Agrotechniques in Industrial Crops

Journal Homepage: https://atic.razi.ac.ir

De Novo Transcriptome Assembly of *Calotropis procera* for Insights into the Natural Rubber Biosynthesis Pathway

Fatemeh Ali-Askari, Manijeh Sabokdast*, Mohammad Reza Naghavi

Department of Agronomy and Plant Breeding, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

ARTICLE INFO

Original paper

Article history:
Received: x Month 202x
Revised: x Month 202x
Accepted: x Month 202x

Keywords:
Gene annotation
KEGG pathway visualization
Latex
RNA-Seq analysis
Terpenoid

ABSTRACT

Calotropis procera has been widely researched as a medicinal plant mainly due to its medicinal properties, including the presence of cardiac glycosides; However, less attention has been paid to its study as a source for natural rubber production. The biosynthesis of natural rubber occurs through the mevalonate (MVA) pathway in the cytoplasm and the methylerythritol phosphate (MEP) pathway in the plastids, with acetyl-CoA being converted to isopentenyl pyrophosphate (IPP), the building block of natural rubber. Given the lack of detailed information on the natural rubber biosynthesis pathway in this plant and the absence of an annotated genome sequence, research based on *de novo* transcriptome assembly is particularly crucial. In this study, the Trinity and rnaSPAdes tools were employed to assemble the *de novo* transcriptome of *C. procera* from the data set available in the SRA database. Following a comprehensive evaluation of the assembly's quality, the transcriptome was annotated using the Hayai-Annotation Plants tool. Subsequently, pathways associated with natural rubber biosynthesis were reconstructed and mapped using the GhostKOALA and KEGG Mapper tools, with pathway visualizations created through the Color tool. According to the results, unigenes received putative annotations in three domains: biological processes, cellular components and molecular functions. Additionally, during the KEGG pathway reconstruction, 19 ortholog genes involved in the terpenoid backbone biosynthetic pathway were identified, including 7 ortholog genes related to the MVA pathway, 8 orthologs related to the MEP pathway, and 4 ortholog genes associated with the biosynthesis of C_{10}-C_{20} isoprenoids. This study enhances our understanding of the natural rubber biosynthesis pathways in *C. procera*, providing a foundation for future research aimed at exploring its biotechnological applications.

DOI: 10.22126/ATIC.2025.12119.1226

1. Introduction

The plant *Calotropis procera*, belonging to the Apocynaceae family, is a perennial, evergreen, and flowering species that naturally occurs in vast regions of Iran, including the provinces of Khuzestan, Bushehr, Hormozgan, Fars, Kerman, and Sistan and Baluchestan (Avazzadeh *et al*., 2018; Kaur *et al*., 2021). The use of this valuable medicinal plant in traditional medicine has been reported for the treatment of fever, leprosy, colds, eczema, asthma, rheumatism, dysentery, skin diseases and dyspepsia (Al-Rowaily *et al*., 2020). This plant is a source of numerous cardiac glycosides and their steroidal components (genins) (Pandey *et al*., 2016). Furthermore, all parts of the plant secrete a white milky latex from the incision sites (Falana and Nurudeen, 2020; Mohamed *et al*., 2017). The chemical composition of the plant's latex is highly complex, containing 25 to 35% poly (*cis*-1,4-isoprene) by weight, which is recognized as the principal compound of natural rubber (Mohamed *et al*., 2017).

Natural rubber, primarily derived from the rubber tree (*Hevea brasiliensis*), is renowned for its elasticity, excellent resistance to pressure, and abrasion. It finds widespread application across various industries, including tire manufacturing and medical equipment. The biosynthesis of this compound occurs through the mevalonate (MVA)-dependent pathway in the cytoplasm and the methylerythritol phosphate (MEP)

---

[*] Corresponding author.
E-mail address: sabokdast@ut.ac.ir

pathway in plastids. These pathways consist of a series of finely regulated enzymatic reactions that convert acetyl-CoA into isopentenyl pyrophosphate (IPP), which serves as the structural building block of natural rubber. (Pütter *et al*., 2017; Yamashita and Takahashi, 2020). The MVA pathway involves 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) as a critical regulatory enzyme, ultimately yielding IPP. Concurrently, Farnesyl Diphosphate Synthase (FPS) in the cytosol generates Farnesyl Pyrophosphate (FPP), which serves as an allylic primer for rubber polymerization. While necessary, its catalytic role is primarily in providing this precursor primer, with the complexity and regulation of the ultimate polymer synthesis residing elsewhere (Cherian *et al*., 2019; Yamashita and Takahashi, 2020).

The actual synthesis of the natural rubber polymer predominantly occurs on the surface of rubber particles, specialized cytoplasmic organelles characterized by a proteolipid unilamellar membrane. Here, *cis*-prenyltransferases (CPTs) catalyze the head-to-tail condensation of IPP molecules onto the FPP primer, elongating the polyisoprene chain with *cis* stereochemistry. Proteins abundant on the rubber particle surface, such as Rubber Elongation Factor (REF) and Small Rubber Particle Protein (SRPP), are intrinsically linked to this process. REF and SRPP form complexes that stabilize the rubber particle membrane, indicating their pivotal role in optimizing rubber biosynthesis beyond mere substrate provision (Fang *et al*., 2025; Yamashita and Takahashi, 2020).

Today, natural rubber faces an increasing global demand, necessitating the exploration of alternative sources. *Taraxacum kok-saghyz* offers a promising alternative due to its high-quality rubber, rapid growth, and adaptability. In this plant, transcriptome analysis identified rubber-related genes and Single Nucleotide Polymorphisms (SNPs), offering tools for improving rubber yield through marker-assisted selection (Luo *et al*., 2017). *H. brasiliensis* has also been studied extensively for similar purposes (Mantello *et al*., 2014). Additionally, *C. procera* is a drought- and salinity-tolerant species that thrives in arid and semi-arid regions. Due to its production of sap containing natural rubber, it is emerging as a promising candidate for this purpose (Kaur *et al*., 2021).

Transcriptome analysis through RNA sequencing serves as a powerful tool for examining gene expression patterns and regulatory networks (Tan *et al*., 2023). This approach has been successfully applied to elucidate the complexities of natural rubber biosynthesis in *H. brasiliensis*, leading to the identification of key genes and pathways in this plant (Mantello *et al*., 2014). Moreover, studies on the transcriptome of *C. procera* have elucidated various biological aspects, including responses to abiotic stresses (Ramadan and Hassanein, 2014), fiber morphogenesis (Khatoon *et al*., 2018), and the profiling of cysteine proteases (Kwon *et al*., 2015), highlighting the power of transcriptome analysis in elucidating molecular mechanisms that underpin key traits in *C. procera*. The discovery of the molecular details of natural rubber biosynthesis in *C. procera* is crucial for enhancing its production and fully exploiting its potential. Moreover, rubber biosynthesis is intricately linked to the terpenoid pathway, which produces essential building blocks for diverse plant compounds. *De novo* transcriptome assembly plays a pivotal role in identifying genes and regulatory elements within these pathways, especially in non-model plants (Ayachit *et al*., 2019; Mokhtari *et al*., 2023).

