

Leaves and Roots of Micropropagated Plants Adopt Different Pathways for Withanolide Biosynthesis in *Withania Coagulans*: A Comparative Transcriptome Study

Swati Gupta

Manipal University - Jaipur Campus

Manoj Kumar Bohra

Manipal University - Jaipur Campus

Kalpna sharma

Manipal University - Jaipur Campus

Sumita Kachhwaha

Rajasthan University: University of Rajasthan

Shanker Lal Kothari

Amity University Jaipur: Amity University - Jaipur Campus

Rohit Jain (✉ rohit.jain@jaipur.manipal.edu)

Manipal University - Jaipur Campus <https://orcid.org/0000-0001-5245-4627>

Research Article

Keywords: Paneer bandh, SSRs, cytochrome p450, methyltransferase, glycosyltransferase, withanogenesis

Posted Date: March 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-299335/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Withania coagulans (Stocks) Dunal commonly known as paneerbandh due to its milk coagulating properties, is a widely used medicinal herb, and micropropagated plants of the plant to produce some novel withanolides. Transcriptome sequencing of leaves and roots of micropropagated plants of *W. coagulans* assembled total 8.08 and 6.35 GB of raw reads into 292,074 and 16,474 high quality reads, out of which 267,119 and 15,758 unigenes were identified in WcL and WcR, respectively. Further, 40.6% WcL and 55.05% WcR unigenes were annotated using more than one database. Metabolic process and cellular components were identified as dominant categories in gene ontology, while total 20,927 WcL and 2,474 WcR unigenes were mapped to different biological pathways. KEGG classification aided in identification of genes involved in biosynthesis of withanolide precursor, 24-methylenecholesterol. All the genes related to withanolide precursor biosynthesis were present only in WcL, indicating *de novo* biosynthesis of withanolides, while absence of some rate limiting enzymes in WcR indicated towards biosynthesis of withanolides through salvage pathways. GTs, MTs and CYP450s were identified as putative genes involved in conversion of 24-methylenecholesterol to different withanolides. Differential expression of these genes further revealed details of enzymes involved in biosynthesis of tissue-specific withanolides. Further, withanolide profiling through HPLC analysis ascertained the differential biosynthesis and accumulation of withanolides in both the tissues in relatively very less concentration confirmed their biosynthesis through salvage mechanism in roots. Therefore, the present study can be fruitful for future research and product development of *W. coagulans* through pathway engineering. Moreover, the SSRs identified in this study can be used in marker assisted breeding and selection of biochemically elite varieties of *W. coagulans*.

Introduction

Withania coagulans (Stocks) Dunal, commonly known as “paneer bandh” or “Indian cheese maker” is one of the highly valued medicinal species of family Solanaceae, used as natural milk coagulant (Jain et al. 2012). The plant has been used in treatment of various diseases such as ulcers, rheumatism, dropsy, diabetes and cancer (Bhandari 1995; Jain et al. 2012). Anti-inflammatory, antitumor, immunomodulatory, cardioprotective, antioxidant, neuroprotective and antimicrobial properties of the plant have been attributed to a specific class of secondary metabolites known as ‘withanolides’ (Maurya et al. 2010; Jain et al. 2012). These ergosterone derivatives are the principle bioactive components of genus *Withania*, and represent a class of naturally occurring C-28 steroidal lactones in which C-22 and C-23, or C-26 and C-23 of C-28 steroidal backbone are oxidised to form either δ - or γ - lactone rings, respectively (Gupta et al. 2015; Agarwal et al. 2017). Withanolides are pharmaceutically active compounds used in formulations for the treatment of cancer, immune disorders and neurodegenerative diseases such as Alzheimer’s and Parkinson’s (Chaurasiya et al. 2012). These critical medicinal properties led to an exponential rise in global demand of withanolides (Sharada et al. 2007).

Variations in the withanolide profile of the plants growing in different agro-climatic conditions has greatly affected the efficacy of the pharmaceutical preparations and limits its uses at industrial scale. The plant

has almost been extinct from its natural habitat due to overexploitation, reproductive failure and poor seed setting (Jain et al. 2009). Thus, there is an urgent need for the development of conservation strategies and enhanced production of bioactive compounds from the plants.

In vitro propagation has provided as an effective approach for both conservation of elite germplasms of endangered species and higher production of bioactive compounds from cultured plant cells (Jain et al. 2009; Nagella and Murthy 2010; Jain et al. 2011). Sangwan et al. (2007) also concluded that *in vitro* cultures could serve as potent alternative to the field plants for production of therapeutically valuable compounds.

Manipulation of culture and growth conditions, use of elicitors, inducers and stress conditions to enhance the production of different bioactive compounds in various plant parts has been well documented (Zhao et al. 2005; Sivanandhan et al. 2014). Identification of regulatory compounds in metabolite biosynthesis is a key component for targeted and increased secondary metabolite production. However, due to limited information about withanolide biosynthesis pathways in *W. coagulans*, it is difficult to identify potential target genes to modulate withanolide biosynthesis process.

With recent advances in sequencing technology, elucidation of biosynthesis pathways and identification of regulatory proteins involved in biosynthesis of secondary metabolites(s) through transcriptome sequencing has become one of the most favored technique (Senthil et al. 2015; Han et al. 2016). Transcriptome sequencing and functional annotation has widely been used to identify regulatory mechanism(s) of important metabolites biosynthesised in different plants such as *Catharanthus roseus* (Shukla et al. 2006), *Chamaemelum nobile* (Liu et al. 2019), *Pueraria lobate* (Wang et al. 2015), *Amaranthus palmeri* (Salas-Perez et al. 2018), *Swertia japonica* (Rai et al. 2016) and *Withania somnifera* (Gupta et al. 2013b). Moreover, expressed sequence tags (ESTs) have also provided a useful tool for gene discovery, particularly in the non-model plants where no reference genome sequences are available (Gupta et al. 2013b).

Transcriptome sequence analysis and pathway mapping studies in *W. somnifera* revealed that withanolides are biosynthesized through isoprenogenesis via mevalonate (MVA) and non-mevalonate (MEP/DOXP) pathways in cytosol and plastids, respectively (Chaurasiya et al. 2012; Dhar et al. 2015). Two isoprene units, namely isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) lead to formation of a triterpenoid, '24-methylene cholesterol' (Chaurasiya et al. 2012; Gupta et al. 2013a). Metabolic divergence from 24-methylene cholesterol into different types of withanolides is primarily mediated through various chemical reactions including cyclization, oxidation, hydroxylation and desaturation (Gupta et al. 2015; Agarwal et al. 2017). However, enzymes catalysing conversion of 24-methylene cholesterol to different types of withanolides is still not clearly known. Studies in past decades have reported significant role of glycosyl transferases, methyl transferases, cytochrome P450s, and transcription factors in conversion of 24-methylene cholesterol to different types of withanolides in leaves and roots of *W. somnifera* (Senthil et al. 2015; Tripathi et al. 2017).

Further, withanolide production has been associated with the morphological differentiation, which corresponds to biosynthesis of tissue-, chemotypes- and species specific withanolides, thereby indicating the variations in their corresponding gene expression profiles (Sharada et al. 2007). Till date, transcriptome studies are only limited to leaf and root tissues of *W. somnifera*, and not *W. coagulans*.

Therefore, this study aimed to perform comprehensive transcriptome profiling of *in vitro* cultured leaf and root tissues of *W. coagulans* with an objective to establish the basic understanding of withanolide biosynthesis pathway(s) in different plant parts and to provide a comparative account of gene expression profiles leading to withanolide biosynthesis. We have reported the total transcriptome data of leaf and root of *W. coagulans* using NGS technology and identified putative genes involved in withanolide biosynthesis for the first time. The EST collection generated in this study has also been analysed for differentially expressed candidate genes which might be involved in withanolide biosynthesis in leaf and root tissues. The transcriptome data was further screened for SSRs, for development of species-specific molecular markers that could facilitate its marker-assisted breeding and selection. The outcome of transcriptome analysis was further validated through comparative withanolide profiling in leaves and roots of micropropagated/*in vitro* raised plants of *W. coagulans*.

Materials And Methods

Explant collection and *in vitro* propagation of *W. coagulans*

Explants (nodal segments) were procured from *W. coagulans* plants grown in the garden located at Manipal University Jaipur, Rajasthan, India. The explants were thoroughly rinsed under running tap water for 5 min and were then washed with 5% (v/v) liquid detergent, Tween®-20 (HiMedia, Mumbai, India) for 5 min followed by washing with running tap water. These explants were then surface sterilised with 0.1% (w/v) HgCl₂ (HiMedia, Mumbai India) for 3 min under aseptic environment, followed by rinsing with sterile distilled water thrice.

Cultures of *W. coagulans* were raised on MS medium (Murashige and Skoog 1962) supplemented with 3% (w/v) sucrose (HiMedia, Mumbai, India), 0.8% (w/v) agar (Plant tissue culture tested; HiMedia, Mumbai, India) and different growth hormones (Duchefa, Netherlands) as per our previously reported protocol (Jain et al. 2009). Precisely, the sterilised nodal explants were inoculated on shoot induction medium (pH 5.8 ± 0.5) composed of MS + 0.5 mgL⁻¹ BA (6-Benzylaminopurine) + 0.5 mgL⁻¹ Kn (Kinetin) and were incubated in plant growth chamber maintained at 25 ± 1°C, 85% relative humidity and 16/8 h photoperiod with 2000 lux luminous intensity.

After 4 wk of incubation, healthy elongated shoots (> 3 cm long) were excised and transferred onto root induction medium containing MS + 0.25 mgL⁻¹ IBA (Indole-3-butyric acid) and were incubated into the plant growth chamber for 4 wks. The rooted plantlets were then used as source of plant material for further transcriptome sequencing and analysis. **Isolation and quality assessment of RNA**

Plant material (leaf and root) was crushed in liq. N₂ using pre-chilled mortar and pestle and followed by total RNA extraction using Xcelgen Plant RNA Kit (Xcleris, Ahmedabad, India) as per manufacturer's protocol. The quality and quantity of the RNA isolated from leaf and root tissues was determined on 1% formaldehyde denaturing agarose gel and Qubit® 2.0 Fluorometer (ThermoFisher Scientific, United States), respectively.

cDNA synthesis, library preparation and quality analysis

Illumina TruSeq Stranded mRNA Library Preparation Kit (Illumina Inc., USA) was used for cDNA synthesis and paired-end library preparation as per instructions given in the manufacturer's manual. Total RNA (~1 µg) was treated with OligodT beads to enrich mRNA fragments which were then purified, fragmented and primed for cDNA synthesis. First and second cDNA strands were synthesised using fragmented mRNA as template. cDNA was subjected to A-tailing, adapter-index ligation and amplification through PCR for library preparation. Quality and quantity of the amplified library was analysed on Bioanalyzer 2100 (Agilent Technologies, USA) using High Sensitivity (HS) DNA Chip (Agilent Technologies, USA).

