Transcriptome analysis of stem wood of *Nothapodytes nimmoniana* (Graham) Mabb. identifies genes associated with biosynthesis of camptothecin, an anti-carcinogenic molecule

BL Manjunatha[1,*], HR Singh[2], G Ravikanth[4], Karaba N Nataraja[1], Ravi Shankar[2], Sanjay Kumar[3] and R Uma Shaanker[1]

[1]*Department of Crop Physiology and School of Ecology and Conservation, University of Agricultural Sciences, GKVK, Bangalore 560 065, India*

[2]*Studio of Computational Biology & Bioinformatics and* [3]*Department of Biotechnology, CSIR-Institute of Himalayan Bioresource Technology (Council of Scientific and Industrial Research), P.O. Box No. 6, Palampur 176 061 Himachal Pradesh, India*

[4]*Ashoka Trust for Research in Ecology and the Environment (ATREE), Royal Enclave, Srirampura, Jakkur Post, Bangalore 560 064, India*

**Corresponding author (Email, manjuatsec@gmail.com)*

Camptothecin (CPT), a monoterpene indole alkaloid, is a potent inhibitor of DNA topoisomerase I and has applications in treating ovarian, small lung and refractory ovarian cancers. Stem wood tissue of *Nothapodytes nimmoniana* (Graham) Mabb. (family Icacinaceae) is one of the richest sources of CPT. Since there is no genomic or transcriptome data available for the species, the present work sequenced and analysed transcriptome of stem wood tissue on an Illumina platform. From a total of 77,55,978 reads, 9,187 transcripts were assembled with an average length of 255 bp. Functional annotation and categorization of these assembled transcripts unraveled the transcriptome architecture and also a total of 13 genes associated with CPT biosynthetic pathway were identified in the stem wood tissue. Four genes of the pathway were cloned to full length by RACE to validate the transcriptome data. Expression analysis of 13 genes associated with CPT biosynthetic pathway in 11 different tissues vis-à-vis CPT content analysis suggested an important role of *NnPG10H*, *NnPSLS* and *NnPSTR* genes in the biosynthesis of CPT. These results indicated that CPT might be synthesized in the leaves and then perhaps exported to stem wood tissue for storage.

[Manjunatha BL, Singh HR, Ravikanth G, Nataraja KN, Shankar R, Kumar S and Shaanker RU 2016 Transcriptome analysis of stem wood of *Nothapodytes nimmoniana* (Graham) Mabb. identifies genes associated with biosynthesis of camptothecin, an anti-carcinogenic molecule. *J. Biosci.* **41** 119–131] **DOI** 10.1007/s12038-016-9591-3

1. Introduction

Camptothecin (CPT), a monoterpene indole alkaloid, has been lauded as one of the most promising anticancer drugs of the 21st century. CPT exhibits a broad spectrum of anti-tumour activity due to its inhibitory activity of DNA topoisomerase I (Hsiang *et al.* 1985). It is also been reported as an inhibitor of human immunodeficiency virus (Oberlies and Kroll 2004). Irinotecan® (CPT11) and Topotecan® (TPT), two water-soluble derivatives of CPT, have been approved by the Food and Drug Administration of USA for treating colorectal and ovarian cancers as well as against several brain tumours in children (Argelia and Craig 2004). Since the discovery of CPT from *Camptotheca acuminata* (Nyssaceae) (Wall *et al.* 1966), the alkaloid has

Keywords. Camptothecin; gene expression; Illumina GA sequencing; *Nothapodytes nimmoniana*; strictosidine synthase; transcriptome

Supplementary materials pertaining to this article are available on the Journal of Biosciences Website.

Published online: 11 February 2016

been reported from several other plant species, such as *Ervatamia heyneana* (Apocynaceae) (Gunasekera *et al.* 1979), *Merrilliodendron megacarpum*, *Ophiorrhiza* sp., fruits of *Miquelia dentata*, *Nothapodytes foetida*, and several species of Icacinaceae (Tafur *et al.* 1976; Arisawa *et al.* 1981; Aimi *et al.* 1989; Suhas *et al.* 2007; Ramesha *et al.* 2008; Uma Shaanker *et al.* 2008; Ramesha *et al.* 2013). *N. nimmoniana*, however, is a major sources of CPT in India (Ramesha *et al.* 2008; Uma Shaanker *et al.* 2008), which also contains many other camptothecines including 10-hydroxy campthothecine and 9-Methoxy campthothecine (Ramesha *et al.* 2008) and hence the species is of importance from this perspective as well.

Despite the importance of CPT, the molecular basis of biosynthesis of CPT in *N. nimmoniana* is not known. Strictosidine, a precursor of terpenoid indole alkaloids (TIA), is considered as the precursor for CPT (Cordell 1974). Over 1800 terpenoid indole alkaloids, including quinine, strychnine, and the anti-cancer compounds vinblastine, and vincristine are derived from strictosidine. Strictosidine is formed by the coupling of secologanin to tryptamine in a reaction catalysed by the enzyme strictosidine synthase (*STR*). This enzyme is located in the vacuole (Yamazaki *et al.* 2003). Tryptamine is formed from the amino acid tryptophan by the action of the cytosolic enzyme tryptophan decarboxylase (*TDC*) while secologanin is a glucoiridoid formed from geraniol (figure 1).

