The Pharma Innovation Journal 2021; 10(7): 31-35

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; 10(7): 31-35

www.thepharmajournal.com
Received: 17-05-2021
Accepted: 21-06-2021

Umaretiya Vidisha R
Department of Biotechnology, College of Agriculture, J.A.U., Junagadh, Gujarat, India

Dr. DR Mehta
Professor and Head, Department of Genetics and Plant Breeding, College of Agriculture, J.A.U., Junagadh, Gujarat, India

Hirani Nimisha V
Department of Biotechnology, College of Agriculture, J.A.U, Junagadh, Gujarat, India

Corresponding Author:
Umaretiya Vidisha R
Department of Biotechnology, College of Agriculture, J.A.U., Junagadh, Gujarat, India

De novo transcriptome analysis of a medicinal plant: *Telosma pallida* L.

Umaretiya Vidisha R, Dr. DR Mehta and Hirani Nimisha V

Abstract
Telosma pallida L. is a member of the Asclepiadaceae family of plants with 2n=22. It is been said to neutralise Tridosha in the past. Next-generation sequencing was carried out in an oxford nanopore sequencer and 571,448, 307,608, 1,320,760, 14,597, and 43,006 sequences were generated in the root, stem, leaf, flower, and fruit, respectively. After *de novo* assembly, N50 measurements of contigs from CLC, CAP3, and CD-HIT were found to be 1557 bp, 1769 bp, and 1827 bp, respectively. The top hit species distribution was against the species *Coffea arabica*, according to a similarity search. Differential gene expression revealed that enzymes involved in the biosynthesis of monoterpenoid indole alkaloids were more prevalent in the stem and leaf than in the root, flower and fruit. It is the first-ever information about this plant and it will help with pharmacological studies of *Telosma pallida* L.

Keywords: Medicinal plant, *Telosma pallida*, *de novo* assembly, differential gene expression

1. Introduction
Telosma pallida L., commonly known as *Pergularia pallida* L., is an undiscovered ayurvedic medicinal plant belongs to the Asclepiadaceae family. In Gujarati, this plant is known as "Radarudi" or "Varshadodi," and in Hindi, it is known as "Surkilla." The somatic chromosome number is 2n=22. This valuable medicinal plant is available and grown in India, Myanmar, Nepal, Pakistan, Thailand, Vietnam etc. The flowering season is July to September i.e. during monsoon. It is traditionally used to balance the Tridosha (vata, pitta, and kapha) energies. The leaves have an astringent flavour and a distinct odour (Kanakhara *et al.*, 2018) [7]. The flower is yellowish-green in colour, scented, and has five petals. Flowers are eaten as a vegetable in India's Saurashtra area. It has long been used to treat whooping cough, the common cold, and asthma.

The basic understanding the transcriptome is very essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues (Wang *et al*., 2009) [13]. Long-read sequencing technologies, such as oxford nanopore technologies, have been introduced into the transcriptomics area and allowing individual reads that cover the whole length of transcripts. The effectiveness of *de novo* transcriptome assemblies in organisms without a reference genome was demonstrated by the availability of well-defined computational tools and well-applied methodology. In the absence of a reference sequence, *de novo* sequencing is often achieved by assembling individual sequence reads into longer contiguous sequences (contigs) or correctly ordered contigs. This sort of sequencing has aided in the knowledge of non-model organisms and has become one of the most prominent methods for gene discovery and expression profiling in non-model species (Grabherr *et al*., 2011) [4].

Materials and Methods
RNA Isolation and Sequencing
Plant material was collected during the July to September months from Joshipura area (21.55711ºN, 70.44772ºE) of Junagadh district, Gujarat, India (Figure 1). Collected samples were immediately stored at -800C. Total RNA was isolated using TRIZOL method from root, flower and fruit, while for stem and leaf samples, the method used was Mornkham *et al*., (2013) [8]. For stem and leaf, total RNA was initially try to extract from 50 mg ground stem and leaf material using five existing extraction methods. These methods are (1) TRIZOL reagent (Invitrogen), (2) Plant RNeasy mini kit (Qiagen, Germany), (3) CTAB method (Chang *et al*., 1993), (4) SDS method (Huded *et al*., 2018) [6] and (5) Method of Ghawana *et al*., (2011) [3]. These initial efforts generated an unsatisfactory amount and purity of total RNA. Most common problem with these methods was DNA contamination. The total RNA extract of root, stem, leaf, flower and fruit were tested for

integrity by 1.0% (w/v) agarose gel electrophoresis. RNA concentration was checked by Qubit™ RNA HS assay kit. For sequencing, Oxford Nanopore sequencer and cDNA-PCR barcoding kit was used.

