Corporation, Carlsbad, California, USA) and Agilent D1000 ScreenTape system (Agilent Technologies, Santa Clara, CA, USA). The library was further sequenced on the Illumina HiSeq 2500 (2 × 150 bp) platform (Illumina, Dedham, MA, USA).

### 2.2. De Novo Assembly and Identification Coding Sequences

The cleaned reads were assembled using Trinity software (version 2.4.0) and TransDecoder v. 3.0.1 (http://transdecoder.sourceforge.net/) [30] was used to identify candidate coding regions within the generated transcripts and look for the open reading frames (ORF) that were at least 100 amino acids long in order to decrease the chances of false positives.

### 2.3. Gene Function Annotation

The transcripts with ORFs were annotated with BLASTX (default parameters, *e*-value cut-off $10^{-5}$) by resemblance counter to NR (non-redundant protein sequences database of NCBI), protein family (Pfam), Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) [31], *e*-value cut-off of $1 \times 10^{-5}$, and cluster of orthologous groups (COG) (https://www.ncbi.nlm.nih.gov/COG/) [32]. The gVolantes server (https://gvolante.riken.jp) [33] was used for the assessment of the completeness of transcriptome assembly via BUSCO_v3 selecting the plant ortholog set. Only transcripts pertaining to plant species were extracted and used for gene ontology. Pfam annotation was done with Hmmerscan, while Blast2Go was used for Gene Ontology (GO) annotation [34,35]. KEGG orthologies were estimated using the KEGG Automated Annotation Server (KAAS) by means of single-directional best hit method (http://www.genome.jp/kegg/kaas/) [36].

### 2.4. Identification of Transcription Factors

Transcription factors families present in the leaves of A. marmelos were identified by searching coding sequences identified by TransDecoder against the plant transcription factor database (PlnTFDB) (http://plntfdb.bio.uni-potsdam.de/v3.0/downloads.php) [37] with an e-value cut-off of $< 1 \times 10^{-10}$.

### 2.5. Identification of Simple Sequence Repeats (SSRs)

The presence of SSRs was determined by using MIcroSAtelliteidentification tool v1.0 (MISA) (http://pgrc.ipk-gatersleben.de/misa/) [38]. Briefly, the transcripts were checked 10 times for monorepeats, six times for direpeats, and five times for tri/tetra/penta/hexarepeats.

## 3. Results

### 3.1. De Novo Assembly, Gene Prediction, and Functional Annotation

RNA-Seq targeting expressed coding sequences has been used successfully in many medicinal and non-model plant species that do not have a reference genome (e.g., *Prosopis cineraria* L. [39], *Andrographis paniculata* Burm.f. [40], *Phyllanthus emblica* L. [41], *Picrorhiza kurroa* [42], and *Azadirachta indica* Royle ex Benth. [43]). Moreover, being a tree, *A. marmelos* can have a large genome size, which further restricts genome sequencing efforts [44].

The pooled RNA sample of *A. marmelos* leaves with RIN values around 8.0 generated a total of 115.92 million paired reads of high quality (Phred score > 30). Trinity assembler was used for the assembly, and after trimming of adapters, there was a total of 133,616 contigs (only from reads of 200 bp and above in length) clustered into 46,345 unigenes (Table 1). The raw data that was obtained as a result of sequencing was submitted to NCBI BioProject (PRJNA433585). The assembly completeness report from gVolante estimated that the transcriptome assembly was 90.15% complete (Figure S1). We scrutinized for an open reading frame that was at least 100 amino acids long in order to decrease the chances of false positives during open reading frames (ORF) predictions. The annotated transcripts with ORFs are listed in Table 2. A total of 90,525 transcripts were annotated to GO terms (Table S1). The transcripts related to plant species were extracted and used for gene ontology (Table 2).

**Table 1.** Assembly statistics of the leaf transcriptome.

| ***A. marmelos* Assembly Statistics** | |
|---|---|
| Number of assembled sequences | 133,616 |
| Total length (nt) | 225,969,847 |
| Range of coding sequence length (nt) | 201–14,853 |
| Mean sequence length (nt) | 1691 |
| Median sequence length (nt) | 1395 |
| N50 sequence length (nt) | 2544 |
| L50 sequence count | 30,077 |
| Number of sequences > 1K (nt) | 81,171 (60.7%) |
| Number of sequences > 10K (nt) | 96 (0.1%) |
| Base composition (%) | A: 29.75 |
| | T: 29.75 |
| | G: 20.36 |
| | C: 20.14 |
| Guanine-Cytosine content (%) | 40.50 |

**Table 2.** Annotation summary of *A. marmelos* leaf transcripts. COG: cluster of orthologous groups, GO: gene ontology, KEGG: Kyoto Encyclopedia of Genes and Genomes, ORFs: open reading frames, Pfam: protein family.

| **Parameters** | **Numbers** |
|---|---|
| Total Transcripts | 133,616 |
| Total ORFs | 165,230 |
| Transcript BLASTX (Plant Species) | 126,101 |
| Transcript annotated with GO terms | 90,525 |
| Transcript BLASTP | 82,445 |
| Transcript BLASTX against Pfam | 67,451 |
| Transcript annotations against KEGG | 83,221 |
| Transcript annotated with COG | 81,636 |

### *3.2. GO Annotation*

In total, 600,642 Gene Ontology (GO) terms were mapped to the *A. marmelos* leaf contigs belonging to all the three possible classes, i.e., biological process (227,921 transcripts), cellular component (188,465 transcripts), and molecular function (184,25 transcripts) in the GO database (Figure 1). The breakdown of the proteins associated with the various biological process, cellular components, and molecular functions is illustrated in Figure 2. The "integral component of membrane" (GO: 0016021) associated with various cellular components, "transcription DNA-templated" (GO: 00006351) associated with biological processes and "ATP binding" (GO: 00005524) associated with molecular function were the most mapped terms in their respective categories (Figure 2.).