A metabolic pathway for biosynthesis of CPT from strictosidine has been proposed (O'Connor and Maresh 2006). This pathway is based on series of oxidation and hydroxylation reactions catalysed by monooxygenases and hydroxylases, belonging to the superfamily of cytochrome P450s (Morant *et al.* 2003; Coon 2005). An aldol-type condensation yields pumiloside, which undergoes allylic isomerization, reduction and oxidation to yield CPT (figure 1). In many TIA producing plants such as *Catharanthus roseus*, strictosidine is deglycosylated to aglycon to form various alkaloids (Samira *et al.* 2007). However, CPT producers have a unique TIA pathway wherein a glucosidase for glycoside hydrolysis is likely to be involved in one of the last steps in the downstream pathway of CPT formation (Samira *et al.* 2007).

In the recent past, several genes involved in strictosidine biosynthesis have been identified and isolated from several alkaloid-producing plants. A few upstream genes like *tryptophan synthase* (*TSB*) and *TDC*, (Lopez-M and Nessler 1997; Lu and McKnight 1999), *3-hydroxy-3-methylglutaryl-CoA synthase* (*HMGR*) (Burnett *et al.* 1994), *1-deoxy-D-xylulose-5-phosphate reductoisomerase* (*DXR*) (Yao *et al.* 2008), and *10-hydroxy geraniol oxidoreductase* (*10HGO*) (Keat *et al.* 2000) involved in secologanin synthesis have been found. *Geraniol 10 hydroxylase* (*G10H*) and *secologanin synthase* (*SLS*), which belong to CYP76B6 and CYP72A1 subfamilies of cytochrome P450 family, respectively have been identified in mono-terpene biosynthesis from *C. roseus*. Miettinen *et al.* (2014) reports the discovery of the last four missing steps of the secologanin biosynthesis through transcriptomic and proteomic approaches. Functionality of all enzymes of the pathway was confirmed by heterologous production of the complex MIA strictosidine in an alternate plant host *Nicotiana benthamiana.* STR catalyses the synthesis of strictosidine, a gene identified from *Rauvolfia serpentina* (Kutchan *et al.* 1988), *C. roseus* (McKnight *et al.* 1990), and *Ophiorrhiza pumila* (Yamazaki *et al.* 2003).

Stem wood of mature *N. nimmoniana* tree accumulates maximum CPT as compared to other tissues (Suhas *et al.* 2007; Uma Shaanker *et al.* 2008). Therefore, it was of interest to study if stem wood tissue, which is considered metabolically less active as compared to leaf and other actively growing tissues, also has genetic machinery for CPT biosynthesis or that it simply acts as tissue for CPT storage. One of the bottlenecks for such analysis was absence of any molecular data for the species. One of the most efficient and comprehensive approach to obtain such data is through transcriptome analysis on next generation sequencing platform (Gahlan *et al.* 2012; Kumari *et al.* 2014). Therefore, the present work was undertaken to analyse transcriptome of the stem wood tissue of *N. nimmoniana*. Also, CPT analysis and gene expression was studied in eleven different tissues to understand capability of different tissues to synthesize and store the molecule in *N. nimmoniana*.

2. Materials and methods

2.1 *Ethics statement*

Nothapodytes nimmoniana is an economically important tree species in the Western Ghats, India. Fieldwork was carried out in the central Western Ghats region of Karnataka, with due permission from the Karnataka Forest Department. Tissue sampling was carried out under the supervision of local foresters and used solely for scientific research. The sampling was non-invasive and does not in any way affect the natural growth of *N. nimmoniana* trees.

2.2 *Plant material and RNA preparation*

Among the different plant parts, stem wood (comprising of inner and outer bark) tissue is known to accumulate higher concentration of CPT in mature trees of *N. nimmoniana* (Suhas *et al.* 2007). Accordingly, stem wood tissue was used for transcriptome analysis. Stem wood was scrapped and collected from 35- to 45-year-old trees of *N. nimmoniana*