Fig 1: Photograph showing the whole plant of *Telosma pallida*

De novo Assembly

Raw data quality analysis was carried by using CLC Genomics Workbench Version 9.5.4. Primary, secondary and tertiary assembly was carried out in CLC, CAP3 and CD-HIT, respectively. CDS prediction was done by Transdecoder 5.3.0 (Haas *et al*., 2013). *Arabidopsis thaliana* was use as reference genome to mapping. All transcript Reads were mapped using CLC software Genomics Workbench Version 9.5.4. Bioconductor 3.13 was used to calculate the RPKM (Reads Per Kilobase of transcript per Million mapped reads) value (Robinson *et al*., 2010) [9].

Gene Function Annotation

Functional annotation was carried out by using BLASTx using omicsbox 2.0.10 against the Uniprot plant database. The panther database and omicsbox 2.0.10 were used to perform the GO annotation (Anonymous, 2019) [1].

Results and Discussion

De novo Assembly, differential gene expression (DEG) and functional annotation

About 571,448, 307,608, 1,320,760, 14,597 and 43,006 sequences were generated in root, stem, leaf, flower and fruit, respectively. All sequences were trimmed and contain good quality reads confirmed by CLC genomic workbench 20.0. Between 25 and 55 per cent GC content was recorded in root, stem, and leaf, While 25 to 60 per cent was observed in the flower; and 25 to 50 per cent was found in the fruit. No ambiguous bases and duplicated sequences were detected. N50 measurement of contigs were found 1557 bp, 1769 bp, 1827 bp, respectively from CLC, CAP3 and CD-HIT (Table 1). Primary assembly of sequences in. Fasta format was carried in CLC genomic work bench 20.0. CLC genomic workbench created contig sequences by using all the information that are in the read sequences and then mapped using the simple contig sequence as a reference. CAP3 program had a capability to clip 5' and 3' low-quality regions of reads and uses base quality values in computation of overlaps between reads and construct of multiple sequence alignments of reads. From CLC, CAP3, and CD-HIT, L50 measurements of contigs were reported as 35694 bp, 4360 bp, 3601 bp as well as L90 measurement of entire assembly observed as 93324 bp, 10883 bp, and 8985 bp, respectively. After mapping all samples with master assembly, the percent of mapped genes in root, stem, leaf, flower, and fruit were 66.87 %, 67.29 %, 72.42 %, 70.2 %, and 70.31 %, respectively. The length of the maximum number of transcripts were between 0-499 bp and 2500-2999 bp. Similarly, the GC content of the assembled combined reads in the *Trachyspermum ammi* L. was 38.35% in 151,115 transcripts with N50 of 1291 bp. (Soltani *et al*., 2018) [11]. Zhao *et al*. (2020) [14] studied *De novo* assembly and characterization of the transcriptome in Chinese endemic *Euphorbia kansui.* After assembly, 58,362 unigenes were recovered in with an N50 length of 1,683 bp and reads had an average GC content of 44.09%

Table 1: Primary, secondary and tertiary assembly statistics of *de novo* transcriptome

	Software		
Assembly	**CLC**	**CAP3**	**CD_HIT**
Contigs (>= 0 bp)	201535	17224	12897
Contigs (>= 1000 bp)	69878	10050	8483
Contigs (>= 5000 bp)	144	27	23
Contigs (>= 10000 bp)	20	2	2
Total contigs	271577	27303	21405
Largest contig (bp)	22400	21964	21964
Total length (bp)	157616056	21359062	18162641
GC (%)	40.49	40.51	40.36
N50	1557	1769	1827
N90	721	891	914
L50	35694	4360	3601
L90	93324	10883	8985

