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

Prashant Kaushik 1 and Shashi Kumar 2

1Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain
2 International Center for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India

**Abstract:** *Aegle marmelos* 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 the minimum and maximum lengths of 201 and 14,853 bp. 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 clementine* 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*; transcripts; transcriptome assembly; simple sequence repeats; transcription factors; cytochrome p450; glycotransferases; metabolic pathway

1. Introduction

*Aegle marmelos* (2n = 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 cultivation as a horticulture plant has been attributed to it has medicinal properties, along with a hardy nature 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 of medical conditions [7,8]. The leaves are most easily accessible and therefore regularly used for the treatments than any other plant part. *A. marmelos* 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].Leaf extract of *A. marmelos* is used as a medication against a number of chronic diseases like diabetes, pancreatic cancer, and arthritis [11–14]. All 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 viz., alkaloids, phenylpropanoids and terpenoid [15,16]. Furthermore, *A. marmelos* leaf extract is used for the green synthesis of gold and silver nanoparticles [17,18].

The sum-total of all 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 expression profiling of non-model organisms [21,22]. The availability of well-defined computational tools along with 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 as compared to other members of Rutaceae, like Citrus sinensis (Sweet Orange) and Citrus clementina (Clementine), both having well-annotated genomes [25,26]. To the best of our knowledge, this is the first detailed report on the transcriptome of this medically important plant. Moreover, only 6 ESTs are available in the National Center for Biotechnology Information (NCBI) database (Accessed on 25 May, 2018). An investigation into the leaf transcriptome of A. marmelos can help to 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 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” (~ 5 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, USA) till further use. RNA was extracted with a TRIZOL-based RNA extraction protocol for plant leaves [27,28]. 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, Montchain, 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., USA) and the library quantification was done using a Qubit Fluorometer (Qubit™ dsDNA HS Assay Kit) and Agilent D1000 ScreenTape system (Agilent Technologies, Santa Clara, CA, USA). The library was further sequenced on the Illumina HiSeq 2500 (2x150 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) [29] and TransDecoder v. 3.0.1(http://transdecoder.sourceforge.net/) was used to identify candidate coding regions within the generated transcripts and to look for open reading frames (ORF) of at least 100 amino acids long 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 (NCBI non-redundant protein sequences database of NCBI), protein family (Pfam), Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/), e-value cut-off of $1 \times 10^{-5}$ and Cluster of Orthologous Groups COG (https://www.ncbi.nlm.nih.gov/COG/) [30]. The gVolantes server (https://gvolante.riken.jp) 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 [31,32]. KEGG orthologies were estimated using KEGG Automated Annotation Server (KAAS) by means of single-directional best hit method (http://www.genome.jp/kegg/kaas/).

2.4. Identification of Transcription Factors

Transcription factors families present in the leaves of *A. marmelos* were identified by searching coding sequences identified by Trans Decoder against the plant transcription factor database (PlnTFDB) (http://plntfdb.bio.uni-potsdam.de/v3.0/downloads.php) 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/). Briefly, the transcripts were checked 10 times for mono-repeats, 6 times for di-repeats and 5 times for tri/tetra/penta/hexa-repeats.

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* [33], *Andrographis paniculate* [34], *Phyllanthus emblica* [35], *Picrorhiza kurroa* [36], *Azadirachta indica* [37]). Moreover, being a tree, *A. marmelos* can have a large genome size which further restricts genome sequencing efforts [38].

The pooled RNA sample of *A. marmelos* leaves of RIN value 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 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 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 percent complete. We scrutinized for an open reading frame of at least 100 amino acids long 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. 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)      | 1,691                                                |
| Median sequence length (nt)    | 1,395                                                |
| N50 sequence length (nt)       | 2,544                                                |
| 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                      |                                                       |
| GC-content (%)                 | 40.50                                                 |
Table 2. Annotation summary of *A. marmelos* leaf transcripts.

| 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 biological, cellular, and molecular functions for a gene product. This is achieved by associating a gene with their ontologies [39,40]. Earlier studies have pointed out a higher metabolic activity in the leaves of *A. marmelos* which is because of the presence of phytochemicals like alkaloids, flavonoids and phenols [41,42]. 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* [43].