Despite the growing recognition of *C. procera* as a promising alternative source for natural rubber production, a comprehensive understanding of its molecular mechanisms remains lacking. While previous studies have provided insights into the plant's medicinal properties and the general chemical composition of its latex, a significant gap exists in detailed molecular studies linking specific genes to the biosynthesis of natural rubber in *C. procera*. Although natural rubber biosynthesis has been well-characterized in *H. brasiliensis* and some other rubber-producing plants, the pathways and key regulatory genes involved in this process in *C. procera* remain largely unexplored, which hinders the full exploitation of its potential for natural rubber production.

Moreover, while the MVA and MEP pathways are known to play a central role in natural rubber production, the specific genes and enzymes responsible for rubber synthesis in *C. procera* have not been fully identified. No comprehensive transcriptome of this plant exists that can provide insights into the gene expression profiles associated with rubber biosynthesis. This lack of data hinders our ability to manipulate the production of natural rubber in this species for industrial and commercial applications.

The goal of this study is to fill this gap by performing *de novo* transcriptome assembly of *C. procera* and identifying the key genes involved in natural rubber biosynthesis. By analyzing the plant's transcriptome, we aim to provide a comprehensive list of candidate genes involved in the biosynthetic pathways of natural rubber, which could pave the way for further functional validation and optimization of rubber production.

2. Materials and methods

2.1. Data collection and preparation

For the *de novo* assembly of the transcriptome of *C. procera*, the raw dataset (accession numbers: SRR1554320, SRR25601465) in FASTQ format was retrieved from the SRA database. The data were uploaded to the Galaxy server (usegalaxy.org, version 24.1.4) using the 'Download and Extract Reads in FASTQ' tool (version 3.1.1). Subsequently, the quality of the reads in the selected SRR samples was assessed using the 'FastQC' tool (version 0.12.1) (Andrews, 2010), with a focus on GC and N content, adapter presence, and other factors. Following this, the 'Trimmomatic' tool (version 0.39) (Bolger *et al.*, 2014) was employed to trim and remove low-quality bases and adapter sequences, utilizing the following parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:36.

2.2. De novo assembly of short reads

Following the read correction process, *de novo* assembly was performed using two tools: 'Trinity' (version 2.15.1) (Grabherr *et al.*, 2011) and 'rnaSPAdes' (version 3.15.5) (Bushmanova *et al.*, 2019), employing the default settings for both. To assess the quality of the assembly generated by the aforementioned tools, the 'rnaQUAST' tool (version 2.3.0) (Bushmanova *et al.*, 2016) was utilized. Additionally, to complement the quality assessment of the assembly, the 'BUSCO' tool (version 5.8.0) (Seppey *et al.*, 2019; Simão *et al.*, 2015) was employed, selecting the Viridiplantea lineage (green plants).

2.3. Candidate gene selection in the rubber biosynthesis pathway

After performing read assembly using the Trinity software, the resulting assembly was searched against plant candidate genes (*REF*, *HMGR*, *CPT*, *FPS*, and *SRPP*) using the 'R' software and the 'matchPDict' function from the 'Biostrings' package (Pagès *et al.*, 2019). Following the identification of high-scoring transcripts with more than 70% similarity, their alignment with the genome of *C. procera* was carried out using the NCBI database. In the next step, gene prediction was performed using the *Arabidopsis thaliana* genome in the online 'FGENESH' software (Solovyev *et al.*, 2006). To validate the predicted results, each predicted gene was searched against the SRA data of *C. procera* using the 'BLASTn' tool.

2.4. Prediction of coding regions

To predict the coding sequence (CDS) in the assembled unigenes, the 'TransDecoder' tool (Haas *et al.*, 2013), version 5.5.0, was utilized. The applied settings included the selection of standard genetic codes, designating 100 amino acids as the minimum protein length, and setting 500 base pairs as the maximum length of the open reading frame (ORF) for training the Hidden Markov Model (HMM).

2.5. Functional annotation of assembled transcriptome

For the functional annotation of the assembled transcriptome of the *C. procera*, the plant-specific tool 'Hayai-Annotation Plants' (Ghelfi *et al.*, 2019), version 2, available at pgdbjsnp.kazusa.or.jp/app/hayai2, was utilized with the default settings. The resulting graphs from the data analysis in this section were generated using Excel (version 2019) and ChiPlot (ChiPlot, 2024) software.

2.6. Prediction of metabolic pathways

For the automatic assignment of K numbers (Knum) and annotation related to orthologs in the KEGG database (KO), as well as the prediction of metabolic pathways involved in rubber biosynthesis in *C. procera*, the predicted protein and peptide sequences generated by the 'TransDecoder' tool were input into the online software Ghostkoala (Kanehisa *et al.*, 2016). Subsequently, the assigned K numbers for enzymes involved in the biosynthetic pathway were reconstructed using the 'Reconstruct' tool within the KEGG Mapper collection (Kanehisa and Sato, 2020; Kanehisa *et al.*, 2022). The results related to the natural rubber biosynthesis pathway were visualized on the map using the 'Color' tool.

3. Results and discussion

3.1. *De novo transcriptome assembly*

The transcriptome of *C. procera* was assembled *de novo* using two tools, 'rnaSPAdes' and 'Trinity', on the Galaxy server. Quality assessment results (Table 1) for the assembled transcriptome, evaluated using 'rnaQuast', revealed that the total number of transcripts assembled with 'rnaSPAdes' was slightly higher than that of 'Trinity'. However, 'Trinity' performed better in assembling transcripts longer than 500 base pairs. On the other hand, 'rnaSPAdes' was more successful in identifying transcripts longer than 1000 base pairs, with a higher average transcript length compared to 'Trinity'. Furthermore, the N50 value, which reflects the quality of transcriptome assembly, was higher for 'rnaSPAdes' than for 'Trinity'. These results generally suggest that 'rnaSPAdes' is more effective in assembling longer transcripts and optimizing data aggregation. In contrast, the quantitative evaluation results using the 'BUSCO' tool showed that both assembly tools performed similarly, with neither tool losing any essential transcripts. However, 'rnaSPAdes' assembled less complete, single-copy transcripts than 'Trinity', while also producing a higher number of duplicate transcripts compared to 'Trinity'. Additionally, 'rnaSPAdes' generated fewer fragmented transcripts than 'Trinity', indicating greater accuracy in simulating the gene structure (Fig. 1).