Cluster generation, sequencing, *de novo* assembly and unigene prediction

cDNA libraries of leaf and root tissues were loaded into Illumina platform (Illumina Inc., USA) for cluster generation and sequencing. The libraries were subjected to paired-end sequencing on Illumina platform (NextSeq500, Illumina Inc., USA) using 2x150 bp chemistry, generating ~ 5 GB data for both leaf and root libraries. To ensure high quality (HQ), the raw reads obtained from Illumina were filtered such that low quality reads and adapter sequences were trimmed using Trimmomatic version 0.36 (Bolger et al. 2014). The HQ reads were then assembled into unique sequences using Trinity Transcriptome Assembler (version 2.1.1) with default (25 k-mers) parameters. Due to unavailability of the reference genome of *W. coagulans*, transcriptome assembly was prepared *de novo*.

The transcripts thus obtained were clustered using CD-HIT (v 4.6.1) package (Fu et al. 2012) for unigenes identification. Short redundant transcripts with 100% overlap and > 90% identity were identified and removed using CD-HIT-EST executable to obtain the non-redundant clustered transcripts termed as "unigenes".

Functional annotation, gene ontology and pathway assignment

The unigenes were annotated against the NCBI non-redundant (Nr) protein (<http://www.ncbi.nlm.nih.gov>), Uniprot (<https://www.uniprot.org>), KOG (<http://www.ncbi.nlm.nih.gov/COG>) and Pfam (<https://pfam.xfam.org>) databases using BLASTx (version 2.2.28+) sequence alignment tool with E-value threshold of 1e-5 such that only top hit for each sequence were recorded. Further, to assign functions to the unigenes, Nr database annotated proteins were subjected to gene ontology (GO) using blast2GO cli (version 1.4.1) software. The unigenes were classified into 3 different domains namely, biological process, cellular component and molecular function, in a way that unigenes associated with similar

functions were assigned the same GO functional group and those having more than one GO term were assigned 2 or 3 GO domains.

Leaf and root unigenes were mapped against KEGG pathway repository and were assigned EC numbers using BLASTx with default bit-score threshold of 60. Category wise distribution of unigenes into metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, brite hierarchies and unclassified processes to identify unigene enriched pathways using KEGG automatic annotation server, KAAS (<https://www.genome.jp/kegg/kaas/>).

Digital gene expression profiling

Transcript level abundance estimates for HQ paired-end reads were computed in terms of transcripts per million (TPM) using RSEM abundance estimation and Bowtie alignment method (Langmead et al. 2009). The TPM values for all the transcripts were summed to obtain gene level TPM score for all enzymes involved in withanolide biosynthesis pathway. Differential gene expression profiling of enzymes involved in terpenoid backbone and steroid biosynthesis pathways along with those belonging to cytochrome P450, methyltransferase and glycosyltransferase gene families was carried out using Web MeV (<http://mev.tm4.org>). Leaf and root TPM values were used in Web MeV and normalised using Deseq method followed by identification and preparation of heat map of differentially expressed genes using LIMMA (Ritchie et al. 2015). Genes with log fold change ≤ -1 or ≥ 1 and q-value ≤ 0.005 were considered differential and heat map of their expression levels was generated. Hierarchical clustering (HCI) of the differentially expressed genes was performed using Euclidean distance matrix and average linkage algorithm, as described previously (Gupta et al. 2013b). Further, a volcano plot of the log fold change vs average expression was also constructed to determine downregulated or upregulated genes in roots with respect to leaves.

Identification of Simple sequence repeats (SSRs)

All the assembled leaf and root transcripts were submitted to microsatellite identification tool, MISA (<https://webblast.ipk-gatersleben.de/misa/index.php>) for identification of SSRs (Beier et al. 2017). The microsatellite identification parameters were defined as minimum 10, 6, 5, 5, 5 and 5 repeats for mono-, di-, tri-, tetra-, penta- and hexa- nucleotide, respectively. For compound microsatellites, 100 bases were set as maximum number between two SSR motifs. The occurrence and distribution frequency of different SSR motifs was also determined using MISA.

Qualitative and Quantitative Profiling of Withanolides

Withanolides were extracted from dried plant material as per previously reported/optimized protocol (Jain et al. 2011).

Qualitative and quantitative profiling of withanolides in leaves and roots was done using HPLC as per our previously reported methods with slight modifications (Jain et al. 2011). 10 ml of each sample was injected through autosampler into 1260 Infinity II HPLC system (Agilent, Germany), and separation was

achieved through reverse phase column (Eclipse Plus C-18, 5 μm , 4.6 x 250 mm) in a solvent gradient of (A) deionized water (Merck Millipore, India) and (B) methanol (Qualigens, Mumbai, India) each containing 0.1% acetic acid (HiMedia, Mumbai, India) at 27°C. The solvent gradient was set as A:B, 60:40 to 25:75 for 0 – 30 min; 10:90 for 31 – 45 min with 0.6 ml min⁻¹ flow rate. The constituent withanolides were detected at 227 nm using UV-Diode array detector (UV-DAD). The characteristic absorbance spectra and retention time of withanolide standards were then used for their identification and quantification in leaf and root extracts (Jain et al. 2011). Five withanolide standards namely, withaferin-A (WF A), withanoside IV (WS IV), withanoside V (WS V), withanolide-B (WL B) and wedelolactone (WDL) procured from Natural Remedies Pvt. Ltd., India were used as markers for profiling.

Results

In vitro propagation of *W. coagulans*

Multiple shoot morphogenesis from the nodal explants with an average of 50 shoots/explant and 100% regeneration efficiency was noted. Callus formation at the cut end and cluster of shoots at each axil of the nodal segments was observed (Fig. 1 ab). Healthy and elongated shoots when transferred to rooting medium resulted in formation of dense network of thin and robust roots (Fig. 1 c).

RNA isolation and cDNA library preparation

Leaf and root tissues yielded 10.4 μg and 6.6 μg total RNA with 694 ng μl^{-1} and 446 ng μl^{-1} qubit concentration, respectively. Leaf and root paired-end libraries were prepared using Illumina library preparation kit and its profiling was done using DNA HS Chip. The average size of leaf and root cDNA libraries were 330 bp and 370 bp, respectively. These libraries were then sequenced on Illumina platform using 2x150 bp PE chemistry.

Sequencing, *de novo* assembly and unigene identification

Sequencing of leaf (WcL) and root (WcR) libraries generated 58,878,942 and 42,683,928 raw reads respectively. The raw read sequences were submitted to NCBI sequence read archive (SRA) and were published with SRX9006975 (leaf) and SRX9006976 (root) accession numbers. Total 8.08 GB (leaf) and 6.35 GB (root) of raw reads were obtained, which were further filtered for HQ reads followed by *de novo* assembly. *De novo* assembly produced 292,074 (WcL) and 16,474 (WcR) transcripts with 410 bp (WcL) and 265.58 bp (WcR) average transcript length. Transcripts length of leaf assembly ranged from 201 bp – 57,369 bp, while that of root assembly ranged from 401 bp – 4637 bp (Table 1). Transcript length distribution pattern indicated that majority of the leaf (~ 61%) and root (~85%) transcripts ranged between 200 – 300 bp (Fig. 2 a).

CD-HIT-EST was used for identification of unigenes from the assembled HQ reads. Total 267,119 and 15,758 unigenes with 392 bp and 265.24 bp average length were identified from leaf and root transcripts, respectively (Table 1). Unigene length distribution pattern for both tissues is represented in Fig. 2 b.

Functional characterization of unigenes

For functional annotation, all the predicted unigenes were aligned against different protein databases using BLASTx program. Among total unigenes annotated (18,8475 – WcL; 9,519 - WcR) using different protein databases, 11,1785 (WcL) and 4,278 (WcR) were annotated using only one database, such that matches for 99% of both WcL and WcR were found only in Nr database, while that for 0.04%, 0.15% and 0.08% of WcL and 0.4%, 0.11%, and 0.09% of WcR were found only in Uniprot, Pfam and KOG databases, respectively (Fig. 3 ab).

The BLAST hits were obtained for 188,174 (70.44%) leaf and 9,448 (59.9%) root unigenes against NCBI Nr protein database, out of which 50.04% (leaf) and 39.65% (root) showed $\geq 99\%$ similarity with known proteins. In contrast, BLAST annotation with Uniprot and Pfam returned matches for 32,976 and 53,908 unigenes, respectively. Lower proportion of leaf unigenes (29.8%) and approximately equivalent proportion of root unigenes (46.2%) were collectively annotated using Uniprot and Pfam (Fig. 3ab). The unigenes were further classified into different protein families based on their annotations derived from Pfam database. Distribution pattern of unigenes across different protein families was determined by identifying number of unigenes coding for similar proteins, and it was revealed that Pkinase, ACR_tran, ABC_tran, Pkinase_Tyr were the most abundant domains in leaf transcriptome, while HSP70, ACR_tran, ABC_tran, GTP_EFTU were most abundant in root transcriptome.

Further, unigenes were annotated against the cluster of eukaryotic orthologous groups (KOG) and were categorised into various KOG functional groups. Out of 40,124 leaf and 2,129 root unigenes identified through KOG database, 4108 (10.2%) leaf and 290 (13.6%) root annotated unigenes belonged to class R (general function prediction only) and class J (translation, ribosomal structure and biogenesis), respectively (Fig. 4 ab).