The GO terms primarily define three categories of functions: namely, the biological, cellular, and molecular functions for a gene product. This is achieved by associating a gene with their ontologies [45,46]. Earlier studies have pointed out a higher metabolic activity in the leaves of *A. marmelos*, which is because of the presence of phytochemicals such as alkaloids, flavonoids, and phenols [47,48]. We have identified a number of GO terms in the leaves of *A. marmelos*; this information could lead to the identification of important pathways of metabolic compounds in *A. marmelos* [49].

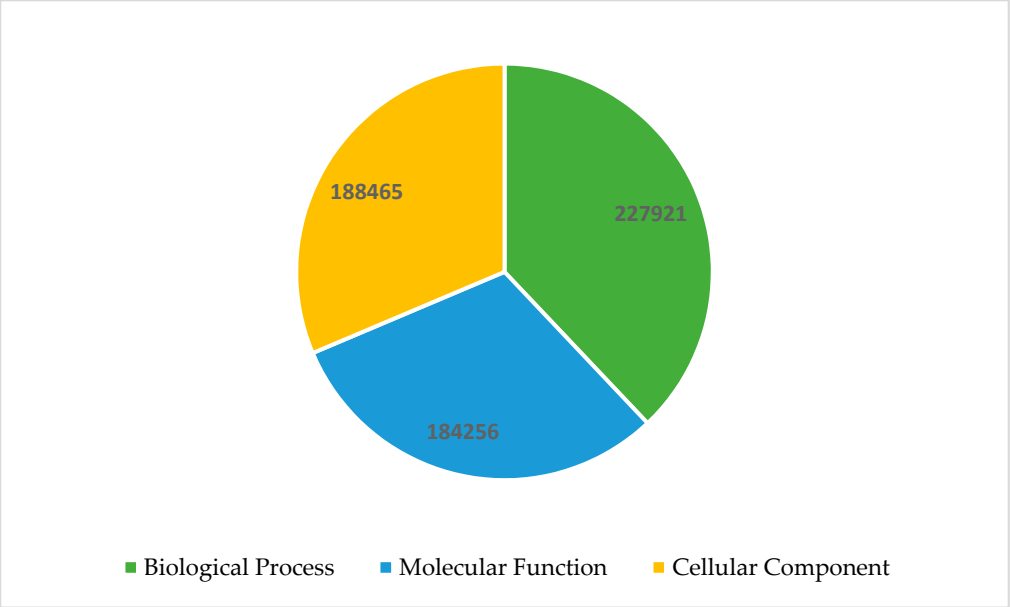


**Figure 1.** Genes associated with the biological process, cellular components, and molecular functions in the *A. marmelos* leaf transcriptome assembly.

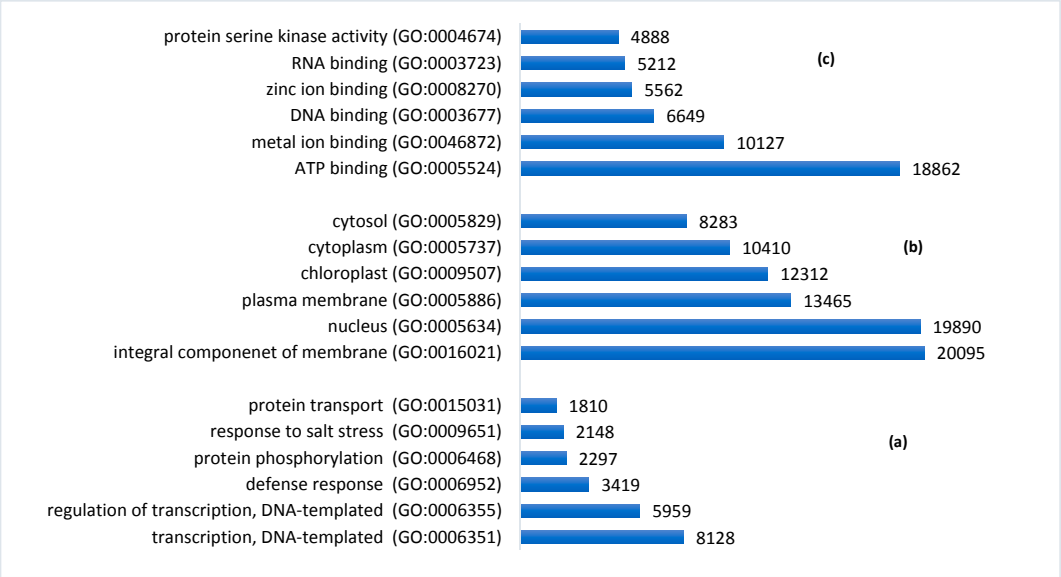


**Figure 2.** Gene Ontology (GO) classification of *A. marmelos* transcripts. GO term are divided in three main categories: biological process (**a**), cellular component (**b**), and molecular function (**c**).

### *3.3. Citrus Database Annotation*

The *A. marmelos* transcripts were also annotated via Phytozome (https://phytozome.jgi.doe.gov/) with reference to the *Citrus clementina* and *Citrus sinensis* genomes. This resulted in the mapping of 78.44% of the transcripts to the *Citrus clementine*, and 79.85% to the *Citrus sinensis* genome (Table 3). An almost similar number of transcripts were annotated with GO terms and KEGG annotation, respectively (Tables S2 and S3). However, recently, an extensive amount of relatedness was observed within the members of genus Citrus of family Rutaceae, which was based on the study performed by using the whole genome sequences of 60 members in the Citrus genus; the authors even pointed out the need for reformulation of the genus [50].

**Table 3.** Annotation summary of *A. marmelos* leaf transcripts with Citrus *sinensis* and *Citrus clementine* genome.

| Genome | *Citrus clementina* | *Citrus sinensis* |
|---|---|---|
| No. of transcripts annotated | 98,917 (78.44) | 100,701 (79.85) |
| No. of transcripts annotated by Gene Ontology | 98,184 | 100,053 |
| No. of transcripts annotated by KEGG | 48,184 | 48,881 |

### *3.4. Simple Sequence Repeats (SSRs) Prediction*

Simple sequence repeats (SSRs), or short tandem repeats or microsatellites, are short repeat motifs that show length polymorphism due to the insertion or deletion mutations of one or more repeat types [51]. We analyzed for the abundance of SSRs of annotated plant transcripts for *A. marmelos* leaf transcripts using the MISA tool, and the predicted SSRs statistics are shown in Figure 3. There were 58,354 transcripts that contained SSRs, and among these, 23,034 had more than one SSRs (Table S4). In total, 94,479 SSRs were identified, of which 65.27% were monorepeats, 19.78% were direpeats, and 13.40% were trirepeats (Figure 3). Tetra, penta and hexarepeats made up 1.01%, 0.26% and 0.24% of the total, respectively (Figure 3). However, out of a total of 94,479 identified SSRs, 11,400 (12.06%) were related to the compound formation.