Table 1. Description of the results from the qualitative assessment of *de novo* assembly of *C. procera* transcripts using the 'rnaQuast' tool.

Statistics	Assembly	
	rnaSPAdes	Trinity
Transcripts	65790	65575
Transcripts > 500 bp	37797	38474
Transcripts > 1000 bp	30074	26515
Average length of transcripts	1310	1160
N50	2426	1986

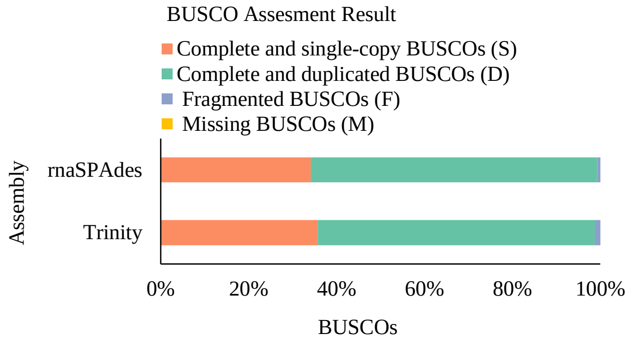


Figure 1. Description of the results from quantitative evaluation of *de novo* assembly of *C. procera* transcriptomes using the 'BUSCO' tool.

De novo transcriptome assembly is a crucial step in RNA-seq data analysis, aimed at reconstructing full-length transcripts from short reads, allowing us to study transcripts without the need for a reference genome (Haas *et al*., 2013). In this study, after performing *de novo* transcriptome assembly for *C. procera*, we utilized two powerful tools, 'rnaQUAST' and 'BUSCO', to conduct a comprehensive quality and quantitative evaluation. 'rnaQUAST' enabled us to precisely assess the assembly quality in terms of completeness, accuracy, and fragmentation. Various metrics are used to evaluate assembly quality, with N50 being one of the most commonly employed. This metric indicates the minimum contig length required to cover half of the total assembled length. In other words, a higher N50 value implies that more reads have overlapped to form larger contigs, resulting in a higher-quality assembly (Bernal-Gallardo and de Folter, 2024). On the other hand, evaluating the quality of assembled sequences is complex and often limited to metrics such as N50. 'BUSCO', by using a set of single-copy orthologous genes, allowed us to assess the completeness of the assembly in terms of essential genes. The fundamental principle of 'BUSCO' is that single-copy orthologous genes are rarely duplicated or lost during evolution, and thus their presence in an assembly indicates its completeness (Simão *et al*., 2015). The results from these two complementary tools provided us with a comprehensive assessment of the quality and quantity of the *C. procera* transcriptome assembly.

3.2. *Functional annotation of transcriptome*

The results obtained from the functional annotation of the transcriptome of *C. procera* with the 'Hayai-Annotation Plants' tool, and the assignment of hypothetical functions to the unigenes, revealed that the enrichment of pathways related to various biological processes, molecular functions and cellular components exhibited diverse and specific distributions among the genes (Fig. 2). The detailed distribution of the identified unigenes is illustrated in Fig. 3, which will be further explained below. In the realm of biological processes, the highest number of unigenes was found in the category of transcriptional regulation, which highlights the significant role of gene expression regulation in cellular processes. This process plays a crucial role, particularly in cell growth, differentiation, and the response to environmental stimuli. Another important biological process is carbohydrate metabolism, which indicates that cellular energy and the regulation of carbohydrate reserves directly influence cell functionality.

In molecular functions, the highest number of unigenes was observed in ATP binding and metal ion binding, emphasizing the importance of energy and the interactions between proteins and metal ions in various cellular processes. Moreover, after DNA and RNA binding activities, various protein-related activities, such as kinase activity, underscore the significance of regulatory processes in cells. These results suggest that proteins and enzymes responsible for metabolic and regulatory processes play a critical role in maintaining cellular homeostasis. In cellular components, membrane-associated components exhibited the highest number of unigenes, clearly reflecting the vital role of the membrane in cellular communication and material transport. Following the membrane, the nucleus contained the highest number of unigenes, emphasizing its importance as the center of genetic information and the regulation of biological processes. Overall, these findings suggest that cells function in a highly intricate collaboration across various biological processes, molecular functions, and cellular components. Understanding these interactions could have numerous applications in various biological fields.

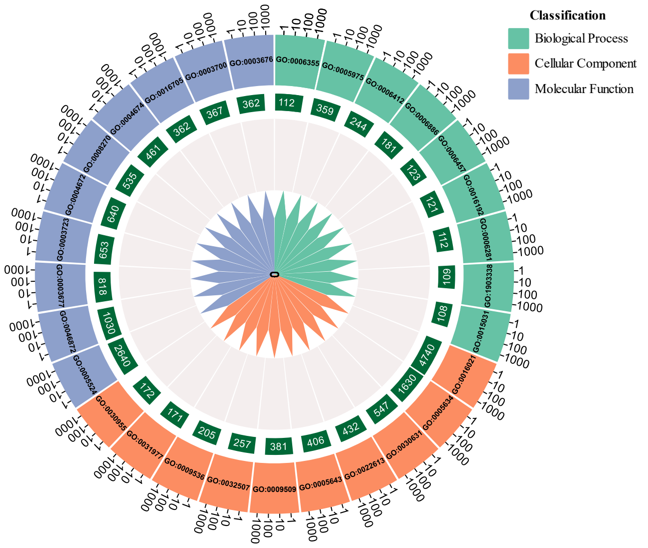


Figure 2. The abundance distribution of three domains —biological processes, cellular components, and molecular functions—associated with unigenes identified in Gene Ontology (GO) annotation of *C. procera* transcripts using 'Hayai-Annotation Plants' tool.