For functional classification, GO annotation of the unigenes identified through Nr database was performed using blast2GO. A total of 102,029 leaf and 3,506 root unigenes were classified into (sub-categories) 47 and 45 functional (sub-categories) groups, respectively (Fig. 5 ab). From the total GO assigned leaf unigenes, 80.1% of them were classified under 'molecular functions', 73.7% under 'biological processes' and 51.5% under 'cellular components'. Similar GO distribution pattern with molecular functions being most enriched GO class with 83% (2910) of total annotated unigenes followed by 76.5% of unigenes in biological processes and 54.3% of unigenes in cellular components, was also recorded for annotated root unigenes. Majority of the unigenes of both tissues were involved in 2 or more functions, while only 26.2% of leaf and 22.8% of root unigenes were classified just under either one of the GO categories. Leaf unigenes were dominantly involved in 'metabolic process', 'catalytic activity', and 'membrane' functions grouped under biological process, molecular functions and cellular component of GO categories, respectively (Fig. 5 ab). Unlike leaf, majority of the root unigenes classified under 'cellular components' were grouped into a different functional category 'cell', while the most enriched functional groups under 'biological process' and 'metabolic functions' were similar to that of leaf (Fig. 5 ab).

Pathway analysis of unigene libraries

Ortholog assignment and mapping of the unigenes to the biological pathways was performed using KAAS. The active biological pathways in leaf and root tissues were identified by mapping the annotated leaf and root unigene sequences to the reference pathways available in KEGG. In total, 20,927 (WcL) and 2,474 (WcR) unigene sequences were assigned 491 (leaf) and 458 (root) different biological pathways (level 3) which were classified into 7 categories (level 1) and further subdivided into 32 (leaf) and 31 (root) subcategories (level 2). Out of the total pathway assigned unigenes, only 49% of leaf and 38% of root unigene sequences represented single pathway, while majority of the unigenes were assigned to as many as 118 and 72 different pathways/processes in leaves and roots, respectively (Fig. 6ab). Total 12298 (25%) WcL and 1682 (23.14%) WcR unigenes were involved in various types of metabolisms, from which 'carbohydrate metabolism' represented largest cluster of unigenes followed by 'amino acid metabolism' and 'energy metabolism' (Fig. 6 ab). Since the study was aimed to identify putative genes involved in steroidal lactone biosynthesis, unigene clusters representing secondary metabolite biosynthesis were also identified. Among all the secondary metabolite biosynthesis pathways, maximum unigenes clustered for "terpenoid backbone synthesis [PATH:ko00900]", followed by "ubiquinone and other terpenoid-quinone biosynthesis [PATH:ko00130]" and "phenylpropanoid biosynthesis [PATH:ko00940]" with 129, 103 and 80 leaf unigenes and 22, 7 and 7 root unigenes, respectively. BRITE functional hierarchies were assigned to 17,188 leaf and 2286 root unigenes, out of which 2935 WcL and 348 WcR were classified under "protein families: metabolisms".

Identification of withanolide precursor biosynthesis genes

The precursor molecule(s) for withanolide biosynthesis are derived from terpenoid backbone and steroid biosynthesis pathways (Gupta, et al. 2013). Pathway annotations of unigenes were used to identify different enzymes leading to biosynthesis of withanolide precursor in both leaf and root tissues. The withanolide precursor, 24-Methylenecholesterol is derived from Farnesyl-pyrophosphate through a 12 step steroid biosynthesis pathway, which is derived from Acetyl-CoA via mevalonate pathway (7 steps) and D-glyceraldehyde-3-phosphate via MEP/DOXP pathway (8 steps) of the terpenoid backbone biosynthesis (Fig. 7). In total, 16 and 13 enzymes are involved in biosynthesis of precursors derived from terpenoid backbone and steroid biosynthesis pathways, respectively. All the 29 enzymes were present in the leaf transcriptome assembly, while only few enzymes (10) could be annotated from the root assembly (Table 2). Moreover, all the enzymes involved were encoded by more than one unigenes annotated from leaf transcriptome, while in root transcriptome, majority of the enzymes were encoded by a single unigene.

Putative genes related to biosynthesis of withanolides

In *W. somnifera*, biosynthesis of different types of withanolides from its precursor (24-methylenecholesterol) is mediated through members of cytochrome P450 (CYP 450), glycosyltransferase (GT) and methyltransferase (MT) gene families (Senthil et al. 2010; Dhar et al. 2015). With the help of functional annotation using different databases total 3642 leaf and 185 root unigenes as members of CYP450, GT and MT gene families were identified, out of which majority of the members were tissue specific and only few (8 - CYP450; 16 – GT; 56 - MT) were present in both leaves and roots. Of the total

unigenes mapped against these gene families, 'cytochrome P450 CYP2 subfamily', 'cytochrome P450 CYP4 CYP19 CYP26 subfamilies', 'glycosyl transferase', 'glycosyl transferase family 1', 'glycosyl transferase family (trehalose-6-phosphate synthase)', 'methyltransferase', 'SAM dependent methyltransferase', 'serine hydroxymethyltransferase' and 'DNA (cytosine-5)-methyltransferase 1' were most abundant in leaves, while 'cytochrome P450', 'cytochrome P450 CYP4 CYP19 CYP26 subfamilies', 'glycosyl transferase', 'glycosyl transferase family 1', 'dolichyl-diphosphooligosaccharide—protein glycosyltransferase', 'SAM dependent methyltransferase', 'methayltransferase', 'RNA methyltrasnferase' and 'glycine hydroxymethylatransferase' were the major class of CYP540, GT and MT gene families.

Differential gene expression analysis

The abundance estimates of leaf and root unigenes were calculated in the form of transcripts per million (TPM), which were then compared for identification of differentially expressed genes. Digital expression profile for the genes commonly expressed in both leaf and root tissues showed significant variations in the expression levels of (putative) genes involved in withanolide precursor biosynthesis and conversion of precursor to different withanolides with average expression ranging from 0.1 – 7 and 0.01 – 9.2, respectively (Fig. 8 ab). Among the commonly expressed genes in both tissues, only 6 withanolide precursor biosynthesis genes and 53 putative withanolide biosynthesis genes expressed differentially between leaf and root tissues. Out of total 59 differentially expressed genes, 34 were upregulated and 25 were downregulated (Fig. 8 cd). Hierarchical clustering further revealed the co-expression pattern of the different genes involved in precursor biosynthesis and conversion of precursor to withanolides (Fig. 8 ab).

SSR markers

Total 28852 leaf and 886 root SSRs were identified through screening of 297,412 leaf and 16,474 root transcripts, respectively. Total 24841 (8.3%) leaf and 772 (4.7%) root transcripts accounted for the total SSRs in both tissues, out of which 11.6% (leaf) and 9.5% (root) sequences harboured more than one SSR. Comparatively higher proportion of root SSRs (10.5%) than that of leaf SSRs (6.7%) were present in compound formation (Table 3). Leaf SSRs were distributed across 6 repeat type classes, while hexa-nucleotide repeats were absent in root transcript sequences. Mono-nucleotide repeats were most abundant (22791 – leaf; 676 – root) in both tissues, whereas hexa-nucleotide repeats (64) in leaf and tetra-nucleotide repeats (4) in root were less (Fig. 9). From the 82 types of SSRs identified from the leaf assembly, A/T, AG/CT, AAG/CTT, AAAG/CTTT, AATGG/ATTCC, AATGGT/ACCATT were most mono-, di-, tri-, tetra-, penta- and hexa- nucleotide repeats. Similarly, 23 different types of repeat units contributed to the SSR pool of the root library, out of which A/T, AC/GT, AAG/CTT, AAAC/GTTT, AAGG/CCTT and AATGG/ATTCC were most common in their respective repeat classes.

Validation of withanolide accumulation through qualitative and quantitative profiling

HPLC analysis revealed significant variations in the withanolide profile of leaves and roots (Figure 10). More than 50 different types of withanolides were detected in leaves, while less than 10 different types of withanolides were present in roots. The relative abundance of various withanolides was significantly

higher in leaves than that of roots (Figure 10). Moreover, none of the 5 withanolide standards were present in root extracts, while WL B (724.83 ng/ μ l of extract), WF A (8.63 ng/ μ l of extract) and WDL (9.74 ng/ μ l of extract) were present in leaf extracts.

Discussion

Withania coagulans has popularly been used as a natural coagulant for a long time and has gained unprecedented attention over past few decades due to its bioactive secondary metabolites, *i.e.*, withanolides. Withanolides are the principle bioactive components of genus *Withania* which possess diverse array of pharmacological properties including immunomodulatory, neuroprotective, cardioprotective, antitumor, *etc* (Jain et al. 2012). Limitations in propagation strategies and variations in its withanolide profile of the plant due to stresses are major concerns in developing novel medicinal formulations. Plant tissue culture strategies have emerged as an efficient tool not only for *ex situ* conservation and mass multiplication of plants but also for enhanced biosynthesis of bioactive metabolites (Jain et al. 2009). Further, biosynthesis of some novel withanolides has also been reported in the *in vitro* cultures of *W. coagulans* (Jain et al. 2011), but due to lack of information about withanolide biosynthesis pathway(s) the underlying gene regulatory phenomenon is difficult to predict. There are various reports on characterization of withanolide biosynthesis pathway(s) genes in different parts of both field grown and micropropagated plants of *W. somnifera* (Senthil et al. 2015; Mishra et al. 2016; Mishra et al. 2020), yet the information about specific enzymes which lead to formation of different withanolides is still incomplete and is also subjected to vary at both species and tissue level.

In the present study, whole transcriptome sequencing and analysis of *in vitro* raised leaf and root tissues of *W. coagulans* was performed to identify the withanolide biosynthesis pathway and putative genes involved in this process. The *in vitro* cultures of the plant were established as per our previously optimized protocol for clonal mass multiplication (Jain et al. 2009; Jain et al. 2011). As per the report in *W. somnifera* (Gupta et al. 2013b), leaf and root samples from the *in vitro* raised plants were harvested for transcriptome analysis to ensure that the transcript libraries represent all the genes involved in withanolide biosynthesis and could indicate if the biosynthesis of withanolides is *de novo* in both tissues or imported from one tissue to other.