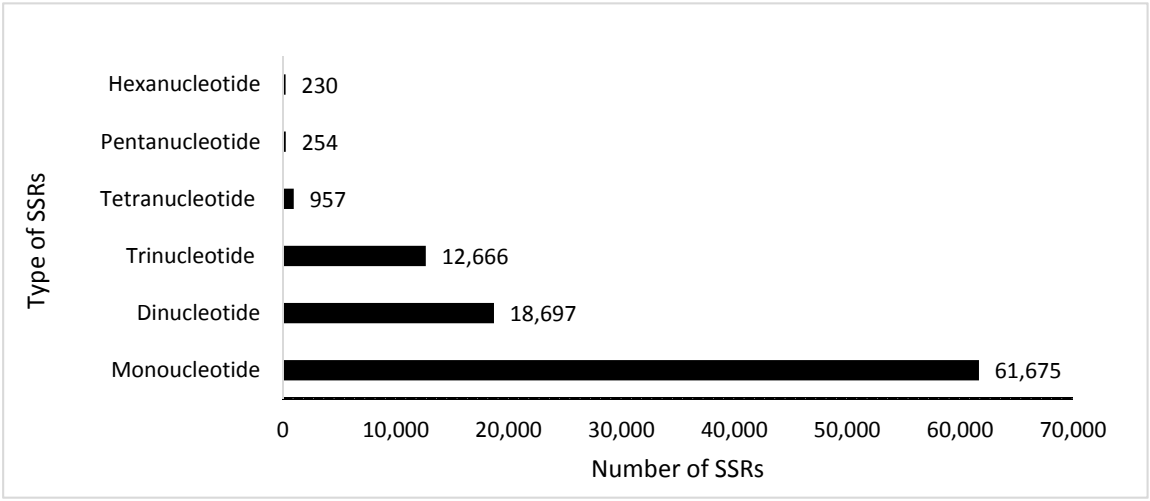


**Figure 3.** Simple sequence repeats (SSRs) classes identified in the leaf transcripts of *A. marmelos*.

SSRs are codominant markers that are well dispersed throughout plant genomes. SSRs are popularly used for marker-assisted selection, fingerprinting, diversity assessment, and quantitative trait loci (QTLs) identification [52]. Routinely, SSRs are identified in the medicinal plants via transcriptome assemblies, because they are more robust and can also be transferred among different species within the same genus. These identified SSRs can also be used for the marker-assisted breeding in *A. marmelos* i.e., to breed this tree for a particular environment or condition. Otherwise, until recently, only diversity-related studies were conducted in *A. marmelos* using universal primers, and researchers were even limited to only 12 RAPDs and 16 universal ISSRs to access diversity among their *A. marmelos* genotypes collection [1,53]. Furthermore, these genomic information-based SSRs can help to identify and differentiate between homozygous and heterozygous individuals. SSRs are also commonly used for the map-based cloning of genes; a close association between genes and their SSRs is crucial in the context of genotyping and haplotyping [51,52].

### *3.5. Transcriptional Factors Identification*

Gene expression patterns are regulated by transcription factors that in turn determine the different biological process [54]. A total of 7002 transcription factors were retrieved from the PlnTFDB.

The 52 transcription factors were unique to the *A. marmelos* leaves; although these were out of a total 6122 that were extant above 100 in the unigenes (Table S5). The most abundant were Auxin response factors (ARFs) (717), myeloblastosis (MYB-related) (562), a basic domain/leucine zipper (bZIP) (437), and basic helix–loop–helix (bHLH) (417), whereas HB-Other (132) and CAMATA (109) were the least abundant (Figure 4).

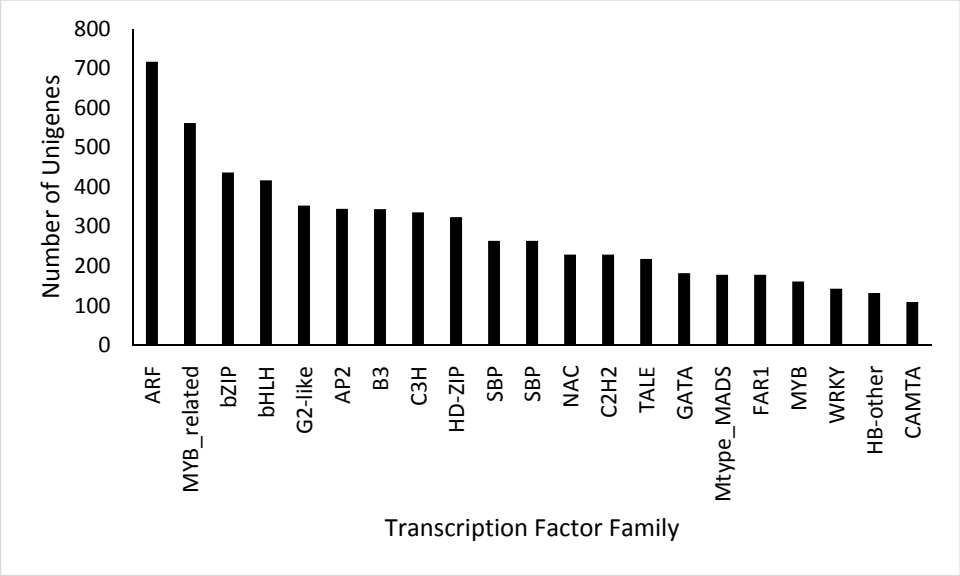


**Figure 4.** Top 21 families of transcription factors identified in the *A. marmelos* leaf.

Auxin is the plant hormone that regulates the different plant processes from growth to senescence. Auxin response factors are necessary for the plant to response to auxin stimuli; they channelize the response via auxin response DNA elements that are present in the primary auxin response genes. ARFs switch the auxin response gene on and off via their transcriptional activation domain or transcriptional repression domain [55,56]. MYB-related transcription factors play many roles like protection against biotic and abiotic stresses. MYB transcription factors also regulate the metabolism of the phenylpropanoid pathway, and are well studied with respect to the regulation of primary and secondary metabolism in the plant [57,58]. Likewise, bZIP and bHLH transcription factors are also involved in the metabolic biosynthesis in plants, especially by activation of phenylpropanoid genes [59,60].