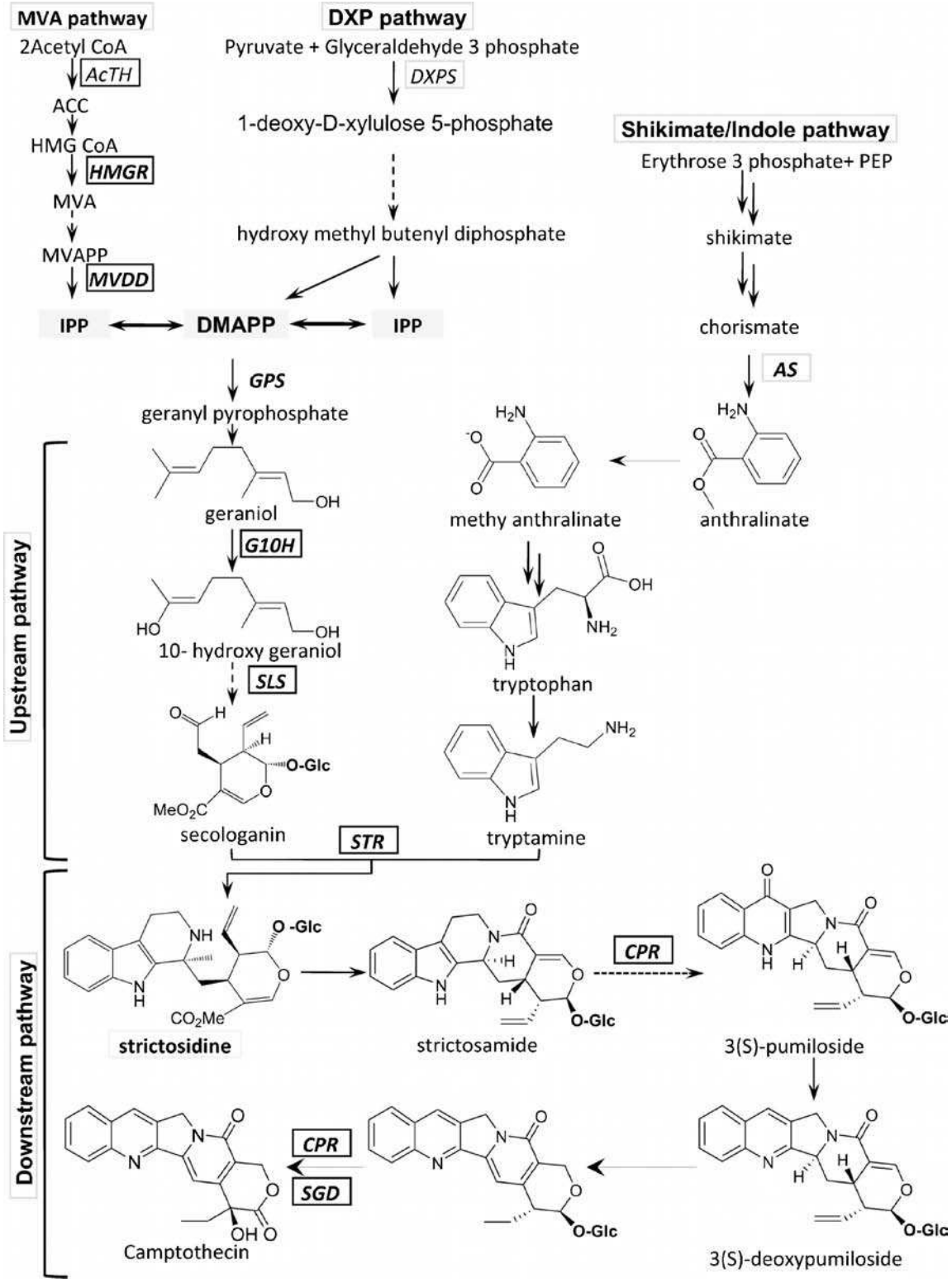

Figure 1. Schematic representation of different pathways leading to formation of CPT in *N. nimmoniana*. Candidate enzymes for which the putative transcripts were identified from the transcriptome dataset are shown in the bold box.

from their natural habitat at Biligiri Ranganswamy Temple Wildlife Sanctuary, Karnataka (12.24 °N, 77 °E), placed into an RNAase free DEPC treated plastic vials (Tarsons, India), immediately frozen in liquid nitrogen, and stored at −80°C till further use. Total RNA was isolated as described by Ghawana *et al.* (2011) and stored at −70°C until processing. Quantity and quality of total RNA were determined with a Nanodrop spectrophotometer (NanoDrop Technologies, USA) and on 1% agarose gel.

2.3 *cDNA library construction and sequencing*

Various steps involved in the preparation of cDNA library were essentially the same as described by Gahlan *et al.* (2012). In brief, poly $(A)^+$ RNA was purified from total RNA pool using Oligotex mRNA midi prep kit (QIAGEN, Germany) followed by repurification using mRNA-Seq 8 Sample Prep Kit (Illumina, USA). Following purification, the mRNA was fragmented in the fragmentation buffer (Illumina, USA) and reverse-transcribed to synthesize first strand cDNA using SuperScript II reverse transcriptase (Invitrogen) and random hexamers (Illumina, USA). Subsequently, second strand cDNA were synthesized using GEX Second Strand Buffer (Illumina, USA), RNase H and DNA polymerase (Illumina, USA). These cDNA fragments were further processed by an end repair using T4 DNA polymerase, Klenow DNA polymerase and T4 polynucleotide kinase (NEB, US), and ligation of adaptors with Illumina's adaptor oligo mix and T4 DNA ligase (Invitrogen). The products were purified from section of approximately 200 bp long using Gel Extraction Kit (Sigma) and enriched with PCR for preparing the sequencing library. Quality control analysis was performed on library to quantify the concentration using Bioanalyzer chip DNA 1200 series II (Agilent, Technologies, USA). Raw sequence reads were generated on Illumina genome analyser IIx platform (Illumina, USA) following manufacturer's instructions.

2.4 *De novo assembly and sequence clustering*

De novo assembly of *N. nimmoniana* transcriptome was carried out on the pipeline of algorithms as reported in our previous communications (Gahlan *et al.* 2012, Kumari *et al.* 2014). Illumina sequencing generated a total of 77,55,978 paired-end (PE) reads. Low quality (LQ) sequence reads with quality scores of less than Q30 were filtered and a total of 53,66,344 good quality reads were retained for de-novo transcriptome. *De novo* assembly was performed using SOAP*denovo* (Li *et al.* 2010) at different k-mer size ranging from 19 to 29 mers with read length of 33 bp to get the most appropriate k-mer for *de novo* transcriptome assembling. SOAPdenovo combined reads of different k-mer length of overlap to form contigs. These contigs were joined into scaffolds using the paired-end reads. The paired-end reads were used again for the gap filling of scaffolds to obtain sequences with the least N's that could not be extended on either end. Percent of 'N' inserted in the assembly during scaffolding was approximately 8.84. Sequence redundancy was removed by searching similar sequences with minimum similarity cut-off threshold of 90% using TGICL-CAP3 clustering program (Pertea *et al.* 2003) based on terminal region matching for at least 40 bp and 90% identity. Due to differences in various clustering algorithms approach, different clustering algorithms tools were combined to give the best results (Li *et al.* 2002). For the same reason we further used another clustering program CD-HIT with 90% similarity cut-off (Li and Godzik 2006). Finally, a total of 9061 transcript sequences were obtained from 9187 assembled transcript sequences (Bioproject ID no. PRJNA207019). GC content of assembled transcript sequence of 9061 was measured using EMBOSS package, geecee program. The GC content of each transcript was calculated and the average GC content of *N. nimmoniana* transcripts was arrived at.