Transcriptome sequencing using Illumina second generation sequencing platform can generate 100x more reads with at least 5x higher read depth than that generated through 454-based sequencers (Xiao et al. 2013). The proportion of transcripts annotated for functional assignment and number of full-length unigenes produced is also relatively higher for Illumina based platforms (Xiao et al. 2013). Therefore, this techniques has been widely used to generate reference libraries and study specialized metabolic pathways through *de novo* assembly and annotation, in various non-model plants including *Asparagus racemosus* (Upadhyay et al. 2014), *Pueraria lobata* (Wang et al. 2015), *Withania somnifera* (Senthil et al. 2015), *Swertia japonica* (Rai et al. 2016), *Chamaemelum nobile* (Liu et al. 2019), *Plumbago zeylanica* (Sundari et al. 2020), *etc*. Based on these recent reports, RNA sequencing of leaf and root tissues harvested from micropropagated plants of *W. coagulans* using Illumina NextSeq sequencing platform

was done. Total 14.43 GB data with 58,878,942 leaf and 42,683,928 root HQ reads were generated, which were then submitted to NCBI's SRA database. Low quality reads and adapter sequences may produce misassemblies or truncated contigs by interfering with the transcriptome assembly (Upadhyay et al. 2014), therefore filtration of the raw reads was performed prior to transcriptome assembly. Since no reference genome was available for *W. coagulans*, the HQ reads were assembled *de novo* into 292,074 leaf and 16,474 root transcripts, out of which 282,877 (267,119 – leaf; 15,758 - root) unigenes were identified. From the total unigenes identified, matches in Nr, Uniprot, Pfam and KOG databases were found for 188,475 (70.6%) leaf and 9,519 (60.4%) root unigenes, From the total annotated unigenes, 40.7% (76,669) leaf and 55% (5,241) root unigenes were mapped in 2 or more databases, indicating extensive coverage of the transcriptome assembly. Upadhyay et al. (2014) reported that unigenes without any matches either represent the untranslated regions and non-coding RNA, or might be due to assembly mistakes. Transcriptome sequencing performed using 454 GC-FLX Titanium platform produced comparatively lesser number of contigs and singletons, out of which only ~40% were annotated against Nr protein database (Gupta et al. 2013b), suggesting lesser coverage of the assembly generated through 454 sequencing platforms over Illumina platforms.

Further, 35% (leaf) and 33% (root) of the total annotated unigenes were assigned to functional domains (5,872 WcL, 1,454 WR) and functional groups (25) using Pfam and KOG databases, respectively. Assignment of these unigenes to large number of functional groups indicates the diversity of the genes encoded in leaf and root transcriptome of the plant. The rich diversity in the gene functions of *W. coagulans* transcriptome was also corroborated by the GO annotations assigned to each unigene mapped using Nr database. Total 102,029 annotated leaf and 3,506 root unigenes were assigned to biological process (3815 – WcL, 1087 - WcR), cellular component (375 – WcL, 493 - WcR) and molecular functions (4493 – WcL, 1237 - WcR), thereby highlighting the diversity of the *W. coagulans* genes identified from its leaf and root transcriptome assembly (Gupta et al. 2013b; Upadhyay et al. 2014). Approximately 74% of GO annotated unigenes were involved in more than one type of functions/processes, probably due to the multiple roles of the functional groups represented. The dominant functions as identified from GO annotations of WcL and WcR did not show any significant similarity with that of *W. somnifera*, thus despite of their close phylogenetic relation and shared morphological & pharmaceutical properties, *W. somnifera* could not be used as a reference plant for transcriptome assembly and annotation of the counterpart species.

The annotated unigenes were mapped to various KEGG biological pathways using KAAS. Total 2,274 and 709 enzymes functional in 491 and 458 pathways of leaves and roots, respectively were identified. Most of the enzymes were encoded by two or more unigenes, as these unigenes are either fragments of a single transcript encoding the enzyme or different members of an enzyme family. Similar studies on *W. somnifera* have identified 124 and 139 functional pathways in tissues of field grown and *in vitro* raised plants (Gupta et al. 2013b; Senthil et al. 2015). Identification of potential unigenes encoding for regulatory enzymes involved in pathways of interest could aid in metabolic engineering of pathways for targeted biosynthesis. From the total annotated unigenes, 1566 (leaf) and 145 (root) unigenes were assigned to 'unclassified metabolic', 'genetic information processing' and 'signaling & cellular process'

pathways, while 603 (leaf) and 34 (root) unigenes were 'poorly characterized'. These unknown and partial unigenes can further be characterized to identify novel genes involved in biosynthesis of different types of withanolides in *W. coagulans* tissues.

24-methylenecholesterol has been identified as a precursor for biosynthesis of withanolides and is derived through steroid biosynthesis and terpenoid backbone biosynthesis (MVA and DOXP) pathways (Chaurasiya et al. 2012; Gupta et al. 2013b). Since *W. coagulans* is closely related to *W. somnifera*, in this study all the unigenes involved in steroid and terpenoid backbone biosynthesis pathways of micropropagated leaves and roots are identified. Unlike in *W. somnifera*, all the 29 enzymes involved in precursor biosynthesis were identified only in leaves, while only few were identified in roots. As per recent reports, enzymes mediating the conversion of 5-dehydroepisterol to 24-methylene cholesterol (DWF5) & 24-methylenecholesterol to 24-methylidesmosterol (sterol isomerase, DWF1), and cyclisation of 2,3-oxidosqualene into cycloartenol (cycloartenol synthase, CAS), are critical for *de novo* withanolide biosynthesis in both leaf and root tissues of *W. somnifera* (Gupta et al. 2015; Mishra et al. 2016; Knoch et al. 2018). In the present study, all these regulatory enzymes (DWF5, DWF1, CAS, etc.) were identified only in leaves, while only DWF5 was present in roots also. This clearly indicated that *de novo* withanolide biosynthesis pathway does not occur in *in vitro* cultured roots. Agarwal et al. (2017) reported that DWF5 and DWF1 are localized in endoplasmic reticulum and are important for biosynthesis of tissue-specific withanolides. Withanolides are biosynthesized in *W. somnifera*, independently through *de novo* pathway in leaf and root tissues of the plant (Senthil et al. 2010; Gupta et al. 2013b). Among the enzymes commonly present in *W. coagulans* leaves and roots, ~43% of them showed differential expression, such that enzyme catalysing the conversion of acetyl Co-A to acetoacetyl Co-A [EC:2.3.1.9] showed highest average expression. Enzyme(s) HMGR (hydroxymethylglutaryl-CoA reductase) was downregulated, while farnesyl diphosphate synthase and 1-deoxy-D-xylulose-5-phosphate synthase were upregulated *w.r.t* that in leaves. Senthil et al. (2015) reported high HMGR expression in roots of *W. somnifera* after 30 days of *in vitro* culture and concluded that the expression level in the cultures varies in a time-dependent manner. Further, co-transformation of a fungal elicitor protein in hairy roots of *W. somnifera* was reported by Mishra et al. (2016), which downregulates HMGR and FPPS enzymes, and reduces withanolide content. Therefore, lower expression levels of HMGR in *in vitro* roots of *W. coagulans* can be attributed to the absence of other pathway enzymes leading to precursor formation. Since, it is more likely that withanolide biosynthesis in *W. coagulans* takes place in salvage mode, potential intermediate metabolites which serve as precursor feed can be predicted depending upon the expression levels and clustering pattern of these enzymes.

Two chemical reactions involving δ -lactonisation between C-22 & C-26 and hydroxylation at C-22 of 24-methylenecholesterol have been rendered essential for biosynthesis of withanolides (Dhar et al. 2015). However, tissue specific divergence of withanolides is mediated through secondary conversions including addition of different side chains, oxidation/reduction, hydroxylation and glycosylation reactions, etc. (Chaurasiya et al. 2012; Gupta et al. 2013b; Gupta et al. 2015). 24-methylenecholesterol has also been identified as key precursor for biosynthesis of another class of steroid derivatives called 'brassinolides'. The precursor is converted into brassinolides through a series of oxidation and hydroxylation reactions,

catalysed by cytochrome P450 enzymes (Bishop 2007; Ohnishi et al. 2009). Cytochrome P450s (CYP or P450) catalyse hydroxylation, oxidative demethylation, desaturation, epoxidation, desaturation, oxidative rearrangement of carbon skeleton and oxidative C-C bond cleavage reactions in various plant cellular and metabolic processes (Rana et al. 2014). Srivastava et al. (2015) reported that P450s belonging to CYP83B1 (WSCYP93Id) & CYP734A1 (WSCYP9Sm) families and CYP71 (WSCYP734B) & CYP72 (WSCYP734) clans account for synthesis of chemotype specific withanolides in *W. somnifera* and suggested involvement of WSCYP93Id and WSCYP9Sm in oxygenation reactions of plants, which account for synthesis of different metabolites. Further, variations in expression profiles of two-A type P450s (*WsCYP98A* and *WsCYP76A*) and two paralogs of cytochrome P450 reductase has also been associated with differential withanolide biosynthesis in *W. somnifera* (Rana et al. 2014). Glycosylation commonly catalysed by glycosyltransferases (GTs), is a modification reaction that is usually related with biosynthesis of secondary metabolites (Dhar et al. 2015). Singh et al. (2016) reported differential biosynthesis of withaferin A, withanolide A and withanoside A is regulated by the expression of sterol glycosyltransferases (SGTs) through a positive feedback mechanism in *W. somnifera*. Further, role of sterol methyltransferase 1 (SMT1) in channelling the intermediates towards first committed step of withanolide biosynthesis in *W. somnifera* has also been reported recently (Pal et al. 2019). In this study, unigenes encoding for P450, MT and GT in leaf assembly (3642) were significantly higher than those in root assembly (185), such that only 8, 16 and 56 members of P450, GT and MT, respectively were common in both tissues. Among common members, only 5 P450s, 7 GTs and 41 MTs were differentially expressed in leaves and roots, out of which 3 P450s, 2 GTs and 18 MTs were downregulated in roots w.r.t those in leaves. Presence of large number of tissue-specific genes and their differential expression indicates diversity of withanolide between both the tissues. Since these putative genes regulate conversion of precursor to different types of withanolides, availability of lesser number of unigenes encoding these genes can be considered as an indication of comparatively less diverse withanolide profile. The putative unigenes thus identified can further be used to elucidate withanolide biosynthesis pathway and mechanism(s) involved in tissue-specific synthesis of withanolides in leaves and roots of *W. coagulans*. This tissue-specific differential accumulation of withanolides was also confirmed through HPLC analysis. Some of the important withanolides including WF A, WL B and WDL were specifically present only in leaves, affirming their tissue-specific accumulation and biosynthesis. Though differential accumulation of withanolides in different tissues of *W. somnifera* has been reported in various studies (Senthil et al. 2015), but same has been reported for the first time in *W. coagulans* in this study. Relatively very low abundance of withanolides in roots and their exclusivity to the tissue further supports our proposed hypothesis of *de novo* and salvage biosynthesis in leaves and roots, respectively.