![Figure 1. Genes associated with the biological process, cellular components, and molecular functions in the *A. marmelos* leaf transcriptome assembly.](image-url)
Figure 2. Gene Ontology (GO) classification of *A. marmelos* transcripts. GO terms are divided into 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 *Citrus clementina* and *Citrus sinensis* genome. This resulted in the mapping of 78.44% of the transcripts to the *Citrus clementina* 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. Though, recently an extensive amount of relatedness was observed within the members of genus Citrus of family Rutaceae 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 Citrus this could also rationalize our almost similar annotation results of *A. marmelos* transcripts with *Citrus clementina* and *Citrus sinensis* genomes [44].
Table 3. Annotation summary of *A. marmelos* leaf transcripts with *Citrus sinensis* and *Citrus clementina* 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 insertion or deletion mutations of one or more repeat types [45]. 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 which contained SSRs and among these, 23,034 had more than one SSRs. In total 94,479 SSRs were identified of which 65.27 % were mono repeats, 19.78 % were di repeats, and 13.40 % were tri repeats (Figure 3). Tetra, penta and hexa repeats were 1.01, 0.26 and 0.24 %, respectively (Figure 3). However, out of total 94,479 SSRs identified 11,400 (12.06 %) were related to the compound formation.

SSRs are codominant markers and are well dispersed throughout plant genomes. SSRs are popularly used for marker-assisted selection, fingerprinting, diversity assessment and QTL identification [46]. Routinely, because of massively parallel sequencing, a large number of 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 SSRs identified for *A. marmelos* can be used for the fingerprinting of different accessions otherwise until recently studies were conducted in *A. marmelos* using RAPDs and ISSRs which are general markers and are not based on previous genomic information [1,47].

![Figure 3](image-url)  
**Figure 3.** Simple sequence repeats (SSRs) classes identified in the leaf transcripts of *A. marmelos*.

3.5. Transcriptional factors identification

Gene expression patterns are regulated by transcription factors which in turn determine the different biological process[48]. A total of 7,002 transcription factors were retrieved PlantTFDB. The fifty-two transcription factors were unique to the *A. marmelos* leaves. Although out of total 6122 were extant above hundred in the unigenes. 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 least abundant (Figure 4).

Auxin is the plant hormone that regulates different plant process 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 present in the primary auxin
response genes. ARFs switch on and off the auxin response gene via their transcriptional
activation domain or transcriptional repression domain[49,50]. MYB-related transcription
factors play many roles in plants from plant defence including 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 [51,52]. Likewise, bZIP and bHLH transcription factors
are also involved in the metabolic biosynthesis in plants especially by activation of
phenylpropanoid genes [53,54].

![Figure 4.](image)

3.6. Transcripts encoding Cytochrome p450s (CYPs) and Glucosyltransferase (GTs)

CYPs help in primary and secondary metabolism of plants by catalysing
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 [55]. There were 477 transcripts in total that were
annotated a cytochrome p450s. Considering their vital role in metabolic pathway, we further
analyzed the abundance of SSRs annotated within these cytochrome p450 transcripts (Table
4). Among the 128 identified SSRs, 85 were with mono repeats, 7 were with di repeats, and
36 were with tri repeats.

The last step in the production of plant secondary metabolites is glycosylation which is
carried out by glucotransferases (GTs) [56–58]. A total of 314 transcripts were annotated as
glucosyltransferase. We analyzed the abundance of SSRs presence in these transcripts, and
among the 247 identified, 109 were mono repeats, 79 were di repeats, 58 were tri repeats and
only 1 was identified as a tetra repeats (Table 4). The SSRs identified using CYPs and GTs
can be of immense potential for identifying genetic diversity among different A. marmelos
accessions with divergent metabolic profiles.
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    |

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 [59]. The transcripts with the highest Fragments per Kilobase per Million mapped reads (FPKM) values were extracted from annotation file along with Kyoto encyclopedia of genes and genomes (KEGG) ID and sorted from the RNA-Seq by Expectation Maximization (RSEM) file that was obtained from assembly, for transcript quantification. Using the KEGG ID, pathways were identified. RSEM is commonly used to obtain information regarding transcript abundance from RNA-Seq data of an organism even without a reference genome [60]. The pathway analysis identified that monoterpenoid biosynthesis and thiamine pathways were the two most expressed pathways present in the *A. marmelos* leaves (Figure 5).