2.5 *Sequence clustering based on significant blast hits and functional domains search for unknown sequences*

The assembled transcript sequences were searched for similarity against non-redundant protein (nr) sequences database at NCBI using BLASTX algorithm with an *E* value threshold of 10^{-5}. For 3243 assembled transcript sequences, which did not show any significant hit through BLASTX, RPS-BLAST (Altschul *et al.* 1990) was performed against the functional domain databases i.e. conserved domain database (CDD) (Aron *et al.* 2011).

2.6 *Functional annotation and classification*

Functional annotation of *N. nimmoniana* transcriptome was performed using annot8r program (Schmid and Blaxter 2008), where all the assembled transcripts were compared with the sequences in public domain databases, which includes NCBI non-redundant protein (Nr) database (*http://www.ncbi.nlm.nih.gov*), Swiss-Prot protein database (*http://www.uniprot.org*), KEGG (Kyoto Encyclopedia of Genes and Genomes) database (*http://www.genome.jp/kegg/*) and EC (Enzyme Commission) database (*http://enzyme.expasy.org*), using the BLASTX (Basic Local Alignment Search Tool) algorithm with an *E* value threshold of 10^{-5}. Since each query sequence exhibited multiple hits, the highest bit score was selected from each query hit of GO, KEGG and EC. The transcripts corresponding to the CPT biosynthetic pathway were identified and the annotation were scored

against existing database for scores over 100 and checked manually.

2.7 *Molecular cloning of ORFs of putative enzyme genes related to CPT biosynthesis*

For complete and partial ORF cloning, cDNA RACE was performed using SMARTer™ RACE cDNA Amplification Kit (Clontech, USA) following manufacturer instructions. All primers used for RACE PCR were designed using Primer 3 Program (*http://primer3.ut.ee*) with an optimal length of 27-30 bp, Tm ≥ 68°C and GC content of 40–60%. Putative transcript sequences recognized from the transcriptome dataset (Bioproject ID no. PRJNA207019) were used for primer designing from Primer 3 input versions (Accelrys, USA) (supplementary table 1). Advantage 2 Polymerase Mix (Clontech, USA) was used for PCR amplification of 3′ ends; 5′ ends and complete ORFs of the four genes. The 5' and 3' fragments were integrated into a pGEM®-T Easy vector (Promega, Madison, USA) and transferred into *E. coli* MAX Efficiency® DH5α™ competent cells (Invitrogen™, India). The isolated clones were sequenced on ABI PRISM® 3100 Genetic Analyzer (ABI, USA).

2.8 *Expression analysis by quantitative RT-PCR*

In a separate experiment, expression of various genes involved in CPT biosynthesis was studied in 11 different tissues namely fruit stalk, immature fruit stalk, young fruits, immature fruits, mature fruits, young leaves, mature leaves, old leaves, inner bark, outer bark and roots of 6 year old saplings of *N. nimmoniana*. The advantage of using younger plants was two fold; first CPT data for different tissue was not reported and secondly, it allowed harvest of tissues under the defined growth conditions. All the samplings used for this study are being maintained at the nursery in the University experimental facility. These materials were raised from seeds collected from the Western Ghats. Initially they were raised in polybags and later transferred to the forest nursery. RNA was isolated as described by Ghawana *et al* (2011) and one µg of total RNA from each tissue was taken for cDNA preparation using SuperScript ™ III First-Strand Synthesis System (Invitrogen™, life technologies) after treating with RNase-free DNase I. qRT-PCR was carried out using SYBR® Premix Ex TaqTM (Perfect Real Time) (TaKaRa, Dalian, China) on an Mx3000P Real-Time PCR Detection System (Agilent technologies, Stratagene). Each qRT-PCR system contained 5 µL SYBR® Green (Applied Biosystems, UK), 0.2 µM forward and reverse primers and 1 µL cDNA template. The PCR amplification program was as follows: 95°C for 10 min; 40 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by a melting-curve program of 95°C for 1 min, 55°C for 30 s and 95°C for 30 s. The gene expression patterns of all genes were normalized to an internal reference gene (*26S rRNA*) (Singh *et al.* 2004). Expression analysis was performed for three biological replicates from each representative samples. The relative gene expression analysis was performed using REST software with the $2^{-\Delta\Delta Ct}$ method (Michael *et al.* 2002). All the Real time quantitative PCR primers (supplementary table 2) were designed using Primer 3 input versions (Accelrys, USA) with suitable parameters.