### *3.6. Transcripts Encoding Cytochrome p450s (CYPs) and Glucosyltransferase (GTs)*

CYPs help in the primary and secondary metabolism of plants by catalyzing monooxygenation reactions. These cytochromes assist in the diversification of metabolic pathways in plants. Currently, these are potential targets for metabolic engineering for the overproduction of metabolites of interest [61]. There were 477 transcripts in total that were annotated cytochrome p450s (Table S6). Considering their vital role inmetabolic pathways, we further analyzed the abundance of SSRs annotated within these cytochrome p450 transcripts (Table 4). Among the 128 identified SSRs, 85 were with monorepeats, seven were with direpeats, and 36 were with trirepeats (Table S7).

**Table 4.** Prediction of simple sequence repeats (SSRs) for the annotated transcripts with cytochrome p450s (CYPs) and glucosyltransferase (GTs).

| Description | (CYPs) | (GTs) |
|---|---|---|
| Total number of annotated transcripts encoding | 477 | 314 |
| Total size of annotated transcripts (bp) | 1,111,258 | 778,664 |
| Number of SSRs identified | 128 | 247 |
| Number of annotated transcripts containing SSR | 94 | 165 |
| Number of annotated transcripts containing more than 1 SSR | 30 | 64 |
| Number of SSRs present in the compound formation | 5 | 16 |

The last step in the production of plant secondary metabolites is glycosylation, which is carried out by glycotransferases (GTs) [62–64]. A total of 314 transcripts were annotated as glucosyltransferase (Table S8). We analyzed the abundance of SSRs that were present in these transcripts (Table S9), and among the 247 identified, 109 were monorepeats, 79 were direpeats, 58 were trirepeats, and only one was identified as a tetrarepeat (Table 4). The SSRs that were identified as using CYPs and GTs can be of immense potential for identifying genetic diversity among different *A. marmelos* accessions with divergent metabolic profiles.

### *3.7. Identification of Biosynthetic Pathways in A. Marmelos Leaf*

*A. marmelos* leaves are used for the treatment of several medical conditions in Ayurveda and Yunani medicine systems [65]. The transcripts with the highest fragments per kilobase per million mapped reads (FPKM) values were extracted from an annotation file along with the Kyoto Encyclopedia of Genes and Genomes (KEGG) ID and sorted from the RNA-Seq by Expectation Maximization (RSEM) file that was obtained from the assembly for transcript quantification. Using the KEGG ID, pathways were identified (Table S10). RSEM is commonly used to obtain information regarding transcript abundance from the RNA-Seq data of an organism, even without a reference genome [66]. The pathway analysis identified that monoterpenoid biosynthesis and thiamine pathways were the two most expressed pathways present in the *A. marmelos* leaves (Figure 5).

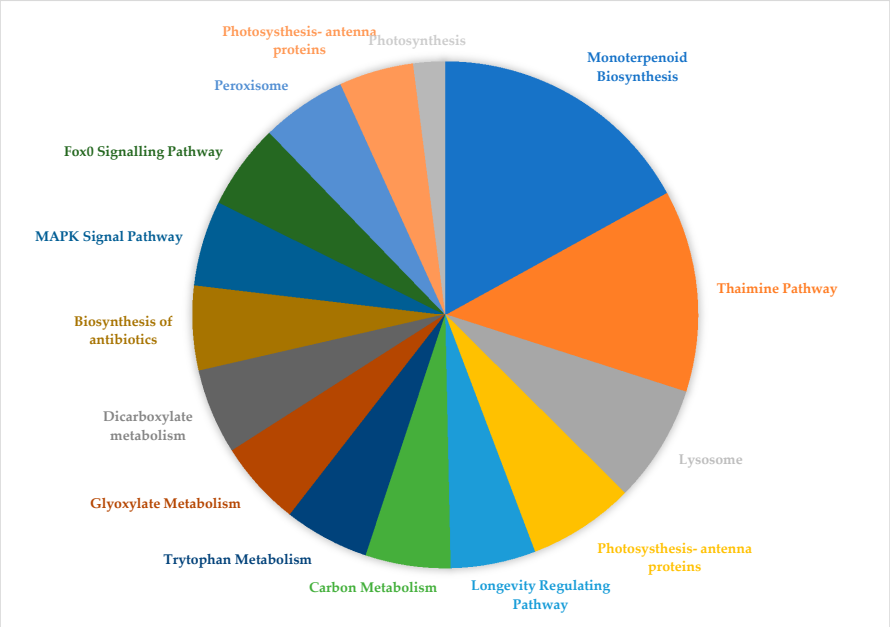


**Figure 5.** The top 15 pathways in the *A. marmelos* leaf.

*A. marmelos* leaves have been reported to contain monoterpenoids as the principal metabolites in the leaves in levels as high as 93.9% [67,68]. Moreover, this monoterpenoid content is not affected by the geographic location of the plant, as it remains unaffected by changes in altitude, unlike many other metabolites [69]. Thiamine is naturally produced in the plants as a sulfur-comprising and water-soluble compound. *A. marmelos* contains thiamine, although thiamine concentration is higher in the fruits than the leaves, and is among the fruits with the highest thiamine content [70,71]. Detailed information regarding the routes, reactions, and encoded enzymes of the two most expressed pathways identified in the leaf transcriptome of *A. marmelos* is provided in Figures 6 and 7. This detailed information generated regarding the monoterpenoid and thiamine biosynthetic pathway will be useful for further genetic analysis of the production of these important metabolites and their pathway engineering.