A comparative analysis of the withanolide biosynthesis pathway and their accumulation mechanism in *W. coagulans* and *W. somnifera* plant parts can also be performed to identify the enzymes essential for *de novo* and salvage synthesis of withanolides and to reveal their phylogenetic origin. Such studies would enable to develop strategies for targeted withanolide biosynthesis through metabolic pathway engineering at gene level.

The leaf and root unigene sequences were also screened for identification of simple sequence repeats (SSRs) which can be used for development of species-, chemotype- and tissue- specific molecular markers. From the total SSRs identified, ~ 17% of them were present in compound formation and ~ 21% of the unigenes harboured more than one SSR. In this study, mononucleotides were the most abundant SSRs in both leaf and root transcriptome assembly. On the contrary, trinucleotide repeats have been identified as most frequently occurring SSRs in leaf and root transcriptome assembly of *W. somnifera* (Gupta et al. 2013b). The differences in marker distribution pattern can be used to design species specific molecular markers for *W. coagulans* and *W. somnifera*.

Conclusion

This study is the first to establish the transcriptome database for *W. coagulans* through NGS technology. Pathway annotation indicated that leaves of *in vitro* raised plantlets possess *de novo* withanogenesis potential. Absence of certain withanogenesis enzymes in the roots revealed that the roots lack *de novo* withanogenic competence and withanolides are biosynthesised through one or more salvage pathway(s). The higher diversity in putative gene families of leaf (108) transcriptome than that of root (81), further corresponded to higher metabolic diversity of withanolides in leaves than in roots. This information can be fruitful in engineering the biosynthesis pathway for enhanced production of bioactive withanolides. Moreover, the metabolic distinctness among the two important medicinal species of *Withania* has also been established for the first time in this study. The SSRs identified can be used to develop chemotype- and species- specific markers, which in turn will help in identification of genotypes rich in specific withanolides and distinction among different varieties of *W. somnifera* and *W. coagulans*. Therefore, sequencing and analysis of *W. coagulans* transcriptome will be useful for further research and development on marker-assisted selection & breeding programs for development of medicinally elite varieties of *Withania*. Additionally, this study has also opened new gateways for establishment of plant cell factories for production of important withanolides at industrial scale.

Declarations

Funding : Enhanced seed grant (Sanction No.

EF/2017-18/QE04-03) by Manipal University Jaipur

Conflicts of interest/Competing interests : Authors have no conflict of interest

Ethics approval : Not applicable

Consent to participate : Not applicable

Consent for publication : Not applicable

Availability of data and material :

Code availability : Not applicable

Author's contributions

Conceptualization RJ; experimental design and methodology RJ; perform experimental studies SG; statistics and data analysis MKB, KS; funding acquisition RJ; writing - manuscript draft preparation SG; writing - review and editing SK; supervision and validation of results SLK. All authors have read and agreed to the published version of the manuscript.

Acknowledgement

Financial support in the form of endowment seed grant no. EF/2017-18/QE04-03 from Manipal University Jaipur is gratefully acknowledged. Senior research fellowship awarded by Indian Council of Medical Research (ICMR, New Delhi) to Swati Gupta is also deeply acknowledged.

References

- Agarwal A, Gupta P, Singh D, Dhar Y, Chandra D, Trivedi P (2017) Comprehensive assessment of the genes involved in withanolide biosynthesis from *Withania somnifera*: chemotype-specific and elicitor-responsive expression. *Funct Integr Genomics* 17(4):477–490. doi:<https://doi.org/10.1007/s10142-017-0548-x>
- Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33(16):2583–2585. doi:<https://doi.org/10.1093/bioinformatics/btx198>
- Bhandari MM (1995) Flora of the Indian desert. MPS Repros Jodhpur, India
- Bishop G (2007) Refining the plant steroid hormone biosynthesis pathway. *Trends Plant Sci* 12(9):377–380. doi:<https://doi.org/10.1016/j.tplants.2007.07.001>
- Bolger A, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120. doi:<https://doi.org/10.1093/bioinformatics/btu170>
- Chaurasiya N, Sangwan N, Sabir F, Misra L, Sangwan R (2012) Withanolide biosynthesis recruits both mevalonate and DOXP pathways of isoprenogenesis in Ashwagandha *Withania somnifera* L.(Dunal). *Plant Cell Rep* 31(10):1889–1897. doi:<https://doi.org/10.1007/s00299-012-1302-4>
- Dhar N, Razdan S, Rana S, Bhat W, Vishwakarma R, Lattoo S (2015) A decade of molecular understanding of withanolide biosynthesis and *in vitro* studies in *Withania somnifera* (L.) Dunal: prospects and perspectives for pathway engineering. *Front Plant Sci* 6:1031. doi:<https://doi.org/10.3389/fpls.2015.01031>
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152. doi:<https://doi.org/10.1093/bioinformatics/bts565>

- Gupta P, Agarwal A, Akhtar N, Sangwan R, Singh S, Trivedi P (2013a) Cloning and characterization of 2-C-methyl-D-erythritol-4-phosphate pathway genes for isoprenoid biosynthesis from Indian ginseng, *Withania somnifera*. *Protoplasma* 250(1):285–295. doi:<https://doi.org/10.1007/s00709-012-0410-x>
- Gupta P, Goel R, Agarwal A, Asif MH, Sangwan N, Sangwan R, Trivedi P (2015) Comparative transcriptome analysis of different chemotypes elucidates withanolide biosynthesis pathway from medicinal plant *Withania somnifera*. *Sci Rep* 5(1):18611. doi:<https://doi.org/10.1038/srep18611>
- Gupta P, Goel R, Pathak S, Srivastava A, Singh S, Sangwan R, Asif M, Trivedi P (2013b) De novo assembly, functional annotation and comparative analysis of *Withania somnifera* leaf and root transcriptomes to identify putative genes involved in the withanolides biosynthesis. *PLoS One* 8(5):e62714. doi:<https://doi.org/10.1371/journal.pone.0062714>
- Han R, Rai A, Nakamura M, Suzuki H, Takahashi H, Yamazaki M, Saito K (2016) De novo deep transcriptome analysis of medicinal plants for gene discovery in biosynthesis of plant natural products *Methods Enzymol.* vol 576. Elsevier, p 19–45
- Jain R, Kachhwaha S, Kothari SL (2012) Phytochemistry, pharmacology, and biotechnology of *Withania somnifera* and *Withania coagulans*: A review. *J Med Plants Res* 6(41):5388–5399. doi:<https://doi.org/10.5897/JMPR12.704>
- Jain R, Sinha A, Jain D, Kachhwaha S, Kothari SL (2011) Adventitious shoot regeneration and *in vitro* biosynthesis of steroidal lactones in *Withania coagulans* (Stocks) Dunal. *Plant Cell Tiss Org Cult* 105(1):135–140. doi:<https://doi.org/10.1007/s11240-010-9840-3>
- Jain R, Sinha A, Kachhwaha S, Kothari SL (2009) Micropropagation of *Withania coagulans* (Stocks) Dunal: A critically endangered medicinal herb. *J Plant Biochem Biotechnol* 18(2):249–252. doi:<https://doi.org/10.1007/BF03263330>
- Knoch E, Sugawara S, Mori T, Poulsen C, Fukushima A, Harholt J, Fujimoto Y, Umemoto N, Saito K (2018) Third DWF1 paralog in Solanaceae, sterol Δ 24-isomerase, branches withanolide biosynthesis from the general phytosterol pathway. *Proc Natl Acad Sci* 115(34):E8096–E8103. doi:<https://doi.org/10.1073/pnas.1807482115>
- Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25. doi:<https://doi.org/10.1186/gb-2009-10-3-r25>
- Liu X, Wang X, Chen Z, Ye J, Liao Y, Zhang W, Chang J, Xu F (2019) De novo assembly and comparative transcriptome analysis: novel insights into terpenoid biosynthesis in *Chamaemelum nobile* L. *Plant Cell Rep* 38(1):101–116. doi:<https://doi.org/10.1007/s00299-018-2352-z>

- Maurya R, Akanksha, Jayendra (2010) Chemistry and pharmacology of *Withania coagulans*: an Ayurvedic remedy. *J Pharm Pharmacol* 62(2):153–160. doi:<https://doi.org/10.1211/jpp.62.02.0001>
- Mishra B, Bose S, Sangwan N (2020) Comparative investigation of therapeutic plant *Withania somnifera* for yield, productivity, withanolide contents, and expression of pathway genes during contrasting seasons. *Ind Crop Prod* 154:112508. doi:<https://doi.org/10.1016/j.indcrop.2020.112508>
- Mishra S, Bansal S, Mishra B, Sangwan R, Jadaun J, Sangwan N (2016) RNAi and homologous over-expression based functional approaches reveal triterpenoid synthase gene-cycloartenol synthase is involved in downstream withanolide biosynthesis in *Withania somnifera*. *PLoS One* 11(2):e0149691. doi:<https://doi.org/10.1371/journal.pone.0149691>
- Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco cultures. *Physiol Plant* 15(3):473–497. doi:<https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>
- Nagella P, Murthy HN (2010) Establishment of cell suspension cultures of *Withania somnifera* for the production of withanolide A. *Biores Technol* 101:6735–6739. doi:<https://doi.org/10.1016/j.biortech.2010.03.078>
- Ohnishi T, Yokota T, Mizutani M (2009) Insights into the function and evolution of P450s in plant steroid metabolism. *Phytochemistry* 70(17–18):1918–1929. doi:<https://doi.org/10.1016/j.phytochem.2009.09.015>
- Pal S, Rastogi S, Nagegowda D, Gupta M, Shasany A, Chanotiya C (2019) RNAi of sterol methyl transferase1 reveals its direct role in diverting intermediates towards withanolide/phytosterol biosynthesis in *Withania somnifera*. *Plant Cell Physiol* 60(3):672–686. doi:<https://doi.org/10.1093/pcp/pcy237>
- Rai A, Nakamura M, Takahashi H, Suzuki H, Saito K, Yamazaki M (2016) High-throughput sequencing and de novo transcriptome assembly of *Swertia japonica* to identify genes involved in the biosynthesis of therapeutic metabolites. *Plant Cell Rep* 35(10):2091–2111. doi:<https://doi.org/10.1007/s00299-016-2021-z>
- Rana S, Bhat W, Dhar N, Pandith S, Razdan S, Vishwakarma R, Lattoo S (2014) Molecular characterization of two A-type P450s, WsCYP98A and WsCYP76A from *Withania somnifera* (L.) Dunal: expression analysis and withanolide accumulation in response to exogenous elicitations. *BMC Biotechnol* 14(1):89. doi:<https://doi.org/10.1186/s12896-014-0089-5>
- Ritchie M, Phipson B, Wu D, Hu Y, Law C, Shi W, Smyth G (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47–e47. doi:<https://doi.org/10.1093/nar/gkv007>