Based on RPKM value, the number of genes expressed in between five samples determined for comparison for DEG. Contig11836.p1 was found to be strongly up- regulated in the root; moderately expressed in the stem, leaf, and fruit, and not expressed in the flower. S-adenosylmethionine synthase 1 is encoded by this gene. Contig7874 coded for vincadifformine 19-hydroxylase, a component of the monoterpenoid indole alkaloids, and had 11377.7 RPKM in the stem. Likewise, Conting11722, which has a 5113.5 RPKM value, was found to be responsible for adaxial/abaxial pattern determination as well as leaf and root morphogenesis. With 731.7 RPKM, RNA-binding protein 1 was substantially expressed in flower. Chalcone synthase played an important role in flavonoid biosynthesis enzyme that was also involved in auxin polar transport. It was widely found in flowers and fruits. Lipid metabolism was one of the most highly expressed contigs in fruit.

Using the BLASTX method, the CDS transcripts were compared to the non-redundant (NR) protein database at the NCBI. *Telosma pallida* sequences were best matched with *Coffea arabica* according to BLAST results (Family: Rubiaceae). In omics box 2.0.10, GO annotation was conducted. After annotation, the findings were sorted by taxonomic classification: flowering plants (taxa:3398, Magnoliopsida). Similarly, Zhang *et al*. (2015) [12] reported that the top-hit species distribution of the homology result of *Gentiana rigescens* against NR databases, showed high homology with sequences from land plants, among which the highest matches were to genes from *Coffea canephora* (36.08%). In terms of GO, the highest GO score was discovered in the integral component of membrane in the cellular component with a value of 105. Transmembrane transport in biological processes has the highest GO score of 35. In molecular function, metal ion binding (65.32) had the highest GO score (Figure 2 to 4). After enzyme code distribution, the maximum sequences were found for Transferases (64 CDS), whereas the smallest sequences were found for Ligases (4 CDS). KEGG (Kyoto Encyclopedia of Genes and Genomes Database) mapping was used to identify genes involved in biological pathways. These genes were determined to be engaged in a variety of functions and

expressed in a variety of places and 77 KEGG pathways were discovered. Genes involved in mRNA surveillance, MAPK signalling in plants, Glycerophospholipid metabolism, Peroxisome, Tyrosine metabolism, Photosynthesis, Glycosylphosphatidylinositol (GP)-anchor production, Glycerolipid metabolism etc. were found prominent. Similar findings were reported by Tian *et al*. (2015) in *Polygala tenuifolia*, candidates genes CYP450s and UGTs obtained, including putative terpenoid backbone and phenylpropanoid biosynthesis pathway. Roy *et al*. (2015) [10] studied transcriptome analysis in flowers, leaves, and roots of Korean medicinal herb *Cirsium japonicum* var. spinossimum. The differentially expressed gene analysis revealed that the expression of genes related to the flavonoid pathway was higher in the flowers, whereas the phenylpropanoid pathway was more highly expressed in the roots.