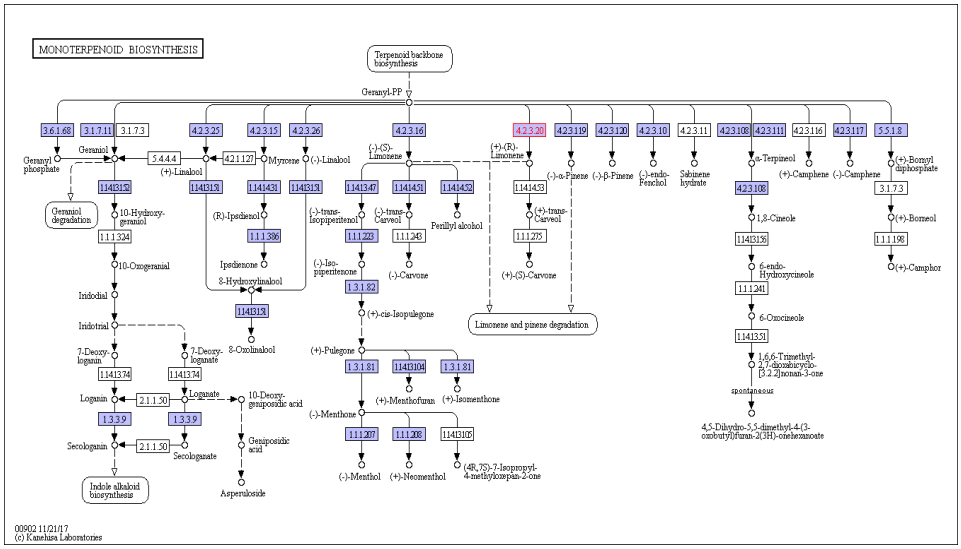


**Figure 6.** *A. marmelos* leaf transcriptome encoded enzymes (highlighted) involved in the monoterpenoid biosynthetic pathway identified in the leaf.

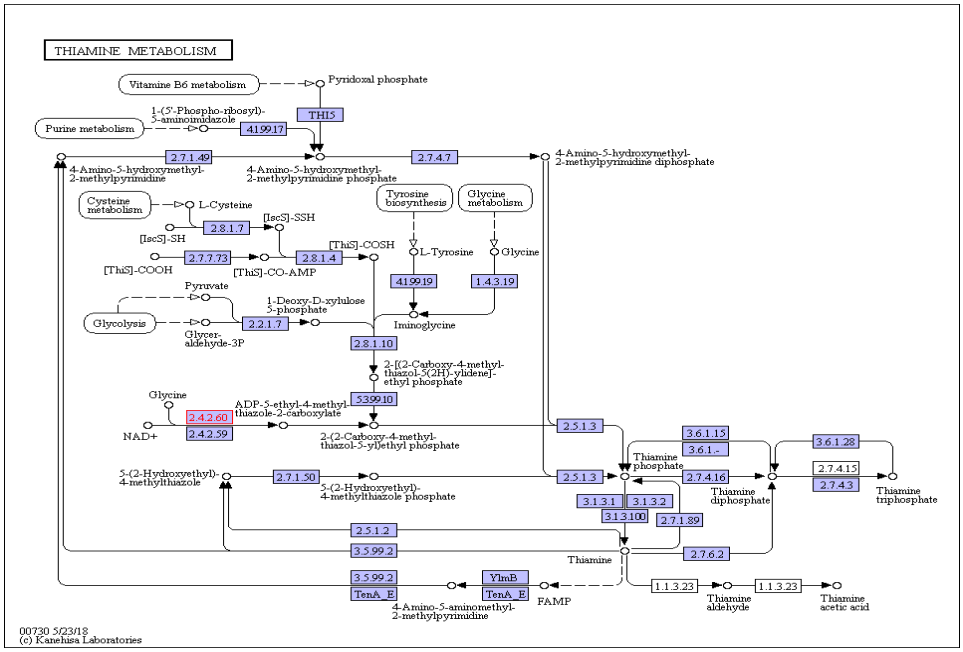


**Figure 7.** *A. marmelos* leaf transcriptome encoded enzymes (highlighted) involved in the thiamine biosynthetic pathway identified in the leaf.

## 4. Conclusions

In cases of underexploited plant species, there is often not enough genomic information available to proceed with their genetic improvement, and subsequently transfer important genes from them to cultivated crops. Transcriptome assembly is a cost-effective alternative to genome sequencing for obtaining the information of expressed genes and assisting in the more effective development of underexploited crops and medicinal plants. RNA-Seq shines a light on genes and their functions, as well as the pathways that are present, and can subsequently lead to evolutionary studies via molecular markers. We have successfully performed the first de novo transcriptome assembly of *A. marmelos*, which is a plant with religious, medicinal, and horticultural importance. It is the first-ever information about this plant, which will be of immense value for evolutionary studies and represents the development of a valuable resource for *A. marmelos*. Also, once a transcriptome reference is available, anchored-based transcriptome assemblies and different types of evolutionary studies can be performed within family Rutaceae involving genus *Aegle*.

**Supplementary Materials:** Supplementary materials can be found at http://www.mdpi.com/1999-4907/9/8/450/s1. Figure S1. Completeness assessment of *A. maremelos* leaf transcriptome assembly using gVolante server. Table S1. *A. marmelos* detailed annotation results. Table S2. Detail list of *A. maremlos* transcripts annotation with *Citrus sinensis* genome. Table S3. Detail result of *A. maremlos* transcripts annotation against *Citrus clementina* genome. Table S4. Detailed of SSRs in the *A. marmelos* leaf transcriptome. Table S5. Transcriptional factors family identified in *A. marmelos* leaf. Table S6. Detail list of cytochrome 450s (CYPs) annotation result. Table S7. SSRs identified in the cytochromes P450s transcripts. Table S8. Detail list of glycosyltransferases (GTs) annotation result. Table S9. SSRs identified in Glycosyltransferases (GTs) transcripts. Table S10. FPKM based top KEGG IDS in *A. maremlos* leaf transcriptome.