*A. marmelos* leaves have been reported to contain monoterpenoids as the principal metabolites in the leaves - as high as 93.9% [61,62]. Moreover, this monoterpenoid content is not affected by the geographic location of the plant as they remain unaffected by changes in altitude, unlike many other metabolites [63]. Thiamine is naturally produced in the plants as a sulphur comprising and water-soluble compound. *A. marmelos* contains thiamine although thiamine concentration is higher in fruits than leaves and is among the fruits with highest thiamine content [64,65].

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

4. Conclusions

In case of underexploited plant species, there is not enough genomic information available to proceed for their genetic improvement and to transfer important genes from them to the cultivated crops. Transcriptome assembly is a cost-effective alternative to genome sequencing for obtaining information of expressed genes and to assist in more effective
development of underexploited crops and medicinal plants. RNA-seq shines light on genes and their functions, pathways present, and can subsequently lead to evolutionary studies via molecular markers. We have successfully performed and generated data of the first de novo transcriptome assembly of *A. marmelos*, a plant with religious, medicinal and horticultural importance [66]. It is the first-ever information about this plant which will be of immense value for evolutionary studies and development of the valuable resource for *A. marmelos*. Also, once transcriptome reference is available, anchored based transcriptome assemblies and different type of evolutionary studies can be performed within family Rutaceae involving genus *Aegle*.

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

**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. *African Journal of Biotechnology* 2013, 12, doi:10.5897/AJB2013.12473.

2. Ajithkumar, D.; Seeni, S. Rapid clonal multiplication through in vitro axillary shoot proliferation of *Aegle marmelos* (L.) Corr., a medicinal tree. *Plant Cell Reports* 1998, 17, 422–426, doi:10.1007/s002990050418.

3. Kishore, K.; Mahanti, K. K.; Samant, D. Phenological growth stages of bael (*Aegle marmelos*) according to the extended Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie scale. *Annals of Applied Biology* 170, 425–433, doi:10.1111/aab.12347.

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 Forests* 2011, 42, 85–100, doi:10.1007/s11056-010-9239-3.

5. Hiwale, S. *Sustainable Horticulture in Semiarid Dry Lands*; Springer India, 2015; ISBN 978-81-322-2243-9.

6. Khare, C. P. *Indian Medicinal Plants: An Illustrated Dictionary*; Springer-Verlag: New York, 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. *Nutrition & Food Science* 2015, 45, 895–919, doi:10.1108/NFS-05-2015-0058.

8. Gutiérrez, S. P.; Sánchez, M. a. Z.; González, C. P.; García, L. A. Antidiarrheal activity of different plants used in traditional medicine. *African Journal of Biotechnology* 2007, 6, doi:10.5897/AJB2007.000-2465.

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, doi:10.1016/j.jep.2004.09.013.

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 Journal of Gastroenterology* 2015, 21, 35, doi:10.4103/1319-3767.151224.

11. Manandhar, B.; Paudel, K. R.; Sharma, B.; Karki, R. Phytochemical profile and pharmacological activity of *Aegle marmelos* Linn. *Journal of Integrative Medicine* 2018, 16, 153–163, doi:10.1016/j.jiom.2018.04.007.

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, doi:10.1177/1534735412451320.
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, doi:10.1016/j.fitote.2006.12.004.

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. *Advances in Integrative Medicine* **2018**, doi:10.1016/j.aimed.2018.03.002.

15. Laphookhieo, S.; Phungpanya, C.; Tantapakul, C.; Techr, S.; Tha-in, S.; Narmdorkmai, W. Chemical Constituents from Aegle marmelos. *Journal of the Brazilian Chemical Society* **2011**, 22, 176–178, doi:10.1590/S0103-50532011000100024.

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.

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, doi:10.1080/19430892.2011.574560.

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 Sustainable Chem. Eng.* **2015**, 3, 483–491, doi:10.1021/acssuschemeng.5b00022.

19. Concise Encyclopaedia of Bioinformatics and Computational Biology, 2nd Edition Available online: https://www.wiley.com/enus/Concise/Encyclopaedia/of/Bioinformatics/and/Computational+Biology%2C+2nd+Edition-p-9780470978719 (accessed on May 06, 2018).