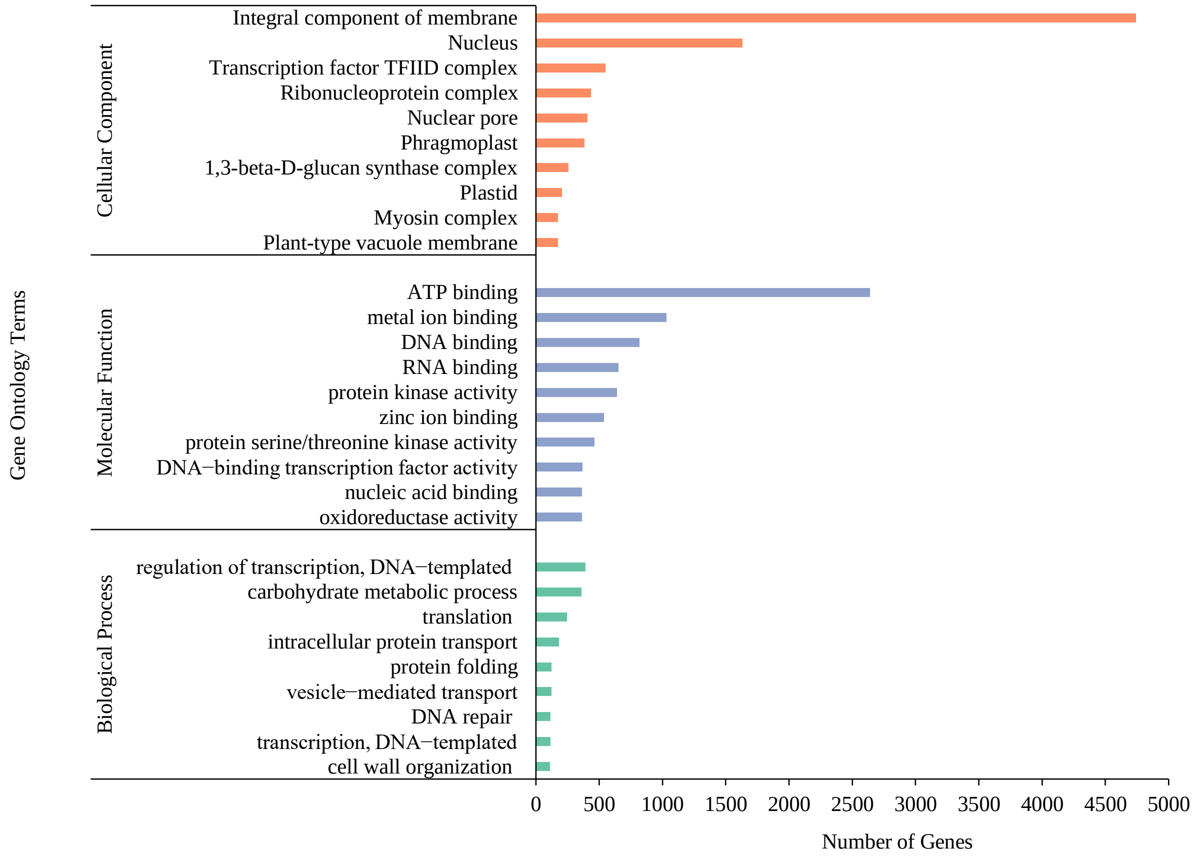


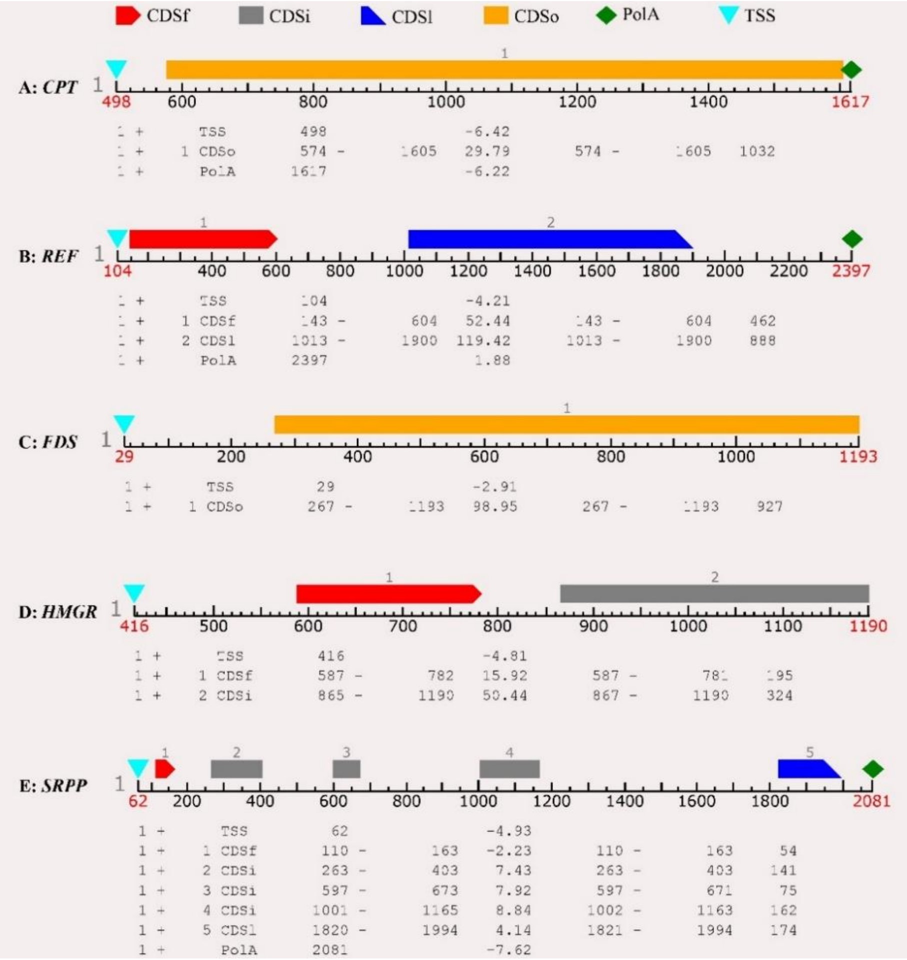
Figure 3. Predicted Gene Ontology (GO) terms associated with biological processes, cellular components, and molecular functions in data from *de novo* assembly annotation of *C. procera* transcriptome using 'Hayai-Annotation Plants' tool.

The 'Hayai-Annotation Plants' tool is designed for the rapid and precise annotation of functional genes in plants, utilizing the 'R' software. This platform combines sequence similarity searches using the 'USEARCH' tool against the 'UniProtKB' database (within the Embryophyta branch), with an additional step dedicated to functional annotation. The tool provides five layers of annotation, including protein names, Gene Ontology (GO) terms encompassing its three main domains (biological process, molecular function, and cellular component), Enzyme Commission (EC) numbers, protein existence levels, and the type of evidence supporting the annotation. The system employs a novel algorithm that prioritizes the protein existence level to expand GO and EC information, allowing for the annotation of representative peptide sequences from *A. thaliana* on a standard personal computer (Ghelfi and Isobe, 2025).