**Author Contributions:** P.K. and S.K. conceived and designed the project. P.K. performed the experiments. P.K. analyzed the data. P.K. and S.K. wrote the paper. Both authors read and approved the final manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** We thank the anonymous reviewers for their careful reading of the manuscript and providing the insightful suggestions. P.K. would like to thank Bengaluru Genomics Centre Pvt. Limited, Bengaluru, India for letting him use their facility.

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

## Abbreviations

| | |
|---|---|
| BLAST | Basic Local Alignment Search Tool |
| COG | Cluster of Orthologous Groups |
| CYP | Cytochrome p450 |
| FPKM | Fragments per kilobase of exon per million fragments mapped |
| GO | Gene Ontology |
| GT | Glucosyltransferase |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| KO | Kyoto Encyclopedia of Genes and Genomes Orthology |
| KOG | euKaryotic Ortholog Groups |
| MISA | MIcroSAtelliteidentification |
| NCBI | National Center of Biotechnology Information |
| NR | Non-redundant |
| Nt | Nucleotide |
| ORF | Open reading frame |
| Pfam | Protein family |
| PlnTFDB | Plant Transcription Factor DataBase |
| RNA-Seq | Ribonucleic acid sequencing |
| RSEM | RNA-Seq expression estimation by expectation maximization |
| SSR | Simple Sequence Repeat |

## References


1. Nayak, D.; Singh, D.R.; Sabarinathan, P.; Singh, S.; Nayak, T. Random amplified polymorphic DNA (RAPD) markers reveal genetic diversity in bael (*Aegle marmelos* Correa) genotypes of Andaman Islands, India. *Afr. J. Biotechnol.* **2013**, *12*. [CrossRef]
2. Ajithkumar, D.; Seeni, S. Rapid clonal multiplication through in vitro axillary shoot proliferation of *Aegle marmelos* (L.) Corr., a medicinal tree. *Plant Cell Rep.* **1998**, *17*, 422–426. [CrossRef]
3. Kishore, K.; Mahanti, K.K.; Samant, D. Phenological growth stages of bael (*Aegle marmelos*) according to the extended Biologische Bundesantalt, Bundessortenamt und Chemische Industrie scale. *Ann. App. Biol.* **2017**, *170*, 425–433. [CrossRef]
4. Singhal, V.; Salwan, A.; Kumar, P.; Kaur, J. Erratum to: Phenology, pollination and breeding system of *Aegle marmelos* (Linn.) Correa (Rutaceae) from India. *New For.* **2011**, *42*, 85–100. [CrossRef]
5. Hiwale, S. *Sustainable Horticulture in Semiarid Dry Lands*; Springer: New Delhi, India, 2015; ISBN 978-81-322-2243-9.
6. Khare, C.P. *Indian Medicinal Plants: An Illustrated Dictionary*; Springer-Verlag: New York, NY, USA, 2007; ISBN 978-0-387-70637-5.
7. Bhardwaj, R.L.; Nandal, U. Nutritional and therapeutic potential of bael (*Aegle marmelos* Corr.) fruit juice: A review. *Nutr. Food Sci.* **2015**, *45*, 895–919. [CrossRef]
8. Gutiérrez, S.P.; Sánchez, M.A.Z.; González, C.P.; García, L.A. Antidiarrhoeal activity of different plants used in traditional medicine. *Afr. J. Biotechnol.* **2007**, *6*, 2988–2994. [CrossRef]
9. Arul, V.; Miyazaki, S.; Dhananjayan, R. Studies on the anti-inflammatory, antipyretic and analgesic properties of the leaves of *Aegle marmelos* Corr. *J. Ethnopharmacol.* **2005**, *96*, 159–163. [CrossRef] [PubMed]
10. Singh, P.; Dutta, S.R.; Guha, D. Gastric mucosal protection by aegle marmelos against gastric mucosal damage: Role of enterochromaffin cell and serotonin. *Saudi J. Gastroenterol.* **2015**, *21*, 35. [CrossRef] [PubMed]
11. Manandhar, B.; Paudel, K.R.; Sharma, B.; Karki, R. Phytochemical profile and pharmacological activity of *Aegle marmelos* Linn. *J. Integr. Med.* **2018**, *16*, 153–163. [CrossRef] [PubMed]
12. Baliga, M.S.; Thilakchand, K.R.; Rai, M.P.; Rao, S.; Venkatesh, P. *Aegle marmelos* (L.) Correa (Bael) and its phytochemicals in the treatment and prevention of cancer. *Integr. Cancer Ther.* **2013**, *12*, 187–196. [CrossRef] [PubMed]
13. Shankarananth, V.; Balakrishnan, N.; Suresh, D.; Sureshpandian, G.; Edwin, E.; Sheeja, E. Analgesic activity of methanol extract of *Aegle marmelos* leaves. *Fitoterapia* **2007**, *78*, 258–259. [CrossRef] [PubMed]
14. Nigam, V.; Nambiar, V.S. *Aegle Marmelos* Leaf Juice As A Complementary Therapy To Control Type 2 Diabetes-Randomised Controlled Trial In Gujarat, India (in press). *Adv. Integr. Med.* **2018**. [CrossRef]
15. Laphookhieo, S.; Phungpanya, C.; Tantapakul, C.; Techa, S.; Tha-in, S.; Narmdorkmai, W. Chemical Constituents from *Aegle marmelos*. *J. Braz. Chem. Soc.* **2011**, *22*, 176–178. [CrossRef]
16. Maity, P.; Hansda, D.; Bandyopadhyay, U.; Mishra, D.K. Biological activities of crude extracts and chemical constituents of Bael, *Aegle marmelos* (L.) Corr. *Indian J. Exp. Biol.* **2009**, *47*, 849–861. [PubMed]
17. Jha, A.K.; Prasad, K. Biosynthesis of Gold Nanoparticles Using Bael (*Aegle marmelos*) Leaf: Mythology Meets Technology. *Int. J. Green Nanotechnol. Biomed.* **2011**, *3*, 92–97. [CrossRef]
18. Rao, K.J.; Paria, S. *Aegle marmelos* Leaf Extract and Plant Surfactants Mediated Green Synthesis of Au and Ag Nanoparticles by Optimizing Process Parameters Using Taguchi Method. *ACS Sustain. Chem. Eng.* **2015**, *3*, 483–491. [CrossRef]
19. Hancock, J.M.; Zvelebil, M.J. (Eds.) *Concise Encyclopaedia of Bioinformatics and Computational Biology*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2014; Available online: https://www.wiley.com/enus/Concise/Encyclopaedia/of/Bioinformatics/and/Computational+Biology%2C+2nd+Edition-p-9780470978719 (accessed on 6 May 2018).
20. Martin, J.A.; Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **2011**, *12*, 671–682. [CrossRef] [PubMed]
21. Papanicolaou, A.; Stierli, R.; ffrench-Constant, R.H.; Heckel, D.G. Next generation transcriptomes for next generation genomes using est2assembly. *BMC Bioinformatics* **2009**, *10*, 447. [CrossRef] [PubMed]
22. Morin, P.A.; Foote, A.D.; Hill, C.M.; Simon-Bouhet, B.; Lang, A.R.; Louis, M. SNP Discovery from Single and Multiplex Genome Assemblies of Non-model Organisms. Methods in Molecular Biology. In *Next Generation Sequencing*; Humana Press: New York, NY, USA, 2018; pp. 113–144. ISBN 978-1-4939-7512-9.

23. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]
24. Geniza, M.; Jaiswal, P. Tools for building de novo transcriptome assembly. *Curr. Plant Biol.* **2017**, *11*, 41–45. [CrossRef]
25. Xu, Q.; Chen, L.L.; Ruan, X.; Chen, D.; Zhu, A.; Chen, C.; Bertrand, D.; Jiao, W.B.; Hao, B.H.; Lyon, M.P.; et al. The draft genome of sweet orange (*Citrus sinensis*). *Nat. Genet.* **2013**, *45*, 59–66. [CrossRef] [PubMed]
26. Terol, J.; Naranjo, M.A.; Ollitrault, P.; Talon, M. Development of genomic resources for Citrus clementina: Characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. *BMC Genom.* **2008**, *9*, 423. [CrossRef] [PubMed]
27. National Center for Biotechnology Information (NCBI) Database. Available online: https://www.ncbi.nlm.nih.gov/nucest/?term=Aegle+marmelos (accessed on 25 May 2018).
28. Portillo, M.; Fenoll, C.; Escobar, C. Evaluation of different RNA extraction methods for small quantities of plant tissue: Combined effects of reagent type and homogenization procedure on RNA quality-integrity and yield. *Physiol. Plant.* **2006**, *128*, 1–7. [CrossRef]
29. Rio, D.C.; Ares, M.; Hannon, G.J.; Nilsen, T.W. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb. Protoc.* **2010**, *2010*, pdb–prot5439. [CrossRef] [PubMed]
30. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef] [PubMed]
31. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2011**, *40*, 109–114. [CrossRef] [PubMed]
32. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [CrossRef] [PubMed]
33. Nishimura, O.; Hara, Y.; Kuraku, S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* **2017**, *33*, 3635–3637. [CrossRef] [PubMed]
34. Conesa, A.; Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, *2008*, 619832. [CrossRef] [PubMed]
35. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef] [PubMed]
36. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **2007**, *35*, W182–W185. [CrossRef] [PubMed]
37. Pérez-Rodríguez, P.; Riaño-Pachón, D.; Corrêa, L.; Rensing, S.; Kersten, B.; Mueller-Roeber, B. PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* **2010**, D822–D827. [CrossRef] [PubMed]
38. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* **2013**, *106*, 411–422. [CrossRef] [PubMed]
39. Rai, M.K.; Shekhawat, J.K.; Kataria, V.; Shekhawat, N.S. De novo assembly of leaf transcriptome, functional annotation and genomic resources development in Prosopis cineraria, a multipurpose tree of Indian Thar Desert. *Plant Gene* **2017**, *12*, 88–97. [CrossRef]
40. Cherukupalli, N.; Divate, M.; Mittapelli, S.R.; Khareedu, V.R.; Vudem, D.R. De novo Assembly of Leaf Transcriptome in the Medicinal Plant Andrographis paniculata. *Front. Plant Sci.* **2016**, *7*, 1203. [CrossRef] [PubMed]
41. Kumar, A.; Kumar, S.; Bains, S.; Vaidya, V.; Singh, B.; Kaur, R.; Kaur, J.; Singh, K. De novo Transcriptome Analysis Revealed Genes Involved in Flavonoid and Vitamin C Biosynthesis in *Phyllanthus emblica* (L.). *Front. Plant Sci.* **2016**, *7*, 481. [CrossRef] [PubMed]
42. Kumar, V.; Sharma, N.; Sood, H.; Chauhan, R.S. Exogenous feeding of immediate precursors reveals synergistic effect on picroside-I biosynthesis in shoot cultures of Picrorhiza kurroa Royle ex Benth. *Sci. Rep.* **2016**, *6*, 29750. [CrossRef] [PubMed]
43. Krishnan, N.M.; Pattnaik, S.; Jain, P.; Gaur, P.; Choudhary, R.; Vaidyanathan, S.; Deepak, S.; Hariharan, A.K.; Krishna, P.B.; Nair, J.; et al. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm Azadirachta indica. *BMC Genom.* **2012**, *13*, 464. [CrossRef] [PubMed]