- Salas-Perez R, Sasaki C, Noorai R, Srivastava S, Lawton-Rauh A, Nichols R, Roma-Burgos N (2018) RNA-Seq transcriptome analysis of *Amaranthus palmeri* with differential tolerance to glufosinate herbicide. PLoS One 13(4):e0195488. doi:<https://doi.org/10.1371/journal.pone.0195488>
- Sangwan RS, Chaurasiya ND, Lal P, Misra L, Uniyal GC, Tuli R, Sangwan NS (2007) Withanolide A biogenesis in *in vitro* shoot cultures of ashwagandha (*Withania somnifera* Dunal): a main medicinal plant in ayurveda. Chem Pharm Bull (Tokyo) 55(9):1371–1375. doi:<https://doi.org/10.1248/cpb.55.1371>
- Senthil K, Jayakodi M, Thirugnanasambantham P, Lee S, Duraisamy P, Purushotham P, Rajasekaran K, Charles S, Roy I, Nagappan A (2015) Transcriptome analysis reveals *in vitro* cultured *Withania somnifera* leaf and root tissues as a promising source for targeted withanolide biosynthesis. BMC Genom 16(1):14–30. doi:<https://doi.org/10.1186/s12864-015-1214-0>
- Senthil K, Wasnik N, Kim Y, Yang D (2010) Generation and analysis of expressed sequence tags from leaf and root of *Withania somnifera* (Ashwagandha). Mol Biol Rep 37(2):893–902. doi:<https://doi.org/10.1007/s11033-009-9696-y>
- Sharada M, Ahuja A, Suri KA, Vij SP, Khajuria RK, Verma V, Kumar A (2007) Withanolide production by *in vitro* cultures of *Withania somnifera* and its association with differentiation. Biol Plant 51(1):161–164. doi:<https://doi.org/10.1007/s10535-007-0031-y>
- Shukla A, Shasany A, Gupta M, Khanuja S (2006) Transcriptome analysis in *Catharanthus roseus* leaves and roots for comparative terpenoid indole alkaloid profiles. J Exp Bot 57(14):3921–3932. doi:<https://doi.org/10.1093/jxb/erl146>
- Singh G, Tiwari M, Singh S, Singh S, Trivedi P, Misra P (2016) Silencing of sterol glycosyltransferases modulates the withanolide biosynthesis and leads to compromised basal immunity of *Withania somnifera*. Sci Rep 6(1):25562. doi:<https://doi.org/10.1038/srep25562>
- Sivanandhan G, Selvaraj N, Ganapathi A, Manickavasagam M (2014) Enhanced biosynthesis of withanolides by elicitation and precursor feeding in cell suspension culture of *Withania somnifera* (L.) Dunal in shake-flask culture and bioreactor. PLoS One 9(8):e104005. doi:<https://doi.org/10.1371/journal.pone.0104005>
- Srivastava S, Sangwan R, Tripathi S, Mishra B, Narnoliya L, Misra L, Sangwan N (2015) Light and auxin responsive cytochrome P450s from *Withania somnifera* Dunal: cloning, expression and molecular modelling of two pairs of homologue genes with differential regulation. Protoplasma 252(6):1421–1437. doi:<https://doi.org/10.1007/s00709-015-0766-9>
- Sundari B, Budhwar R, Dwarakanath B, Thyagarajan S (2020) De novo transcriptome analysis unravels tissue-specific expression of candidate genes involved in major secondary metabolite biosynthetic pathways of *Plumbago zeylanica*: implication for pharmacological potential. 3 Biotech 10:271. doi:<https://doi.org/10.1007/s13205-020-02263-9>

Tripathi S, Sangwan R, Narnoliya L, Srivastava Y, Mishra B, Sangwan N (2017) Transcription factor repertoire in Ashwagandha (*Withania somnifera*) through analytics of transcriptomic resources: Insights into regulation of development and withanolide metabolism. *Sci Rep* 7(1):16649. doi:<https://doi.org/10.1038/s41598-017-14657-6>

Upadhyay S, Phukan U, Mishra S, Shukla R (2014) De novo leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in *Asparagus racemosus*. *BMC Genom* 15(1):746. doi:<https://doi.org/10.1186/1471-2164-15-746>

Wang X, Li S, Li J, Li C, Zhang Y (2015) De novo transcriptome sequencing in *Pueraria lobata* to identify putative genes involved in isoflavones biosynthesis. *Plant Cell Rep* 34(5):733–743. doi:<https://doi.org/10.1007/s00299-014-1733-1>

Xiao M, Zhang Y, Chen X, Lee E, Barber C, Chakrabarty R, Desgagné-Penix I, Haslam T, Kim Y, Liu E (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J Biotechnol* 166(3):122–134. doi:<https://doi.org/10.1016/j.jbiotec.2013.04.004>

Zhao J, Davis LC, Verpoorte R (2005) Elicitor signal transduction leading to production of plant secondary metabolites. *Biotechnol Adv* 23(4):283–333. doi:<https://doi.org/10.1016/j.biotechadv.2005.01.003>

Figures

Figure 1



Figure 1

In vitro propagation of *W. coagulans*. (a-b) Multiple shoot morphogenesis from nodal explants cultured on shoot proliferation medium, (c) Rooting of the in vitro regenerated shoots

Figure 2

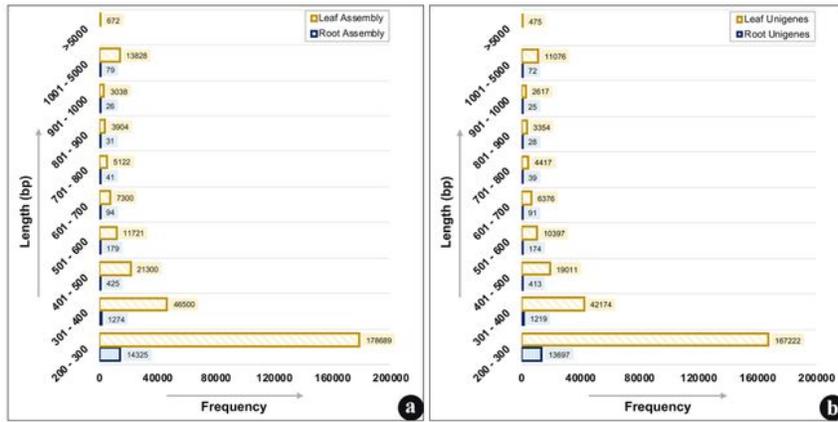


Figure 2

Length distribution pattern of (a) transcripts, and (b) unigenes of WcL and WcR

Figure 3

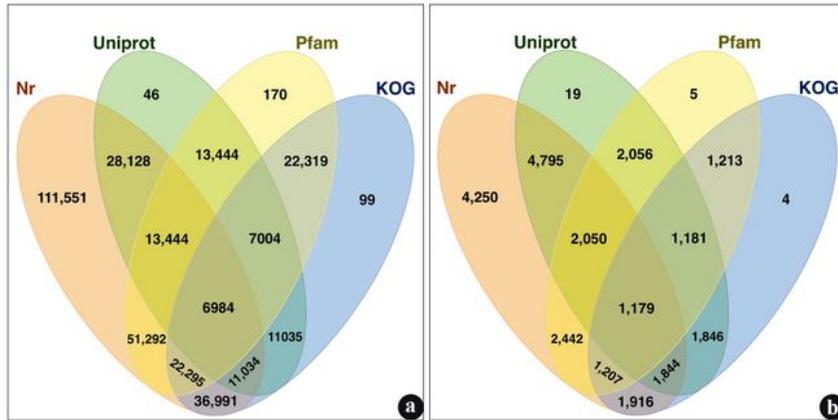


Figure 3

Annotation pattern of (a) WcL, and (b) WcR unigenes against different protein databases