2.9 *Sample preparation and CPT extraction*

Representative replicate samples of fruit stalk, immature fruit stalk, young fruits, immature fruits, mature fruits, young leaves, mature leaves, old leaves, inner bark, outer bark and roots of 6-year-old saplings (as described in section 2.8) were dried to constant moisture content at 60°C for 96 h in a hot air oven. The dried samples were ground to fine powder using a pestle and mortar. A sample (100 mg) of powder of each of the samples was extracted in 61% EtOH (10 mL) at 60 °C for 90 min in a shaking water bath. After cooling to room temperature, the extract (1 mL) was centrifuged at 10,000 rpm for 10 min at 10°C. The supernatant was passed through a 0.2 µ filter (Tarsons, India) and subjected to HPLC analysis (Suhas *et al.* 2007; Ramesha *et al.* 2008; Uma Shaanker *et al.* 2008)

2.10 *High-performance liquid chromatographic analysis*

CPT was analysed on a RP-18 column (4.6×250 mm, 5µm) with a PD-M 20A photo diode array detector. The HPLC (Shimadzu, Japan) conditions were: 254 nm as the detector wavelength, 1.6 mL/min flow rate and 20 µL sample loop. The mobile phase consisted of: 40% CH_3CN and 60% H_2O + 0.1% CF_3CO_2H in an isocratic mode (t_R=3.5 min). One mg/ml stock of CPT (Sigma Aldrich Inc, USA) standard solution was prepared using DMSO and methanol (1:3 v/v). Standard curve was developed for the concentration range of 50 μgmL^{-1} to 250 μgmL^{-1} with an interval of 50 μgmL^{-1}. The best fit (R^2=0.99) was used in calculating the amount of CPT in the sample (Suhas *et al.* 2007; Ramesha *et al.* 2008; Uma Shaanker *et al.* 2008).

3. Results and discussion

3.1 *Illumina sequencing and de novo assembly*

Reads were generated from the stem wood tissues of *N. nimmoniana* using Illumina paired-end sequencing technology. The 36 base pairs (bp) short PE reads were converted to *qseq* format using *bcl* converter program, where

qseq files consist of reads, coordinates number, tile number and quality encodings. However, for every 36 bp reads only 33 bp reads were considered for downstream analysis since 3′ ends of reads were prone to sequencing error, and therefore 3 bases at 3′ end were excluded. A total of 77,55,978 PE reads was generated and after quality filtering a total of 53,66,344 paired end (PE) reads were obtained (table 1).

The parameters such as average coverage, average length, maximum length, total transcripts, percentage of transcripts of 1000 bp and above, contig N50 and scaffold N50 were used to justify the best k-mer size. Data showed that 23 k-mer sizes emerged to be the best k-mer for assembling (table 2). Assembled transcripts were enriched at higher k-mer with much higher coverage. However, total number of assembled transcripts decreased linearly with increase in k-mer size as was also reported by Gahlan *et al* (2012). Over-representation of assembled transcripts was more at lower k-mer and under-representation at higher k-mers. For assembling the transcriptome, only those reads were considered which produced high frequency k-mer. For more sensitive assembling, PE module was used with approximate distance of 200bp between PE reads. For all the assembling steps, minimum length cut-off for assembled transcripts was set to 100 bp. A total of 9,187 transcripts were obtained, of which 1.13% were above 1,000 bp, yielding average coverage of 62.52, average length of 254.92 bp and maximum length of 3,162 bp (table 2). Finally, a total of 9,061 transcript sequences were obtained from 9,187 assembled transcripts by clustering program CD-HIT with 90% similarity cut-off (Li and Godzik 2006). The low levels of total assembled transcripts may be due to the relative difficulty in isolation of mRNA from the wood tissue as well as the low abundance of total transcript pools in the wood tissue compared to other parts of the plant.

The average GC content of *N. nimmoniana* transcripts was found to be 47.7%, which is within the range of GC content levels of dicot plant coding sequences (44-47%) (Vinogradov 2003). Sixty-five percent of *N. nimmoniana* transcripts had a GC content ranging between 40% and 49% (figure 2). The average GC content of 47.7% is higher than the GC content reported for *Arabidopsis* (42.5%) (Riki and Julia 2005).

3.1.1 *Sequence clustering based on significant blast hit and functional domains search for unknown sequences:* A total of 5,818 sequences exhibited significant hit by BLAST, while no significant hit was found for 3,243 sequences. Since the actual unique genes are expected to be much lower than the total assembled sequences produced from any *de novo* assembling tool, clustering algorithms was deployed as described by Gahlan *et al.* (2012) which yielded a total of 5138 unigenes from 9061 transcripts. The grouped sequences are available in supplementary table 3. A total of 290 assembled transcript sequences showed significant functional domain against CDD. Among the significant domains, pre-mRNA-splicing factor 38 proteins, fibrillarin, protein of unknown functions, extensin-like region and topoisomerase II-associated protein were most represented (figure 3).

Table 1. Summary of transcriptome sequencing and assembly of stem wood tissues of *N. nimmoniana*

Total number of paired end reads	77,55,978
No. of reads obtained after quality filtering	53,66,344
No. of assembled transcripts	9,187
Average length of transcripts (in base pair)	255
Average coverage (%)	63

3.1.2 *Functional annotation and classification:* The unigenes were analysed with those existing in NCBI non-redundant protein (Nr), Uniprot/SwissProt and KEGG databases. A total of 4804, 2937 and 1998 significant hits were obtained from 9187 transcripts and represents the best possible hits against GO, KEGG and EC database respectively, these genes were further classified into two major categories, namely, biological process and molecular function using plant specific GO slims. Highly represented groups in the biological process category included those for metabolic processes, response to stimulus and nucleobase-containing compounds (supplementary figure 1A). Besides these, genes for other biological processes such as multicellular organismal development, cellular process, cell differentiation, cell communication, cell death and transport were also identified. Genes involved in DNA binding, catalytic and transferase activities were highly represented in molecular function category (supplementary figure 1B). Other genes such as those encoding for antioxidant activity, motor activity, isomerase activity, enzyme regulator activity, signal transducer activity were also identified in the molecular function through GO annotation. A total of 1733, 1506 and 1565 unigenes were classified into molecular function, biological process and cellular component categories, respectively.