3.3. Prediction of candidate genes in the rubber biosynthesis pathway

Among the five candidate genes explored in this study—*REF*, *HMGR*, *CPT*, *FDS*, and *SRPP*—three of the latter genes were identified with 100% similarity to the targeted genes. The gene prediction conducted using the 'FGENESH' software also confirmed these genes as valid candidate genes. The results of the gene prediction indicated that all of them are located on the positive DNA strand. The coding sequence of the *cis-prenyl transferase* (*CPT*) gene contains a single exon, spanning nucleotides 574 to 1605 (Fig. 4A). Additionally, the predicted protein of this gene is 343 amino acids in length. Regarding the *rubber elongation factor* (*REF*) gene, two exons were observed, with exon one spanning nucleotides 143 to 604 and exon two spanning nucleotides 1013 to 1900. The predicted protein length for this gene is 449 amino acids (Fig. 4B). Furthermore, the coding sequence of the *farnesyl diphosphate synthase* (*FDS*) gene consists of a single exon, ranging from nucleotides 267 to 1193 (Fig. 4C), and its predicted protein is 308 amino acids long. For the key *3-hydroxy-3-methylglutaryl-CoA reductase* (*HMGR*) gene, two exons were observed, the first exon spans nucleotides 587 to 782, and the second exon spans nucleotides 865 to 1190, with the predicted protein consisting of 174 amino acids (Fig. 4D). Finally, for the small *rubber particle protein* (*SRPP*) gene, which encodes a small rubber particle protein, five exons were found, with lengths of 54, 141, 75, 162, and 174 nucleotides, respectively. The predicted protein length for this gene is 203 amino acids (Fig. 4E). Furthermore, searches conducted using the BLASTn tool against the SRA data of *C. procera* confirmed that the five identified genes are structurally and functionally consistent with genes found in the *C. procera* genome.

The biosynthesis of natural rubber involves two distinct metabolic pathways based on their synthesis locations: the MVA pathway, which occurs in the cytoplasm, and the MEP pathway, which takes place in plastids. Both pathways contribute to the synthesis of isopentenyl pyrophosphate (IPP), which is subsequently converted to its isomer, dimethylallyl pyrophosphate (DMAPP) (Yamashita and Takahashi, 2020). The MVA pathway begins with acetyl-CoA and continues through a series of enzymatic reactions, with a key step catalyzed by HMGR. On the other hand, the MEP pathway starts with pyruvate and glyceraldehyde-3-phosphate and includes a set of enzymes, including the rate-limiting enzyme 1-deoxy-*D*-xylulose 5-phosphate synthase (DXS) (Pütter *et al*., 2017). IPP and DMAPP are then utilized to synthesize larger prenyl pyrophosphate precursors such as geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP), and geranylgeranyl pyrophosphate (GGPP). These precursors serve as building blocks for various isoprenoids, including natural rubber. The final step in natural rubber biosynthesis involves the polymerization of IPP units by *cis*-prenyltransferase (CPT). CPT is a membrane-bound enzyme located within rubber particles, specialized organelles where rubber biosynthesis and storage occur (Cherian *et al*., 2019). The *rubber elongation factor* (*REF*) gene plays a crucial role in stabilizing and maintaining rubber particles in the biosynthetic pathway, although its precise function remains unclear (Yamashita and Takahashi, 2020). Additionally, the SRPP protein plays a key role in the biosynthesis of natural rubber. This protein acts as a cofactor in the rubber production process. Recent studies have shown that SRPP expression is influenced by various transcription factors, which could potentially enhance natural rubber production (Li *et al*., 2020). The rubber transferase (RT) complex is also considered a pivotal enzyme involved in catalyzing rubber biosynthesis. However, the exact composition of the RT complex in rubber-producing plants remains ambiguous and unknown (Tan *et al*., 2023). A comprehensive understanding of the rubber biosynthesis pathway enables researchers to explore metabolic engineering approaches aimed at enhancing both the yield and quality of rubber in plants. It has been demonstrated that the upregulation of key enzymes in the MVA and MEP pathways, such as HMGR and DXS, leads to increased rubber production in specific plants. Furthermore, manipulating the expression of *CPT* and other genes related to rubber biosynthesis also holds promise for enhancing rubber biosynthesis (Cherian *et al*., 2019).



```
1 +      TSS        498               -6.42
1 +    1 CDSo       574 -      1605   29.79       574 -      1605    1032
1 +      PolA      1617               -6.22
```

```
1 +      TSS        104               -4.21
1 +    1 CDSf       143 -       604   52.44       143 -       604     462
1 +    2 CDSl      1013 -      1900  119.42      1013 -      1900     888
1 +      PolA      2397                1.88
```

```
1 +      TSS         29               -2.91
1 +    1 CDSo       267 -      1193   98.95       267 -      1193     927
```

```
1 +      TSS        416               -4.81
1 +    1 CDSf       587 -       782   15.92       587 -       781     195
1 +    2 CDSi       865 -      1190   50.44       867 -      1190     324
```

```
1 +      TSS         62               -4.93
1 +    1 CDSf       110 -       163   -2.23       110 -       163      54
1 +    2 CDSi       263 -       403    7.43       263 -       403     141
1 +    3 CDSi       597 -       673    7.92       597 -       671      75
1 +    4 CDSi      1001 -      1165    8.84      1002 -      1163     162
1 +    5 CDSl      1820 -      1994    4.14      1821 -      1994     174
1 +      PolA      2081               -7.62
```

Figure 4. Prediction of potential genes in *A. thaliana* genomic DNA using the 'FGENESH' tool. The abbreviations CDSf, CDSi, CDSl, CDSo, PolA, and TSS represent key components of a gene structure. Specifically, CDSf denotes the first coding sequence (including the start codon), CDSi refers to an internal coding sequence (internal exon), CDSl represents the last coding sequence (including the stop codon), and CDSo stands for a single coding sequence. PolA indicates the polyadenylation signal sequence, while TSS marks the transcription start site.

3.4. Natural rubber biosynthesis pathway reconstruction in C. procera

According to the results obtained from 'GhostKoala', approximately 19,350 unigene sequences (47.2%) were annotated with KO assignments. Among these, the highest number of annotations were related to protein families involved in genetic information processing, carbohydrate metabolism, and those involved in signaling pathways and cellular processes (Fig. 5). The biosynthetic pathway reconstruction results, using the 'Reconstruct' tool based on the output from 'GhostKoala' in the 'KEGG Mapper', identified a total of 424 pathways in the assembled transcriptome of *C. procera*. Out of these, 19 ortholog genes were related to the biosynthesis of terpenoids. Furthermore, this part of the study showed that the unigene sequences receiving GhostKoala annotations belonged to 96 distinct modules, among which the biosynthesis modules for terpenoids and polyketides were identified. Notably, the C_5 isoprenoid biosynthesis submodule from the MVA pathway (7 orthologs), the C_5 isoprenoid biosynthesis submodule from the MEP pathway (8 orthologs), and the C_{10}-C_{20} isoprenoid biosynthesis submodule in plants (4 orthologs), all of which are associated with the biosynthesis of natural rubber, were identified (Fig. 6).