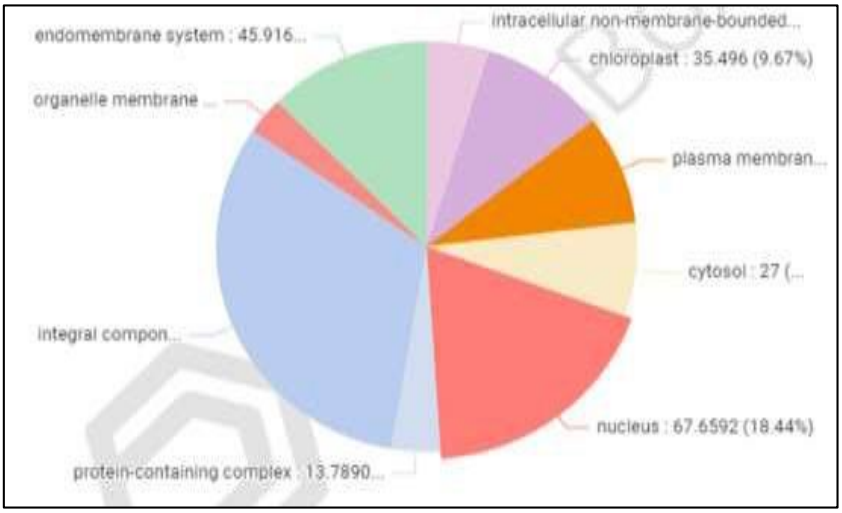

Fig 2: GO score distribution for cellular component

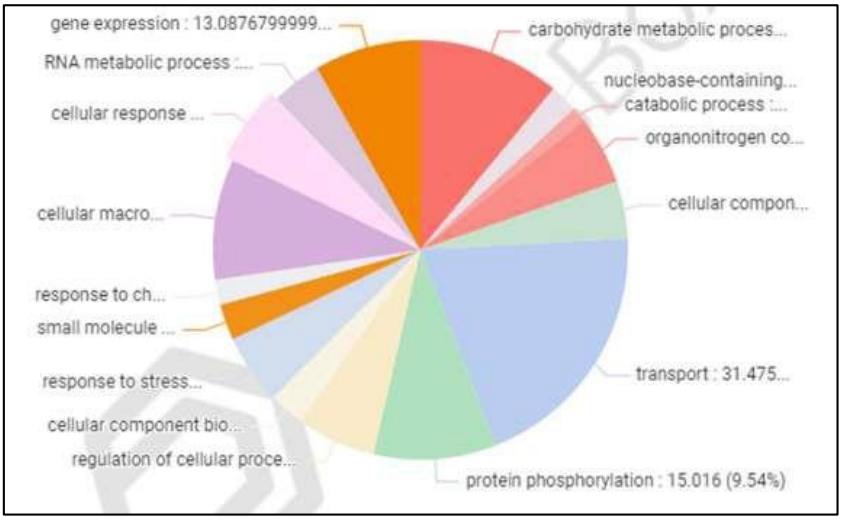

Fig 3: GO score distribution for biological process

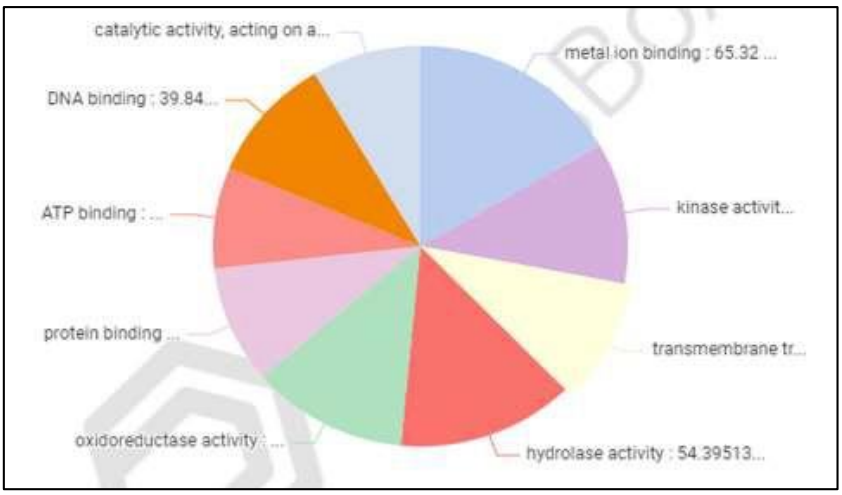

Fig 4: GO score distribution for molecular function

Reference mapping with *Arabidopsis thaliana*

In total, 22,208 genes were uniquely mapped, out of which 16,939 genes were non- specifically mapped, and 39,147 genes were presented as total mapped genes. In the stem, 14,524 genes were uniquely mapped; 15,857 genes were non-specifically mapped, and 30,381 genes were reported as total mapped genes.

There were 82,038 genes uniquely mapped in leaf; 46,498 non-specifically mapped genes, and 128,536 total mapped genes were reported. In flower, 2,311 genes were uniquely mapped and 3,224 genes were non- specifically mapped, with 5,535 genes were overall mapped. In fruit, there were 2,760 genes that were uniquely mapped, while 974 that were non-specifically mapped, with total of 3,734 mapped genes. Maximum genes mapped were present in leaf samples.

Differential expression analysis of significantly responsive genes were determined. LHB1B2, a photosystem II light harvesting complex gene, was the most upregulated gene in stem as compared to root with a factor change of 9.90. Similarly, PANK2 was found to be responsible for regulating the synthesis of coenzyme A in the leaf and was increased by 6.75 fold.