Best KEGG and EC classification were obtained for 2,937 and 1,998 unigenes, respectively. Supplementary figure 2A shows the top 20 KEGG pathways observed for *N. nimmoniana* transcriptome. A large number of assembled transcripts belonged to ribosomal genes. Genes reflecting other important pathways such as those associated with plant signal transduction, spliceosome, RNA transport, protein processing in endoplasmic reticulum, glycolysis/gluconeogenesis were also identified. Supplementary figure 2B shows the top 20 abundant enzyme classes with highest number of transcripts belonging to serine/threonine protein kinase followed by RNA helicase enzyme. These

Table 2. Effect of *k-mer* size on *de novo* assembly of transcriptome of stem wood tissues of *N. Nimmoniana*

K-mer	Average coverage	Average length (bp)	Maximum length (bp)	Total transcripts	% transcripts≧ 1000 (bp)	Contig N50 (bp)	Scaffold N50 (bp)
19	56.32	245.09	2,290	15,796	0.85	158	319
21	58.63	254.09	2,554	12,534	1.24	158	336
23	**62.52**	**254.92**	**3,162**	**9,187**	**1.13**	**158**	**340**
25	65.94	248.77	4,156	6,168	1.56	157	339
27	73.13	249.30	2,854	3,377	1.69	157	339
29	88.66	239.55	2,529	1,314	1.22	154	338

Note: Boldface indicates the best k-mer size selected for transcriptome assembly

play a crucial role in the regulation of mRNA quality control and gene expression in plant development. Hardie (1999) described that the network of protein serine/threonine kinases in plant cells appear to act as a 'central processor unit' (cpu), accepting input information from receptors that sense environmental conditions, phytohormones, and other external factors, and convert it into appropriate outputs such as changes in metabolism, gene expression, and cell growth and division.

3.1.3 *Transcripts involved in the upstream pathway leading to the formation of strictosidine:* The proposed biosynthetic pathway leading to CPT formation via strictosidine is shown in figure 1. Geraniol is the basic isoprenoid moiety, which is involved in the biosynthesis of a large number of secondary metabolites including CPT. Plants use mevalonate (MVA) and methyl D-erythritol 4-phosphate (MEP), either in isolation or in combination, to synthesize terpenes. Studies in *Ophiorrhiza pumila* showed that majority of CPT is synthesized from MEP pathway though dimethylallyl pyrophosphate (DMAPP) (Eisenreich *et al*. 2004). DMAPP is catalysed by the enzyme GPS to form geraniol (figure 1), which serves as a precursor for the synthesis of secologanin. Tryptamine synthesized from the shikimate pathway is condensed with secologanin by the key enzyme STR to form strictosidine (Stockigt and Ruppert 1999).

From the assembled data set, a total of 35 transcripts representing 13 genes involved in the CPT biosynthesis and transport were discovered (supplementary table 4). Of these, 10 putative genes belong to the upstream CPT biosynthetic pathway leading to the formation of key precursor molecule strictosidine were identified, which include those encoding three important enzyme *G10H*, *SLS*, and *STR* (supplementary table 4). None of these genes had been reported previously from *N. nimmoniana*. Based on the assembled transcripts in transcriptome data set from *N. nimmoniana,* four putative complete open reading frames (ORFs) coding for *NnPMVDD* (GenBank accession no. KF589937), *NnPG10H* (GenBank accession no. KF589936), *NnPSLS* (GenBank accession no. KF589935), and *NnPSTR* (GenBank accession no. KF589934) were cloned to full length through rapid amplification of cDNA ends (RACE), and analysed. All the amino acid sequences representing each ORFs were analysed for their homology and catalytic domain's with the respective proteins at NCBI database. Four putative partial ORFs namely *NnPAcTH* (GenBank accession no. KF589939), *NnPDXPS* (GenBank accession no. KF589942), *NnPCPR* (GenBank accession no. KF589941) and *NnPAS* (GenBank accession no. KF589941)

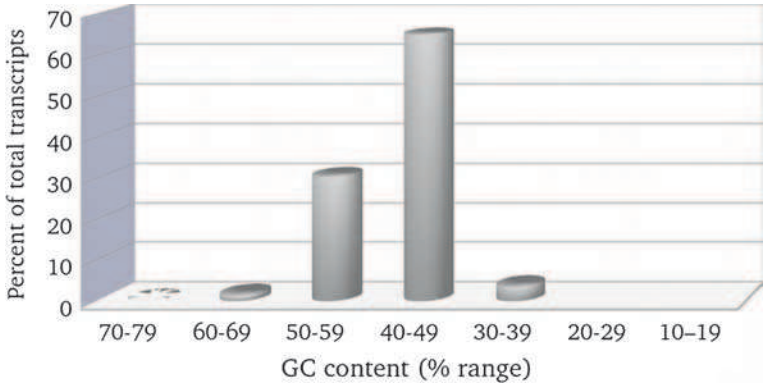

Figure 2. Guanine-cytosine (GC) content analysis of *N. nimmoniana* transcripts.