In the cytoplasmic pathway, acetyl-CoA is the primary precursor. This pathway consists of six main steps, facilitated by the enzymes acetyl-CoA acetyltransferase (ACAT, [EC:2.3.1.9]), HMG-CoA synthase (HMGS, [EC:2.3.3.10]), HMGR, [EC:1.1.1.34], mevalonate kinase (MVK, [EC:2.7.1.36]), phosphomevalonate kinase (PMK, [EC:2.7.4.2]), and mevalonate diphosphate decarboxylase (MVD, [EC:4.1.1.33]). Among these enzymes, HMGR has been identified as the key rate-limiting enzyme in this pathway (Cherian *et al*., 2019). In contrast, the plastidial MEP pathway utilizes eight consecutive enzymes to produce IPP and DMAPP from precursors. The first step in this pathway is rate-limiting and is catalyzed by the enzyme DXS, [EC:2.2.1.7], which converts the precursors pyruvate and glyceraldehyde-3-phosphate into 1-deoxy-*D*-xylulose 5-phosphate (DXP). All genes encoding the enzymes of the MEP pathway in plants have a prokaryotic origin (Cordoba *et al*., 2009).

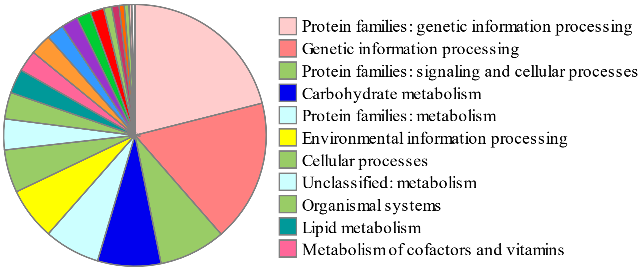


Figure 5. Results of 'GhostKOALA's functional annotation tool on data obtained from 'TransDecoser' tool.

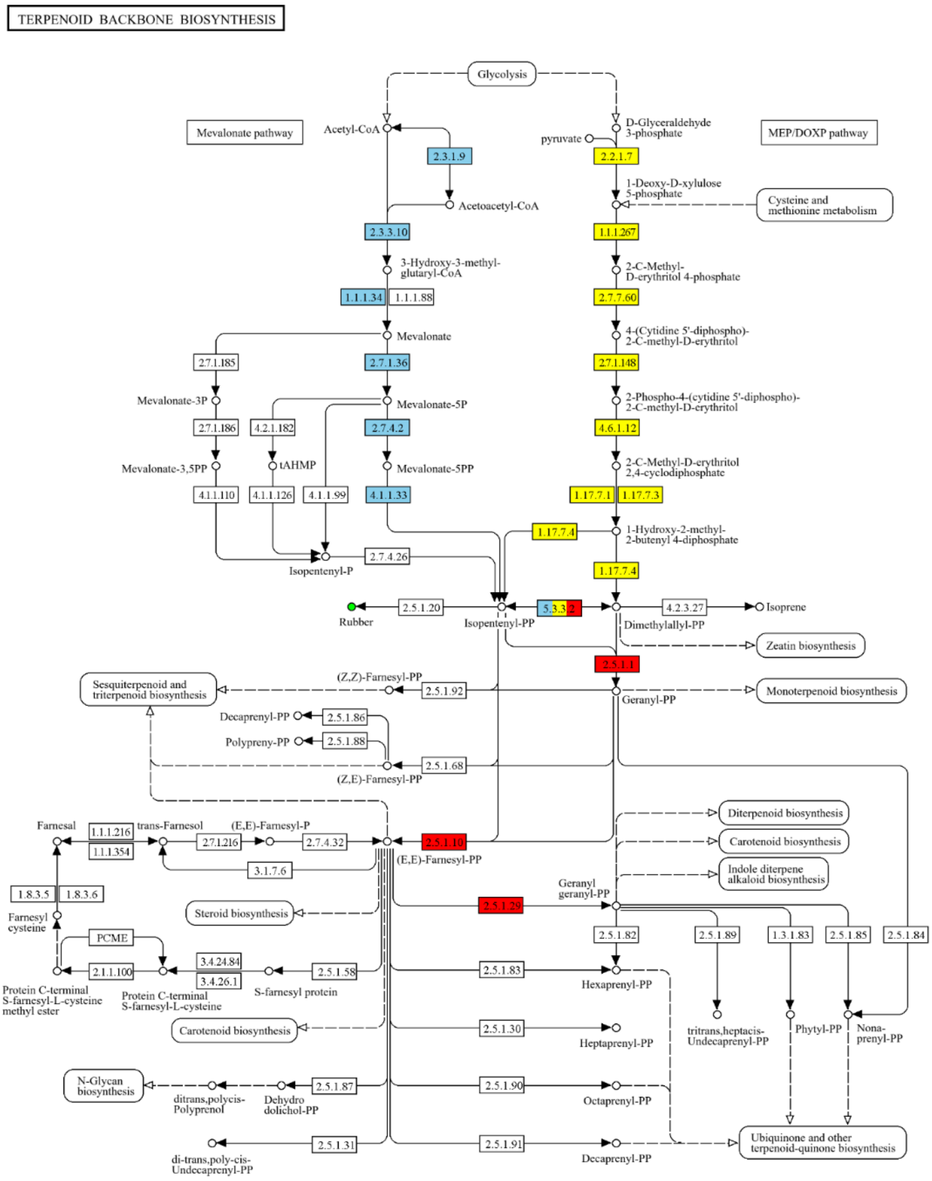


Figure 6. The backbone of the terpenoid biosynthetic pathway (map00900) and display of annotated unigenes (colored rectangles) involved in the rubber biosynthetic pathway. Blue, yellow, and red rectangles represent the MVA, MEP, and C10-C20 isoprenoid biosynthesis pathways in plants, respectively (Kanehisa *et al*., 2022).