20. Martin, J. A.; Wang, Z. Next-generation transcriptome assembly. *Nature Reviews Genetics* **2011**, 12, 671–682, doi:10.1038/nrg3068.

21. Morris, P. A.; Fooie, 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. In *Next Generation Sequencing*; Methods in Molecular Biology; Humana Press, New York, NY, 2018; pp. 113–144 ISBN 978-1-4939-7512-9.

22. Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, 29, 644–652, doi:10.1038/nbt.1883.

23. Geniza, M.; Jaiswal, P. Tools for building de novo transcriptome assembly. *Current Plant Biology* **2017**, 11–12, 41–45, doi:10.1016/j.cpb.2017.12.004.

24. Xu, Q.; Chen, L.-L.; Ruan, X.; Chen, D.; Zhu, A.; Chen, C.; Bertrand, D.; Jiao, W.-B.; Hao, B.-H.; Lyon, M. P.; Chen, J.; Gao, S.; Xing, F.; Lan, H.; Chang, J.-W.; Ge, X.; Lei, Y.; Hu, Q.; Miao, Y.; Wang, L.; Xiao, S.; Biswas, M. K.; Zeng, W.; Guo, F.; Cao, H.; Yang, X.; Xu, X.-W.; Cheng, Y.-J.; Xu, J.; Liu, J.-H.; Luo, O. J.; Tang, Z.; Guo, W.-W.; Kuang, H.; Zhang, H.-Y.; Roose, M. L.; Nagarajan, N.; Deng, X.-X.; Ruan, Y. The draft genome of sweet orange (*Citrus sinensis*). *Nature Genetics* **2013**, 45, 59–66, doi:10.1038/ng.2472.

25. 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 Genomics* **2008**, 9, 423, doi:10.1186/1471-2164-9-423.

27. 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. *Physiologia Plantarum* **128**, 1–7, doi:10.1111/j.1399-3054.2006.00716.x.

28. 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, doi:10.1101/pdb.prot5439.

29. Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **2013**, 8, 1494–1512, doi:10.1038/nprot.2013.084.

30. 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.

31. Conesa, A.; Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* **2008**, 2008, 619832, doi:10.1155/2008/619832.

32. Finn, R. D.; Clements, J.; Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, 39, W29–W37, doi:10.1093/nar/gkr367.

33. 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, doi:10.1016/j.plgene.2017.09.002.

34. 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, doi:10.3389/fpls.2016.01203.

35. 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, doi:10.3389/fpls.2016.01610.

36. 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, doi:10.1038/srep29750.

37. Krishnan, N. M.; Pattnaik, S.; Jain, P.; Gaur, P.; Choudhary, R.; Vaidyanathan, S.; Deepak, S.; Hariharan, A. K.; Krishna, P. B.; Nair, J.; Varghese, L.; Valivarthi, N. K.; Dhas, K.; Ramaswamy, K.; Panda, B. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm Azadirachta indica. *BMC Genomics* **2012**, 13, 464, doi:10.1186/1471-2164-13-464.

38. Schatz, M. C.; Witkowski, J.; McCombie, W. R. Current challenges in de novo plant genome sequencing and assembly. *Genome Biol.* **2012**, 13, 243, doi:10.1186/gb4015.

39. Blake, J. A. Ten Quick Tips for Using the Gene Ontology. *PLOS Computational Biology* **2013**, 9, e1003343, doi:10.1371/journal.pcbi.1003343.

40. Rhee, S. Y.; Wood, V.; Dolinski, K.; Draghici, S. Use and misuse of the gene ontology annotations. *Nature Reviews Genetics* **2008**, 9, 509–515, doi:10.1038/nrg2363.
41. Mujeeb, F.; Bajpai, P.; Pathak, N. Phytochemical Evaluation, Antimicrobial Activity, and Determination of Bioactive Components from Leaves of Aegle marmelos Available online: https://www.hindawi.com/journals/bmri/2014/497606/ (accessed on May 12, 2018).

42. Pati, R.; Muthukumar, M. Genetic Transformation of Bael (Aegle marmelos Corr.). In Biotechnology of Neglected and Underutilized Crops; Springer, Dordrecht, 2013; pp. 343–365 ISBN 978-94-007-5499-7.