Figure 4

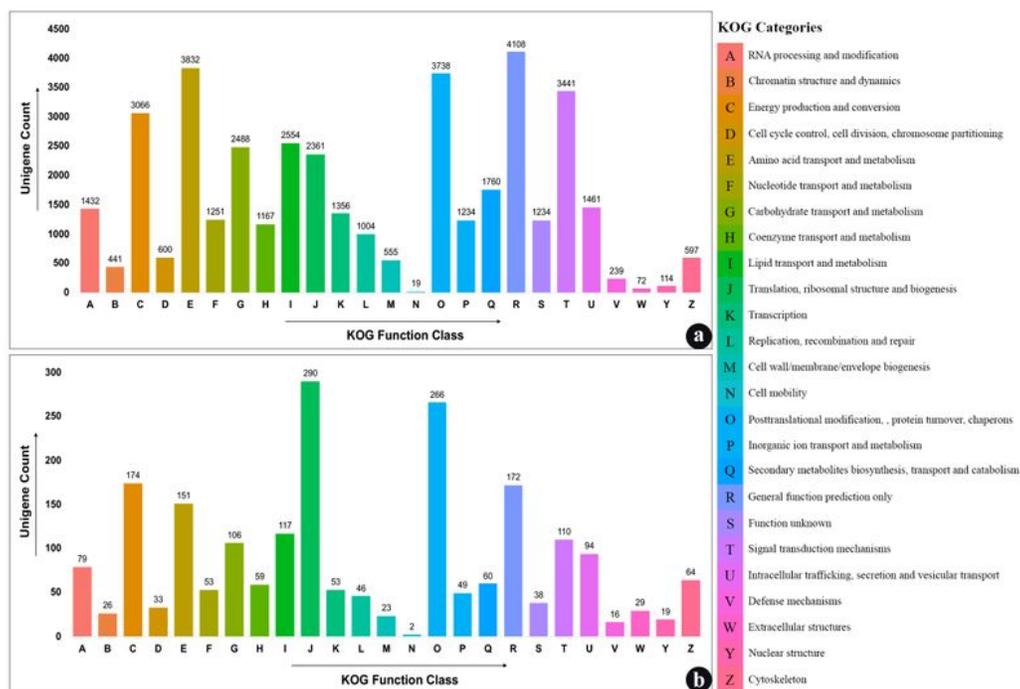


Figure 4

KOG classification and distribution pattern of (a) WcL, and (b) WcR unigenes

Figure 5

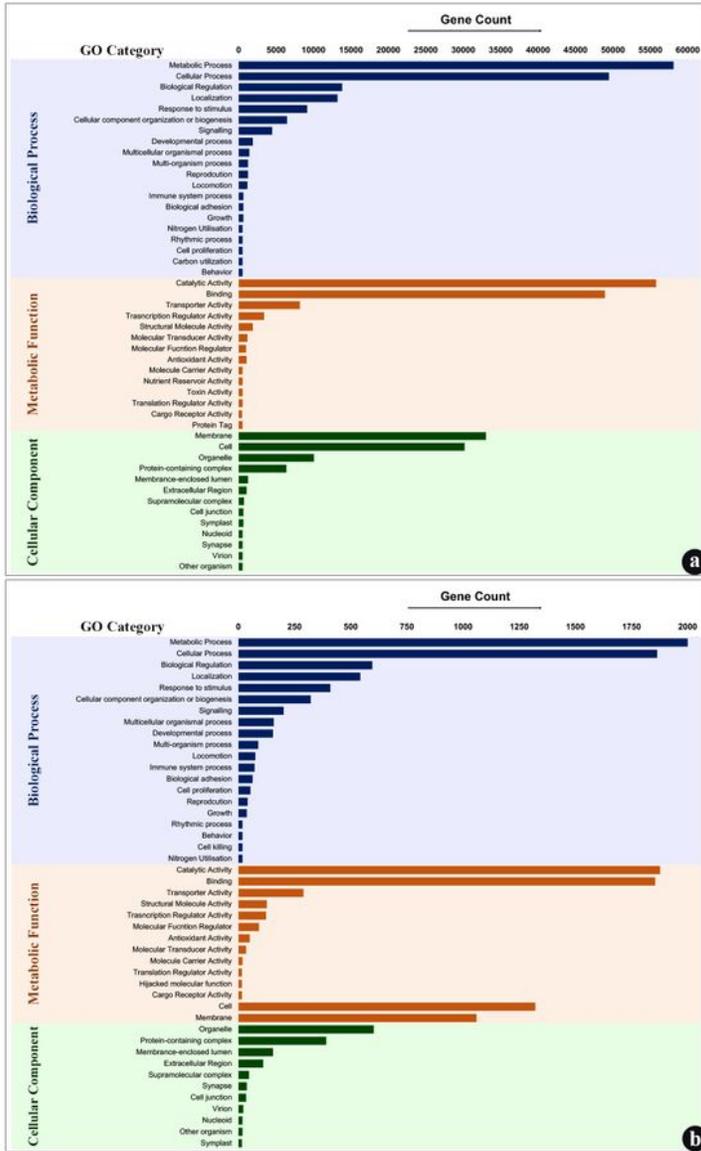


Figure 5

Gene ontology of Nr annotated unigenes of (a) WcL, and (b) WcR

Figure 6

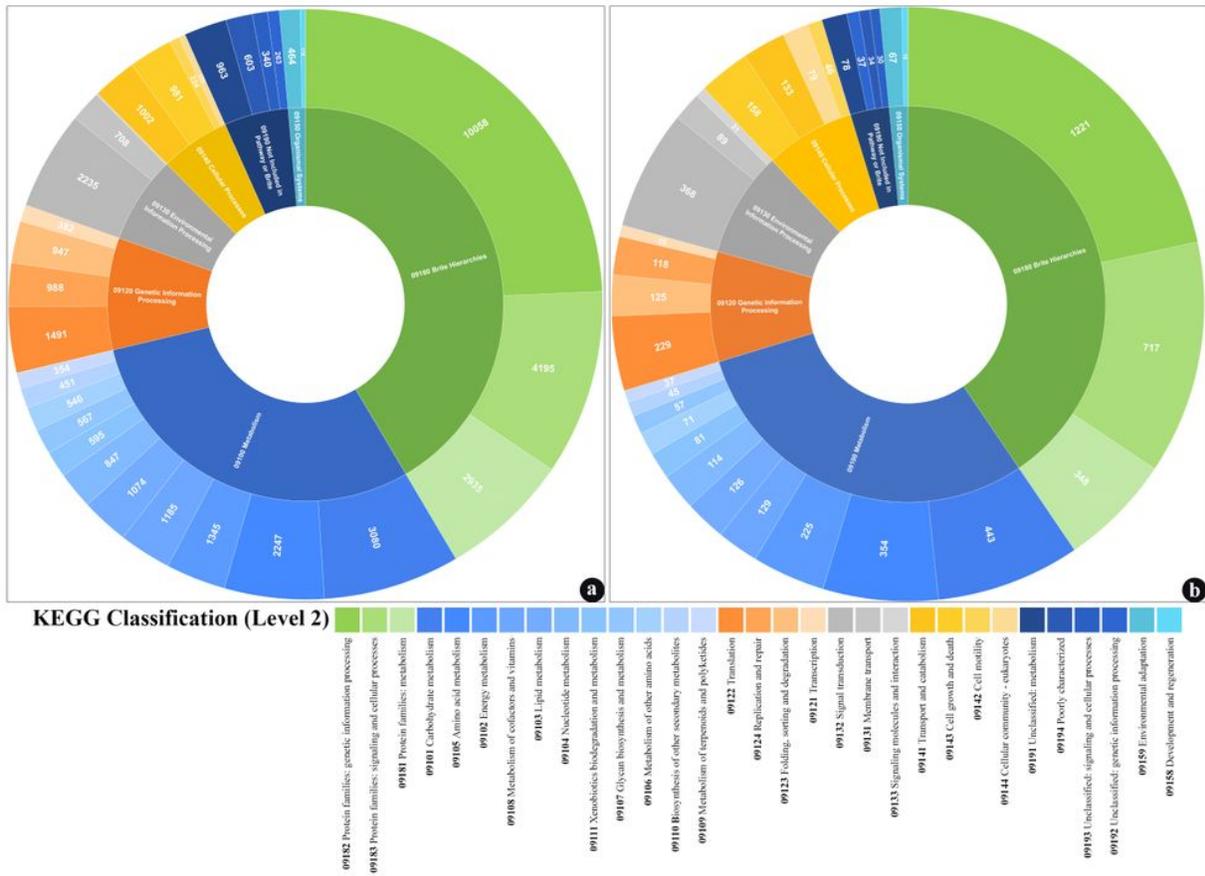


Figure 6

Classification of (a) WcL, and (b) WcR Nr annotated unigenes into different KEGG categories

Figure 7

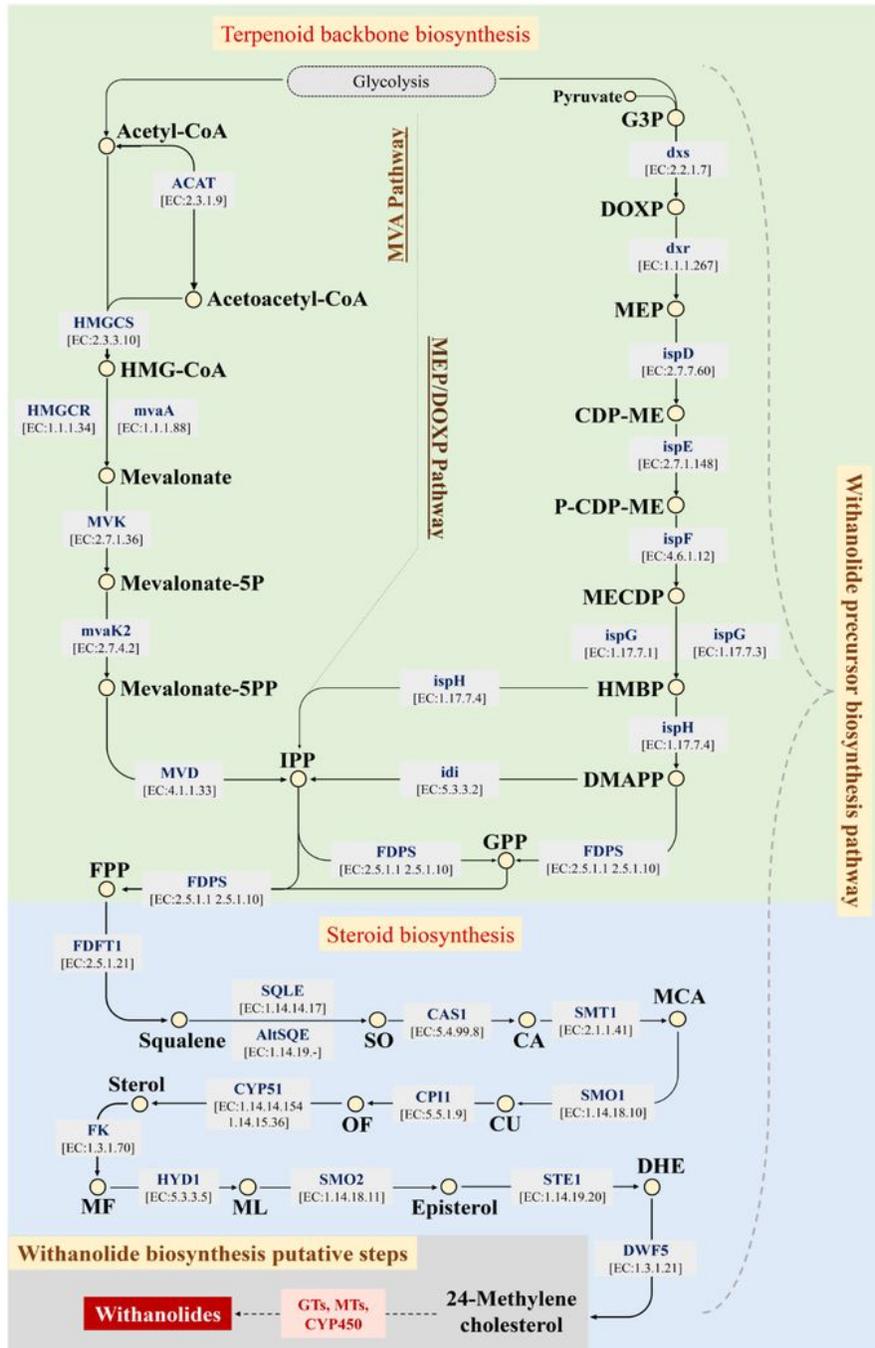


Figure 7

Putative withanolide biosynthesis pathway in *W. coagulans*. Abbreviations: HMG-CoA 3-hydroxy-3-methylglutaryl-CoA, G3P D-glyceraldehyde 3-phosphate, DOXP 1-deoxy-D-xylulose 5-phosphate, MEP 2-C-methyl-D-erythritol 4-phosphate, CDP-ME 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol, P-CDP-ME 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol, MECDP 2-C-methyl-D-erythritol 2,4-cyclodiphospho, HMBP 1-hydroxy-2-methyl-2-butenyl 4-diphosphate, DMAPP dimethylallyl