COB11 was involved in the deposition of apical pectin cap and cellulose microfibrils in pollen tubes and COBL11 was discovered to be increased in flowers. DA1, a gene involved in fruit shape determination, was discovered to be upregulated in fruit. Because protein synthesis activity was higher than protein breakdown, ubiquitin-associated genes were mainly downregulated in all regions of the plant.

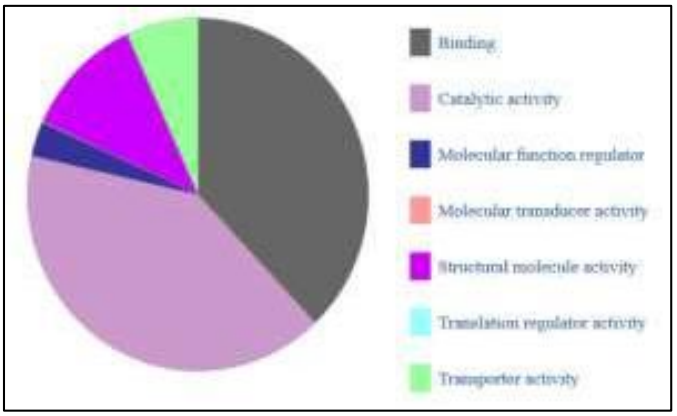

Fig 5: Overall molecular function (MF) of genes

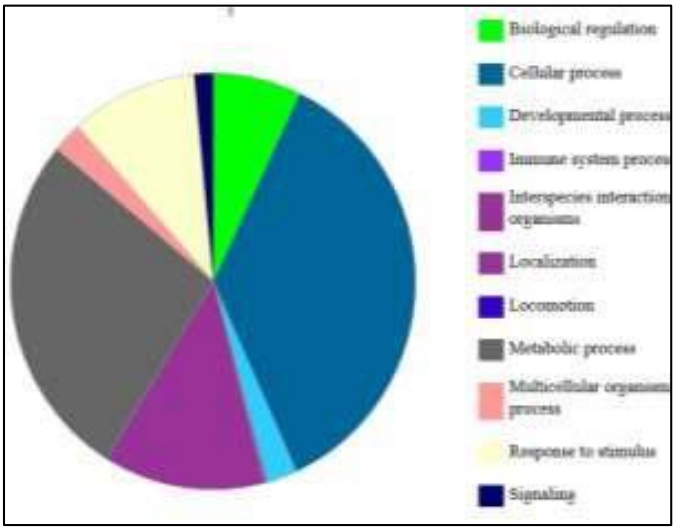

Fig 6: Overall biological process (BP) of genes

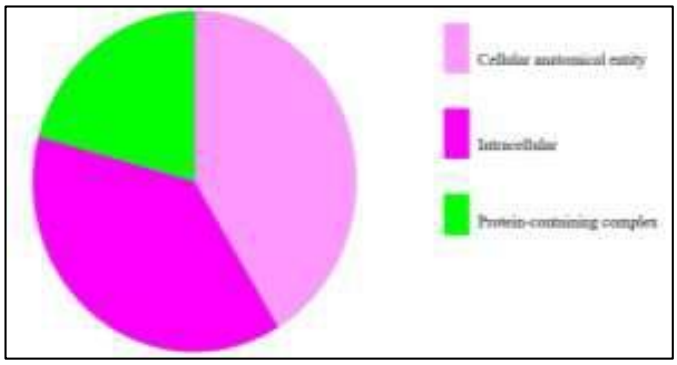

Fig 7: Overall cellular component (CC) of genes

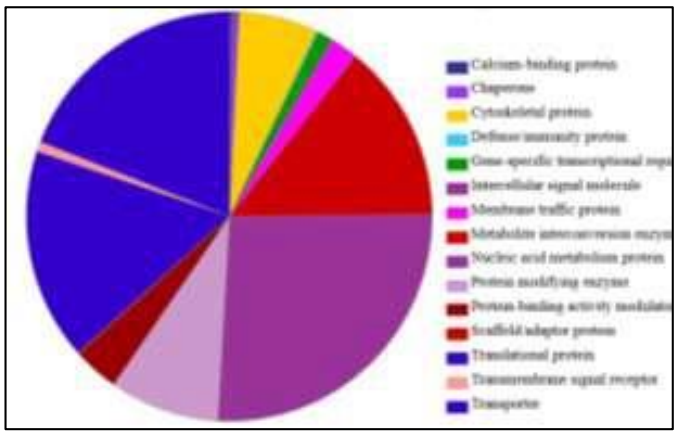

Fig 8: Overall protein class

After mapping with Arabidopsis thaliana, significant genes were identified for GO. The cellular anatomical entity was the most common cellular component discovered (67%). Likewise, catalytic activity was discovered to have a major molecular function (42 %). The cellular process was the most prominent biological process discovered (55 %). Major protein class expressed were nucleic acid metabolism proteins (Figure 5 to 8).

Conclusions

RNA-sequencing provides information on genes and their functions, as well as potential pathways. *Telosma pallida* L. root, stem, leaf, flower, and fruit tissues were subjected to a complete RNA-sequencing. In this work, 571,448, 307,608, 1,320,760, 14,597, and 43,006 sequences were generated in the root, stem, leaf, flower, and fruit, respectively. Genome sequencing is way expensive, thus transcriptome assembly is a good option. The first *de novo* transcriptome assembly of *Telosma pallida* L. was successfully completed. Assignment of GC content and functional categories using GO annotation was also carried out to facilitate and accelerate future genome-wide studies in this plant. KEGG pathway mapping revealed several essential plant pathways, including mRNA surveillance, MAPK signalling in plants, Glycerophospholipid metabolism, Peroxisome, Tyrosine metabolism, Photosynthesis, and others. It is the first-ever information about this plant and it will help with pharmacological studies of *Telosma pallida* L.

References

1. Anonymous. OmicsBox - Bioinformatics made easy. BioBam Bioinformatics (Version 2.0.10) 2019. www.biobam.com/omicsbox.