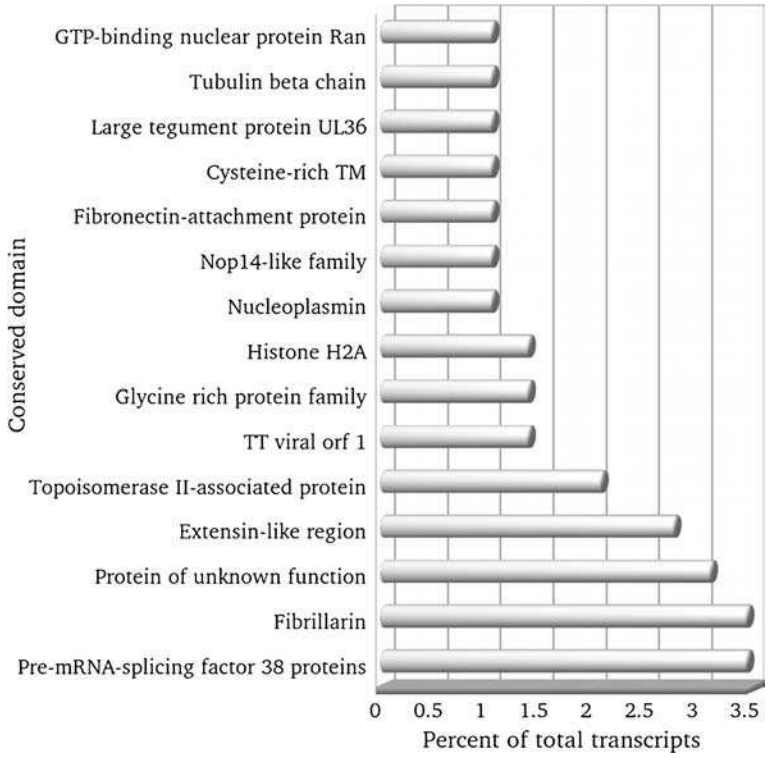

Figure 3. Highly represented functional conserved domains found in unknown sequences.

were cloned and analysed. All these putative genes of CPT pathway had >98% sequence similarlity with their corresponding assembled trancripts used for ORF cloning.

3.1.4 *Transcripts related to downstream pathway leading to synthesis of CPT and its transport:* The genes involved in the downstream pathway leading to CPT formation from strictosidine are yet to be deciphered (O'Connor and Maresh 2006). The proposed biosynthetic pathway suggests the involvement of cytochrome P450s in the intermediate steps from strictosamide to pumiloside formation and in the last steps of CPT biosynthetic pathway (figure 1). Recently, three unique NADPH cytochrome P450 reductase's (CPRs) were isolated and characterized from *N. foetida* cDNA library (Huang *et al.* 2012). It is likely that these are the genes involved in the intermediate steps of CPT biosynthetic pathway. Four putative *CPR* transcripts were also identified from the data set, which showed significant similarities with the CPRs reported from *N. foetida*. These could be the possible candidates involved in the downstream pathway of CPT biosynthesis in *N. nimmoniana.*

Three transcripts (C50327_30.0, scaffold130_36.8 and C55545_39.0) representing the superfamily glucosidases were identified from the assembled datasets. These are involved in the plant defense and secondary metabolism and catalyse the action of deglycosylation (Ketudat and Esen 2010). This gene is likely to be participating in the last step of CPT biosynthesis (figure 1). Two scaffolds (scaffold1029_35.5 and scaffold2323_92.4) corresponding to the multidrug resistance protein (MDR) transporters belonging to the ATP-binding cassette (ABC) transporter family were also identified. These have been earlier shown to be involved in alkaloid translocation (Sakai *et al.* 2002; Terasaka *et al.* 2003) and it is likely that these are involved in the CPT transport in *N. nimmoniana*. All the transcripts representing the genes related to CPT biosynthetic pathway and its transport are being reported for the first time from *N. nimmoniana*.

3.2 *Expression analysis of putative transcripts of CPT biosynthetic pathway in different tissues of N. nimmoniana*

Reports have suggested that stem wood tissue of mature trees (35 to 45 years old) of *N. nimmoniana* contained the maximum concentration of CPT as compared to its other tissues (Uma Shaanker *et al.* 2008). Transcriptome data was thereby obtained for stem wood tissue of *N. nimmoniana*, which established the presence of all the genes responsible for synthesizing CPT, suggesting the capability of the tissue to synthesize the molecule.

Biosynthesis of CPT involves precursors from MVA, MEP and shikimate/indole pathways (figure 1), which are present

optimally in metabolically active tissues. Therefore, 11 different tissues of six-year-old saplings of *N. nimmoniana* were analysed for CPT content vis-à-vis gene expression. Results indicate that these saplings had high CPT content in fruits (young, immature and mature), young leaves and roots (figure 4A) as opposed to adult trees. The high content of CPT in actively growing tissues such as young leaves and roots could be under selection to defend these from herbivore attack. The tissues themselves however are not affected by the high endogenous CPT as they possess critical point mutations on the CPT binding domain of Topoisomerase-1 that confer self-resistance to CPT (Sirikantaramas *et al.* 2015).