44. Schatz, M.C.; Witkowski, J.; McCombie, W.R. Current challenges in de novo plant genome sequencing and assembly. *Genome Biol.* **2012**, *13*, 243. [CrossRef] [PubMed]
45. Blake, J.A. Ten Quick Tips for Using the Gene Ontology. *PLOS Comput. Biol.* **2013**, *9*, e1003343. [CrossRef] [PubMed]
46. Rhee, S.Y.; Wood, V.; Dolinski, K.; Draghici, S. Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* **2008**, *9*, 509–515. [CrossRef] [PubMed]
47. Mujeeb, F.; Bajpai, P.; Pathak, N. Phytochemical Evaluation, Antimicrobial Activity, and Determination of Bioactive Components from Leaves of *Aegle marmelos*. *BioMed Res. Int.* **2014**, *2014*, 1–11. Available online: https://www.hindawi.com/journals/bmri/2014/497606/ (accessed on 12 May 2018). [CrossRef] [PubMed]
48. Pati, R.; Muthukumar, M. Genetic Transformation of Bael (*Aegle marmelos* Corr.). In *Biotechnology of Neglected and Underutilized Crops*; Springer: Dordrecht, The Netherlands, 2013; pp. 343–365. ISBN 978-94-007-5499-7.
49. Terol, J.; Tadeo, F.; Ventimilla, D.; Talon, M. An RNA-Seq-based reference transcriptome for Citrus. *Plant Biotechnol. J.* **2016**, *14*, 938–950. [CrossRef] [PubMed]
50. Wu, G.A.; Terol, J.; Ibanez, V.; López-García, A.; Pérez-Román, E.; Borredá, C.; Domingo, C.; Tadeo, F.R.; Carbonell-Caballero, J.; Alonso, R.; et al. Genomics of the origin and evolution of Citrus. *Nature* **2018**, *554*, 311–316. [CrossRef] [PubMed]
51. Nadeem, M.A.; Nawaz, M.A.; Shahid, M.Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N.; et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotec. Eq.* **2018**, *32*, 261–285. [CrossRef]
52. Vieira, M.L.C.; Santini, L.; Diniz, A.L.; de Freitas Munhoz, C. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **2016**, *39*, 312–328. [CrossRef] [PubMed]
53. Mujeeb, F.; Bajpai, P.; Pathak, N.; Verma, S.R. Genetic Diversity Analysis of Medicinally Important Horticultural Crop *Aegle marmelos* by ISSR Markers. *Methods Mol. Biol.* **2017**, *1620*, 195–211. [CrossRef] [PubMed]
54. Liu, L.; White, M.J.; MacRae, T.H. Transcription factors and their genes in higher plants. *Eur. J. Biochem.* **1999**, *262*, 247–257. [CrossRef] [PubMed]
55. Boer, D.R.; Freire-Rios, A.; van den Berg, W.A.M.; Saaki, T.; Manfield, I.W.; Kepinski, S.; López-Vidrieo, I.; Franco-Zorrilla, J.M.; de Vries, S.C.; Solano, R.; et al. Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell* **2014**, *156*, 577–589. [CrossRef] [PubMed]
56. Finet, C.; Berne-Dedieu, A.; Scutt, C.P.; Marlétaz, F. Evolution of the ARF gene family in land plants: Old domains, new tricks. *Mol. Biol. Evol.* **2013**, *30*, 45–56. [CrossRef] [PubMed]
57. Ambawat, S.; Sharma, P.; Yadav, N.R.; Yadav, R.C. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* **2013**, *19*, 307–321. [CrossRef] [PubMed]
58. Baldoni, E.; Genga, A.; Cominelli, E. Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms. *Int. J. Mol. Sci.* **2015**, *16*, 15811–15851. [CrossRef] [PubMed]
59. Hartmann, U.; Sagasser, M.; Mehrtens, F.; Stracke, R.; Weisshaar, B. Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. *Plant Mol. Biol.* **2005**, *57*, 155–171. [CrossRef] [PubMed]
60. Baxevanis, A.D.; Vinson, C.R. Interactions of coiled coils in transcription factors: Where is the specificity? *Curr. Opin. Genet. Dev.* **1993**, *3*, 278–285. [CrossRef]
61. Kumar, S. Engineering Cytochrome P450 Biocatalysts for Biotechnology, Medicine, and Bioremediation. *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*, 115–131. [CrossRef] [PubMed]
62. Chang, C.W.T. Predictable Enzymatic Glycosylation. *Chem. Biol.* **2009**, *16*, 579–580. [CrossRef] [PubMed]
63. Keegstra, K.; Raikhel, N. Plant glycosyltransferases. *Curr. Opin. Plant Biol.* **2001**, *4*, 219–224. [CrossRef]
64. Vogt, T.; Jones, P. Glycosyltransferases in plant natural product synthesis: Characterization of a supergene family. *Trends Plant Sci.* **2000**, *5*, 380–386. [CrossRef]
65. Rishabha, M.; Ajay, K.; Anupama, S.; Gt, K. Pharmacological Screening, Ayurvedic values and Commercial Utility of *Aegle Marmelos*. *Int. J. Drug Dev. Res.* **2012**, *4*, 28–37.
66. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [CrossRef] [PubMed]

67. Kaur, H.P.; Garg, S.N.; Sashidhara, K.V.; Yadav, A.; Naqvi, A.A.; Khanuja, S.P.S. Chemical Composition of the Essential Oil of the Twigs and Leaves of *Aegle marmelos* (L.) Correa. *J. Essent. Oil Res.* **2006**, *18*, 288–289. [CrossRef]
68. Raju, P.M.; Agarwal, S.S.; Ali, M.; Velasco-Negueruela, A.; Pérez-Alonso, M.J. Chemical Composition of the Leaf Oil of *Aegle marmelos* (L.) Correa. *J. Essent. Oil Res.* **1999**, *11*, 311–313. [CrossRef]
69. Verma, R.S.; Padalia, R.C.; Chauhan, A. Essential oil composition of *Aegle marmelos* (L.) Correa: Chemotypic and seasonal variations. *J. Sci. Food Agric.* **2013**, *94*, 1904–1913. [CrossRef] [PubMed]
70. Tarwadi, K.; Agte, V. Antioxidant and micronutrient potential of common fruits available in the Indian subcontinent. *Int. J. Food Sci. Nutr.* **2007**, *58*, 341–349. [CrossRef] [PubMed]
71. Baliga, M.S.; Bhat, H.P.; Joseph, N.; Fazal, F. Phytochemistry and medicinal uses of the bael fruit (*Aegle marmelos* Correa): A. concise review. *Food Res. Int.* **2011**, *44*, 1768–1775. [CrossRef]