2. Chang S, Puryear J, Cairney J. A simple and efficient

method for isolating RNA from pine trees. Plant Molecular Biology Reporter 1993;11:113-116.
3. Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Bhardwaj PK *et al*. An RNA isolation system for plant tissues rich in secondary metabolites. BioMed Central research notes 2011;4:e85.
4. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 2011;29:644-652.
5. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J. *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nature Protocols 2013,8(8):494-512.
6. Huded A, Jingade P, Mishra MK. A rapid and efficient SDS-based RNA isolation protocol from different tissues of coffee. 3 Biotech 2018;8(3):183-186.
7. Kanakhara RD, Rudrappa HC, Shukla VJ, Acharya R. A detailed pharmacognostical and phytochemical profile on stem of *Telosma pallida* (L.) Kurz.: An extrapharmacopoeial drug. International J. Research in Ayurveda and Pharmacy 2018;9(1):36-43.
8. Mornkham T, Wangsomnuk P, Fu Y, Wangsomnuk P, Jogloy S, Patanothai A. Extractions of high quality RNA from the seeds of *Jerusalem artichoke* and other plant species with high levels of starch and lipid. Plant, 2013;2:302-316.
9. Robinson MD, McCarthy DJ, Smyth GK. Edge R: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-140.
10. Roy P, Abdulsalam FI, Pandey DK, Bhattacharjee A, Eruvaram NR, Malik T. Evaluation of antioxidant, antibacterial, and antidiabetic potential of two traditional medicinal plants of India: Swertia cordata and Swertia chirayita. Pharmacognosy Research 2015;7(1):57-62.
11. Soltani HM, Noori SA, Shariati JV. Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. Scientific reports 2018,8:e13405.
12. Tian H, Xu X, Zhang F, Wang Y, Guo S, Qin X, Du G. Analysis of *Polygala tenuifolia* transcriptome and description of secondary metabolite biosynthetic pathways by illumina sequencing. International J. Genomics, 2015:e782635.
13. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 2009,10:57-63.
14. Zhao X, Wang M, Chai J, Li Q, Zhou Y, Li Y, Cai X. *De novo* assembly and characterization of the transcriptome and development of microsatellite markers in a Chinese endemic *Euphorbia kansui*. Biotechnology and Biotechnological Equipment 2020;34(1):562-574.