Transcriptomic analyses shed light on the mechanisms of terpenoid biosynthesis. *Gymnema sylvestre* and *Valeriana officinalis* transcriptome analyses identified numerous genes involved in the terpenoid pathway (Ayachit *et al*., 2019; Mokhtari *et al*., 2023). Additionally, *in silico* transcriptome analysis has proven essential in identifying microRNAs (miRNAs) as crucial regulators, with miR5021 and miR838 in *Taraxacum* spp. being linked to terpene synthesis and influencing key enzyme expression for rubber precursors (Karimi *et al*., 2022).

Further research is essential to fully elucidate the rubber biosynthesis pathway and its regulatory mechanisms in *C. procera* and other alternative rubber-producing plants. The identification and characterization of new genes and enzymes involved in rubber biosynthesis, along with a comprehensive understanding of the regulatory mechanisms controlling rubber production, will be crucial for developing effective approaches to enhance rubber yield and quality in these plants. Additionally, exploring the potential of synthetic biology approaches to design and engineer novel rubber biosynthesis pathways could pave the way for the sustainable development of this product (Ning *et al*., 2023).

4. Conclusion

The *de novo* assembly approach is essential for constructing comprehensive transcriptomes. In this study, given the nature of *C. procera* as a non-model plant, a *de novo* transcriptome assembly approach was employed. This approach was crucial as it allowed for the characterization of the plant's transcriptome without relying on an existing reference genome. RNA-seq data generates vast amounts of information, enabling the identification of numerous transcripts. Assigning functional annotations to the predicted unigenes, including GO analysis and exploring the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, was invaluable in identifying genes involved in the biosynthetic pathway of natural rubber. While the primary biosynthetic pathways of natural rubber, namely the MVA and MEP pathways, are well-established, many genetic and regulatory details of these pathways remain poorly understood, particularly in plants such as *C. procera*. Thus, this study plays a crucial role in complementing existing knowledge by identifying key candidate genes and candidate enzymes involved in these pathways. The analysis revealed the presence of key genes and enzymes related to the MVA and MEP pathways, which are essential for the production of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), precursors for rubber biosynthesis. This research not only aids in the identification and investigation of key candidate genes in the natural rubber biosynthetic pathway but also, by introducing bioinformatics tools and innovative methods, paves the way for further research in this field. Ultimately, the findings of this study contribute to a deeper understanding of the molecular mechanisms involved in the biosynthetic pathway of natural rubber production in *C. procera* and could serve as a foundation for the development of more sustainable and novel methods for its production.

Conflict of interests

All authors declare no conflict of interest.

Ethics approval and consent to participate

No humans or animals were used in the present research. The authors have adhered to ethical standards, including avoiding plagiarism, data fabrication, and double publication.

Consent for publications

All authors read and approved the final manuscript for publication.

Availability of data and material

All the data are embedded in the manuscript.

Authors' contributions

All authors contributed to the study conception and design. F. Ali-Askari was responsible for investigation, data curation, formal analysis, and writing the original draft as well as reviewing and editing the manuscript. M. Sabokdast provided supervision, while M.R. Naghavi served as the advisor for the study.

Informed consent

The authors declare not to use any patients in this research.

Funding/Support

This study was supported by the University of Tehran, Karaj, Iran.

Acknowledgement

This article was achieved based on the material and equipment of Department of Agronomy and Plant Breeding, College of Agriculture and Natural Resources, University of Tehran, that the authors thanks it.

References

Al-Rowaily S.L., Abd-ElGawad A.M., Assaeed A.M., Elgamal A.M., Gendy A., Mohamed T.A., Dar B.A., Mohamed T.K., Elshamy A.I. 2020. Essential oil of *Calotropis procera*: comparative chemical profiles, antimicrobial activity, and allelopathic potential on weeds. Molecules 25(21): 5203. https://doi.org/10.3390/molecules25215203

Andrews S. 2010. FastQC a quality control tool for high throughput sequence data. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Avazzadeh A.M., Saeidi H., Bagheri A. 2018. Leaf morphological study of *Calotropis procera* (Aiton) R.br. in Iran. Paper presented at the 20th National and 8th International Congress on Biology of Iran, Maragheh. https://civilica.com/doc/850409

Ayachit G., Shaikh I., Sharma P., Jani B., Shukla L., Sharma P., Bhairappanavar S.B., Joshi C., Das J. 2019. *De novo* transcriptome of *Gymnema sylvestre* identified putative lncRNA and genes regulating terpenoid biosynthesis pathway. Scientific Reports 9(1): 14876. https://doi.org/10.1038/s41598-019-51355-x

Bernal-Gallardo J.J., de Folter S. 2024. Plant genome information facilitates plant functional genomics. Planta 259(5): 117. https://doi.org/10.1007/s00425-024-04397-z

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Bushmanova E., Antipov D., Lapidus A., Prjibelski A.D. 2019. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. GigaScience 8(9): giz100. https://doi.org/10.1093/gigascience/giz100

Bushmanova E., Antipov D., Lapidus A., Suvorov V., Prjibelski A.D. 2016. rnaQUAST: a quality assessment tool for *de novo* transcriptome assemblies. Bioinformatics 32(14): 2210-2212. https://doi.org/10.1093/bioinformatics/btw218

Cherian S., Ryu S.B., Cornish K. 2019. Natural rubber biosynthesis in plants, the rubber transferase complex, and metabolic engineering progress and prospects. Plant Biotechnology Journal 17(11): 2041-2061. https://doi.org/10.1111/pbi.13181

ChiPlot. 2024. ChiPlot Retrieved December 28, 2024. https://www.chiplot.online/

Cordoba E., Salmi M., León P. 2009. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. Journal of Experimental Botany 60(10): 2933-2943. https://doi.org/10.1093/jxb/erp190

Falana M.B., Nurudeen Q.O. 2020. Evaluation of phytochemical constituents and in vitro antimicrobial activities of leaves extracts of *Calotropis procera* against certain human pathogens. Notulae Scientia Biologicae 12(2): 208-221. https://doi.org/10.15835/nsb12210699

Fang F., Yuan B., He L., He M., Wang X. 2025. Roles of rubber elongation factor and small rubber particle protein in rubber particles. Plant Molecular Biology 115(3): 71. https://doi.org/10.1007/s11103-025-01593-7

Ghelfi A., Isobe S. 2025. Hayai-Annotation: a functional gene prediction tool that integrates orthologs and gene ontology for network analysis in plant species. Computational and Structural Biotechnology Journal 27: 117-126. https://doi.org/10.1016/j.csbj.2024.12.011

Ghelfi A., Shirasawa K., Hirakawa H., Isobe S. 2019. Retracted: Hayai-Annotation Plants: an ultra-fast and comprehensive functional gene annotation system in plants. Bioinformatics 35(21): 4427-4429. https://doi.org/10.1093/bioinformatics/btz380