43. Terol, J.; Tadeo, F.; Ventimilla, D.; Talon, M. An RNA-Seq-based reference transcriptome for Citrus. Plant Biotechnology Journal 2016, 14, 938–950, doi:10.1111/pbi.12447.

44. 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.; Curk, F.; Du, D.; Ollitrault, P.; Roose, F. G.; Rokhsar, D. S.; Talon, M. Genomics of the origin and evolution of Citrus. Nature 2018, 554, 311–316, doi:10.1038/nature25447.

45. Nadeem, M. A.; Nawaz, M. A.; Shahid, M. Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N.; Özkan, H.; Chung, G.; Baloch, F. S. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology & Biotechnological Equipment 2018, 32, 261–285, doi:10.1080/13102818.2017.1400401.

46. Vieira, M. L. C.; Santini, L.; Diniz, A. L.; Munhoz, C. de F. Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol 2016, 39, 312–328, doi:10.1590/1678-4685-GMB-2016-0027.

47. 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, doi:10.1007/978-1-4939-7060-5_14.

48. Liu, L.; White, M. J.; MacRae, T. H. Transcription factors and their genes in higher plants. European Journal of Biochemistry 262, 247–257, doi:10.1046/j.1432-1327.1999.00349.x.

49. 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.; Weijers, D.; Coll, M. Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. Cell 2014, 156, 577–589, doi:10.1016/j.cell.2013.12.027.

50. 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, doi:10.1093/molbev/mss220.

51. 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, doi:10.1007/s12298-013-0179-1.

52. Baldoni, E.; Genga, A.; Cominelli, E. Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms. Int J Mol Sci 2015, 16, 15811–15851, doi:10.3390/ijms160715811.

53. Hartmann, U.; Sagasser, M.; Mehtrens, 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, doi:10.1007/s11103-004-6910-0.
54. Baxevanis, A. D.; Vinson, C. R. Interactions of coiled coils in transcription factors: where is the specificity? *Current Opinion in Genetics & Development* **1993**, *3*, 278–285, doi:10.1016/0959-437X(93)90035-N.

55. Kumar, S. Engineering Cytochrome P450 Biocatalysts for Biotechnology, Medicine, and Bioremediation. *Expert Opin Drug Metab Toxicol* **2010**, *6*, 115–131, doi:10.1517/17425250903431040.

56. Chang, C.-W. T. Predictable Enzymatic Glycosylation. *Chemistry & Biology* **2009**, *16*, 579–580, doi:10.1016/j.chembiol.2009.06.001.

57. Keegstra, K.; Raikhel, N. Plant glycosyltransferases. *Curr Opin Plant Biol* **2001**, *4*, 219–224, doi:10.1016/S1369-5266(00)00164-3.

58. Vogt, T.; Jones, P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* **2000**, *5*, 380–386.

59. Rishabha, M.; Ajay, K.; Anupama, S.; Gt, K. Pharmacological Screening, Ayurvedic values and Commercial Utility of *Aegle Marmelos*. *International Journal of Drug Development and Research* **2012**, *4*, 28–37.

60. Li, B.; Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, *12*, 323, doi:10.1186/1471-2105-12-323.

61. 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. *Journal of Essential Oil Research* **2006**, *18*, 288–289, doi:10.1080/10412905.2006.9699091.

62. 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. *Journal of Essential Oil Research* **1999**, *11*, 311–313, doi:10.1080/10412905.1999.9701141.

63. Verma, R. S.; Padalia, R. C.; Chauhan, A. Essential oil composition of *Aegle marmelos* (L.) Correa: chemotypic and seasonal variations. *Journal of the Science of Food and Agriculture* **2013**, *94*, 1904–1913, doi:10.1002/jsfa.6510.

64. 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, doi:10.1080/09637480701243905.

65. 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 Research International* **2011**, *44*, 1768–1775, doi:10.1016/j.foodres.2011.02.008.

66. Kaushik, P.; Kumar, S. Data of de novo assembly of the leaf transcriptome in *Aegle marmelos*. *Data in Brief* **2018**, *19*, 700-703, doi:10.1016/j.dib.2018.05.095.