pyrophosphate, IPP isopentyl pyrophosphate, GPP geranyl pyrophosphate, FPP farnesyl pyrophosphate, SO (S)-squalene-2,3-epoxide, CA cycloartenol, MCA 24-methylenecycloartenol, CU cycloeucalenone, OF obtusifoliol, MF 4 α -methylfecosterol, ML 24-methylenelophenol, DHE 5-dehydroepisterol

Figure 8

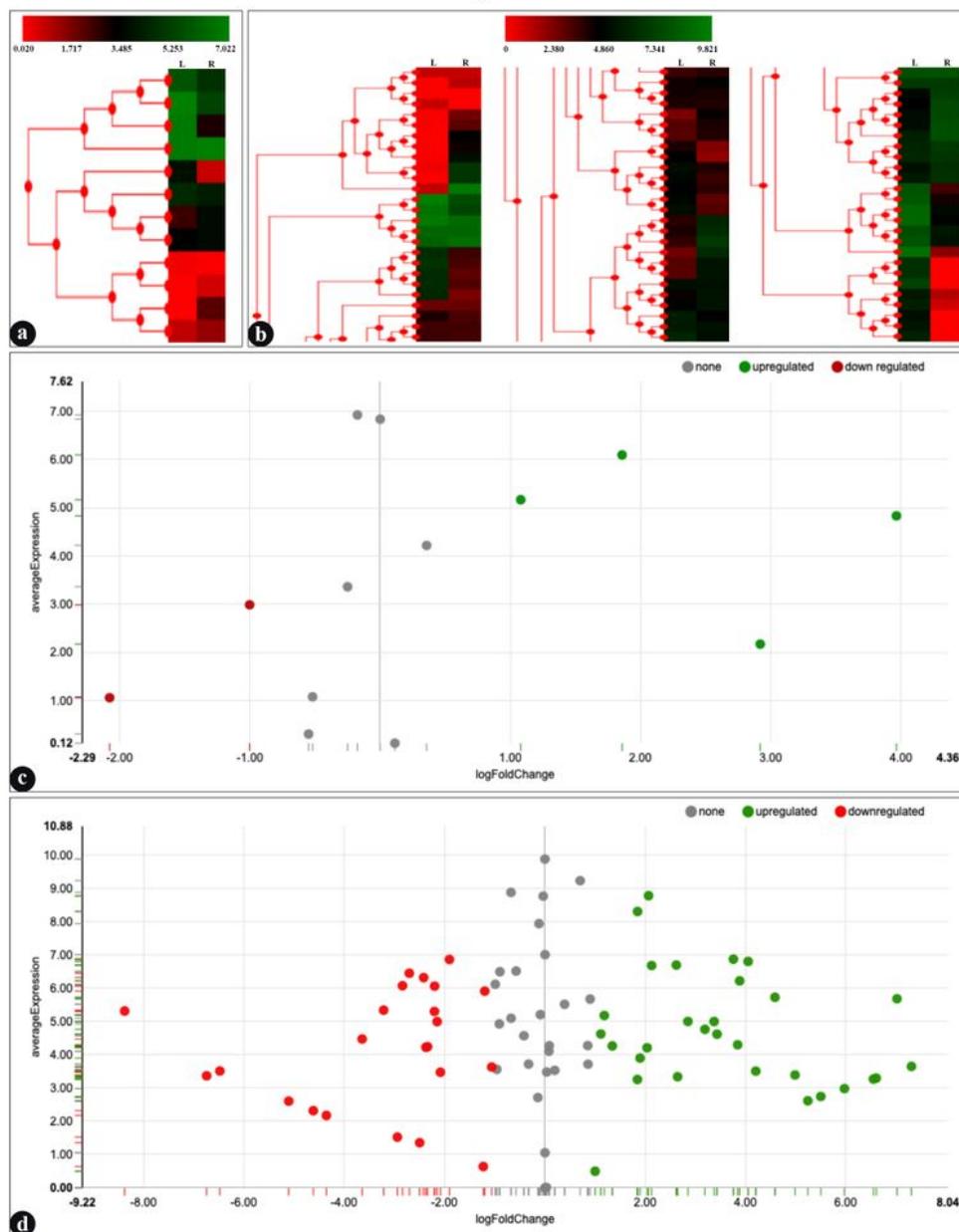


Figure 8

Heat map representing differential gene expression and hierarchical clustering of (a) enzymes involved in biosynthesis of withanolide precursor, and (b) putative enzymes involved in conversion of 24-

methylenecholesterol to withanolides in WcL and WcR. Volcano plot showing down- and up- regulation of (c) enzymes involved in biosynthesis of withanolide precursor, and (d) putative enzymes involved in conversion of 24-methylenecholesterol to withanolides in WcR w.r.t WcL

Figure 9

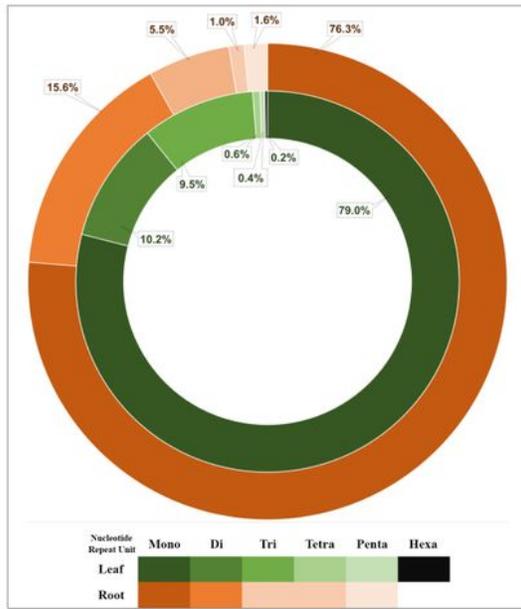


Figure 9

SSR distribution pattern in WcL and WcR transcriptome

Figure 10

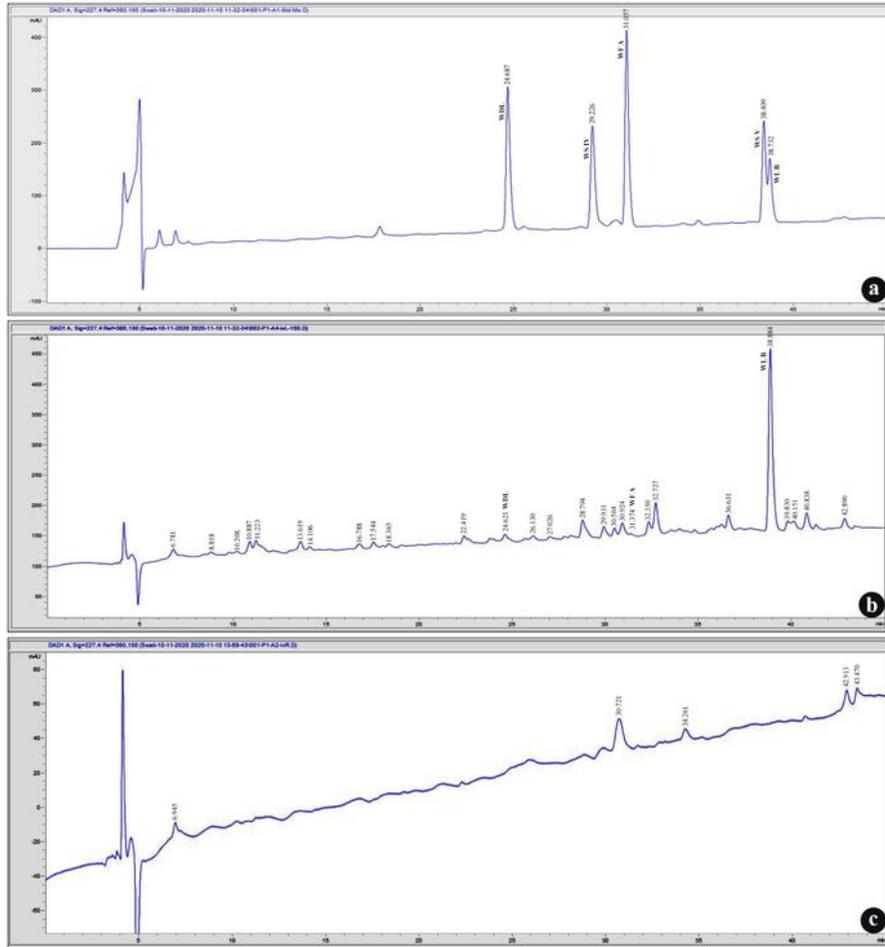


Figure 10

HPLC-DAD chromatogram representing withanolide profile of (a) withanolide standards (WF A, WDL, WL B, WS IV, WS V), (b) leaves, and (c) shoots of in vitro regenerated *W. coagulans*.