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. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29(7): 644-652. https://doi.org/10.1038/nbt.1883

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. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8(8): 1494-1512. https://doi.org/10.1038/nprot.2013.084

Kanehisa M., Sato Y. 2020. KEGG mapper for inferring cellular functions from protein sequences. Protein Science 29(1): 28-35. https://doi.org/10.1002/pro.3711

Kanehisa M., Sato Y., Kawashima M. 2022. KEGG mapping tools for uncovering hidden features in biological data. Protein Science 31(1): 47-53. https://doi.org/10.1002/pro.4172

Kanehisa M., Sato Y., Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology 428(4): 726-731. https://doi.org/10.1016/j.jmb.2015.11.006

Karimi A.A., Naghavi M.R., Peyghambari S.A., Sobhani A., Rasoulnia A. 2022. Identification of miRNAs and their target genes in *Taraxacum* spp. Journal of Agricultural Science and Technology 24(6): 1457-1471. https://doi.org/10.52547/jast.24.6.1457

Kaur A., Batish D.R., Kaur S., Chauhan B.S. 2021. An overview of the characteristics and potential of *Calotropis procera* from botanical, ecological, and economic perspectives. Frontiers in Plant Science 12: 690806. https://doi.org/10.3389/fpls.2021.690806

Khatoon A., Iqbal N., Asif M., Cheema H.M., Saeed M., Bashir A. 2018. Comparative analysis of fiber morphogenesis genes of *Calotropis procera* and *Gossypium hirsutum*. International Journal of Agriculture and Biology 20: 288-296. https://doi.org/10.17957/ijab/15.0492

Kwon C.W., Park K.M., Kang B.C., Kweon D.H., Kim M.D., Shin S.W., Je Y.H., Chang P.S. 2015. Cysteine protease profiles of the medicinal plant *Calotropis procera* R. Br. revealed by *de novo* transcriptome analysis. PLoS One 10(3): e0119328. https://doi.org/10.1371/journal.pone.0119328

Li H.L., Qu L., Guo D., Wang Y., Zhu J.H., Peng S.Q. 2020. Histone deacetylase interacts with a WRKY transcription factor to regulate the expression of the small rubber particle protein gene from *Hevea brasiliensis*. Industrial Crops and Products 145: 111989. https://doi.org/10.1016/j.indcrop.2019.111989

Luo Z., Iaffaldano B.J., Zhuang X., Fresnedo-Ramírez J., Cornish K. 2017. Analysis of the first *Taraxacum kok-saghyz* transcriptome reveals potential rubber yield related SNPs. Scientific Reports 7(1): 9939. https://doi.org/10.1038/s41598-017-09034-2

Mantello C.C., Cardoso-Silva C.B., da Silva C.C., de Souza L.M., Scaloppi Junior E.J., de Souza Gonçalves P., Vicentini R., de Souza A.P. 2014. *De novo* assembly and transcriptome analysis of the rubber tree (*Hevea brasiliensis*) and SNP markers development for rubber biosynthesis pathways. PloS One 9(7): e102665. https://doi.org/10.1371/journal.pone.0102665

Mohamed N., Ismail M., Abdel-Mageed W., Shoreit A. 2017. Biodegradation of natural rubber latex of *Calotropis procera* by two endophytic fungal species. Journal of Bioremediation & Biodegradation 8(1): 380. https://doi.org/10.4172/2155-6199.1000380

Mokhtari A., Omidi M., Ebrahimi M., Alizade H., Sobhani A. 2023. Evaluation of the transcriptome of valerian (*Valeriana officinalis*) to identify genes involved in terpenoids biosynthesis pathway. Modern Genetics Journal 18(2): 111-123. (In Farsi). http://mg.genetics.ir/article-1-1793-en.html

Ning Y., Yang D.D., Yu X.C., Cao X. 2023. Multi-omics-driven development of alternative crops for natural rubber production. Journal of Integrative Agriculture 22(4): 959-971. https://doi.org/10.1016/j.jia.2023.03.007

Pagès H., Aboyoun P., Gentleman R., DebRoy S. 2019. Biostrings: efficient manipulation of biological strings. R package version 2.78.0. https://doi.org/doi:10.18129/B9.bioc.Biostrings

Pandey A., Swarnkar V., Pandey T., Srivastava P., Kanojiya S., Mishra D.K., Tripathi V. 2016. Transcriptome and metabolite analysis reveal candidate genes of the cardiac glycoside biosynthetic pathway from *Calotropis procera*. Scientific Reports 6(1): 34464. https://doi.org/10.1038/srep34464

Pütter K.M., van Deenen N., Unland K., Prüfer D., Schulze Gronover C. 2017. Isoprenoid biosynthesis in dandelion latex is enhanced by the overexpression of three key enzymes involved in the mevalonate pathway. BMC Plant Biology 17: 88. https://doi.org/10.1186/s12870-017-1036-0

Ramadan A.M., Hassanein S.E. 2014. Characterization of *P5CS* gene in *Calotropis procera* plant from the de novo assembled transcriptome contigs of the high-throughput sequencing dataset. Comptes Rendus Biologies 337(12): 683-690. https://doi.org/10.1016/j.crvi.2014.09.002

Seppey M., Manni M., Zdobnov E.M. 2019. BUSCO: assessing genome assembly and annotation completeness. Gene prediction: Methods and Protocols (pp. 227-245). https://doi.org/10.1007/978-1-4939-9173-0_14

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210-3212. https://doi.org/10.1093/bioinformatics/btv351

Solovyev V., Kosarev P., Seledsov I., Vorobyev D. 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biology 7(Suppl 1): S10. https://doi.org/10.1186/gb-2006-7-s1-s10

Tan Y., Cao J., Tang C., Liu K. 2023. Advances in genome sequencing and natural rubber biosynthesis in rubber-producing plants. Current Issues in Molecular Biology 45(12): 9342-9353. https://doi.org/10.3390/cimb45120585

Yamashita S., Takahashi S. 2020. Molecular mechanisms of natural rubber biosynthesis. Annual Review of Biochemistry 89(1): 821-851. https://doi.org/10.1146/annurev-biochem-013118-111107

HOW TO CITE THIS ARTICLE

Ali-Askari F., Sabokdast M., Naghavi M.R. 202x. *De Novo* Transcriptome Assembly of *Calotropis procera* for Insights into the Natural Rubber Biosynthesis Pathway. *Agrotechniques in Industrial Crops* x(x): xx-xx. 10.22126/ATIC.2025.12119.1226