Expression of putative genes *NnPAcTH, NnPHMGR* and *NnPMVDD* of the MVA pathway was found to be high in actively growing tissues (immature fruit stalk, immature fruits, young fruits, young leaves and roots) (figure 5A) with nearly 2- to 4-fold higher expression as compared to old leaves and stem wood (bark tissues). The expression of *NnPDXPS* of the MEP pathway (encodes DXPS that catalyses the formation of DXP) (figure 1) was also found to be high in actively growing tissues (young leaves, 5.12-fold; young fruits, 4.31-fold; roots, 3.61-fold) as compared to old leaves (figure 5B). Expression of an upstream gene *NnPAS* of the shikimate pathway, involved in the reaction associated with diversion of chorismate towards tryptamine synthesis, was high in young fruits (6.34-fold), young leaves (4.96-fold) and roots (3.34-fold) as compared to matured and old leaves (<1.36-fold) and inner and outer bark tissues (<0.4-fold) (figure 5B). The significantly higher expression of these genes in younger tissues indicates the preeminence of this pathway in basic biological process as well as in secondary metabolite synthesis (Eisenreich et al. 2004).

3.2.1 *Genes involved in the CPT biosynthesis:* Secologanin and tryptamine are two substrates synthesized from the secoiridoid and shikimate pathways, respectively (figure 1). Expression levels of *NnPG10H, NnPCPR* and *NnPSLS* of the secoiridoid pathway were analysed in different parts of *N. nimmoniana. NnPG10H* and *NnPCPR* showed higher folds of expression in fruit stalk, young fruits, and young leaves. Huang *et al.* (2012) also reported abundant CPR transcripts in flower buds, seeds and leaves of *N. foetida*. The expression of *NnPSLS,* which is downstream to *NnPG10H* and *NnPCPR* and responsible for the formation of secologanin was highly specific to fruit stalk, immature fruit stalk and roots (figure 5C). In the downstream pathway of CPT biosynthesis, *NnPSTR* condenses secologanin and tryptamine to strictosidine, the key precursor of MIA pathway. The expression of *NnPSTR* was also analysed and found to have a 5-fold increase in fruit stalk and roots, indicating that the young tissues might be the sites of strictosidine synthesis which then might go on to produce CPT (figure 5C).

Although the downstream pathway of CPT biosynthesis is deduced (O'Connor and Maresh 2006), the enzymes involved in the conversion of strictosidine to CPT remains unknown. It was hypothesized that a *glucosidase* is involved in the deglycosylation of a precursor molecule prior to the formation of CPT (figure 1). Few studies have reported a ß-

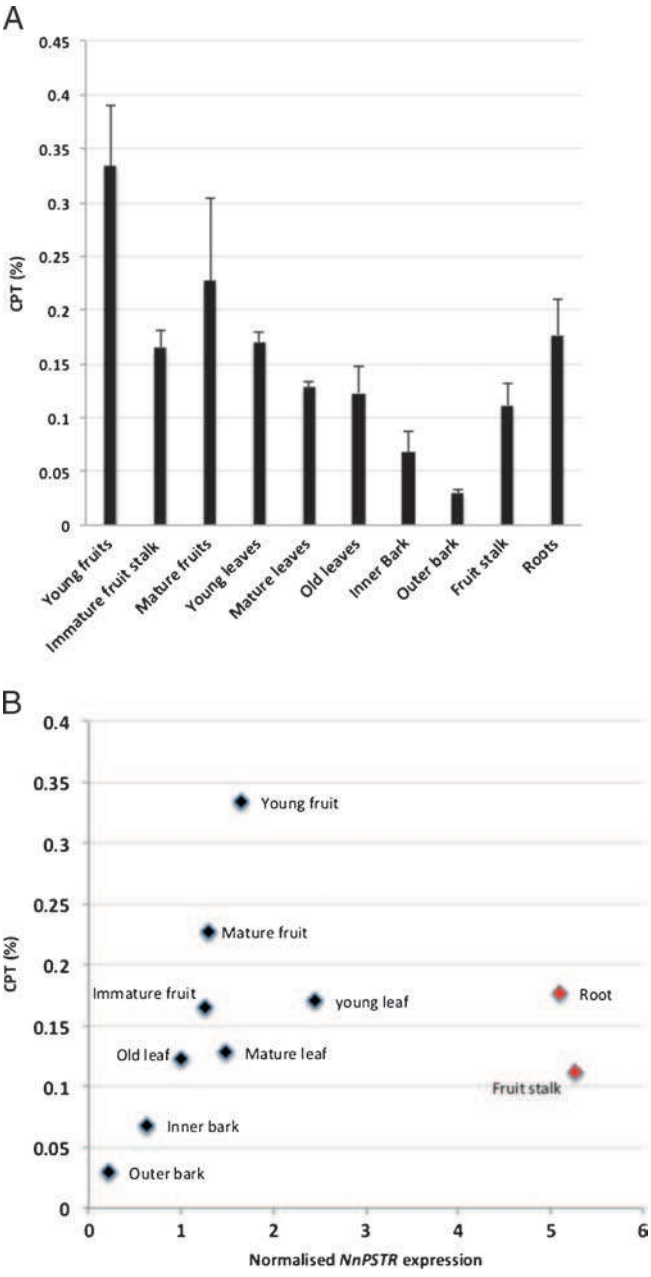

Figure 4. Camptothecin (CPT) content and its correlation with *NnPSTR* expression: (**A**) Mean percent CPT per gram dry weight in different tissues of *N. nimmoniana*; (**B**) Correlation between CPT content (%/g dry wt) and normalized fold expression of *NnPSTR* in different tissues of *N. nimmoniana*. Correlation co-efficient r=0.1505, and on excluding root and fruit stalk r = 0.627 at p-value 0.10 (supplementary table 5).

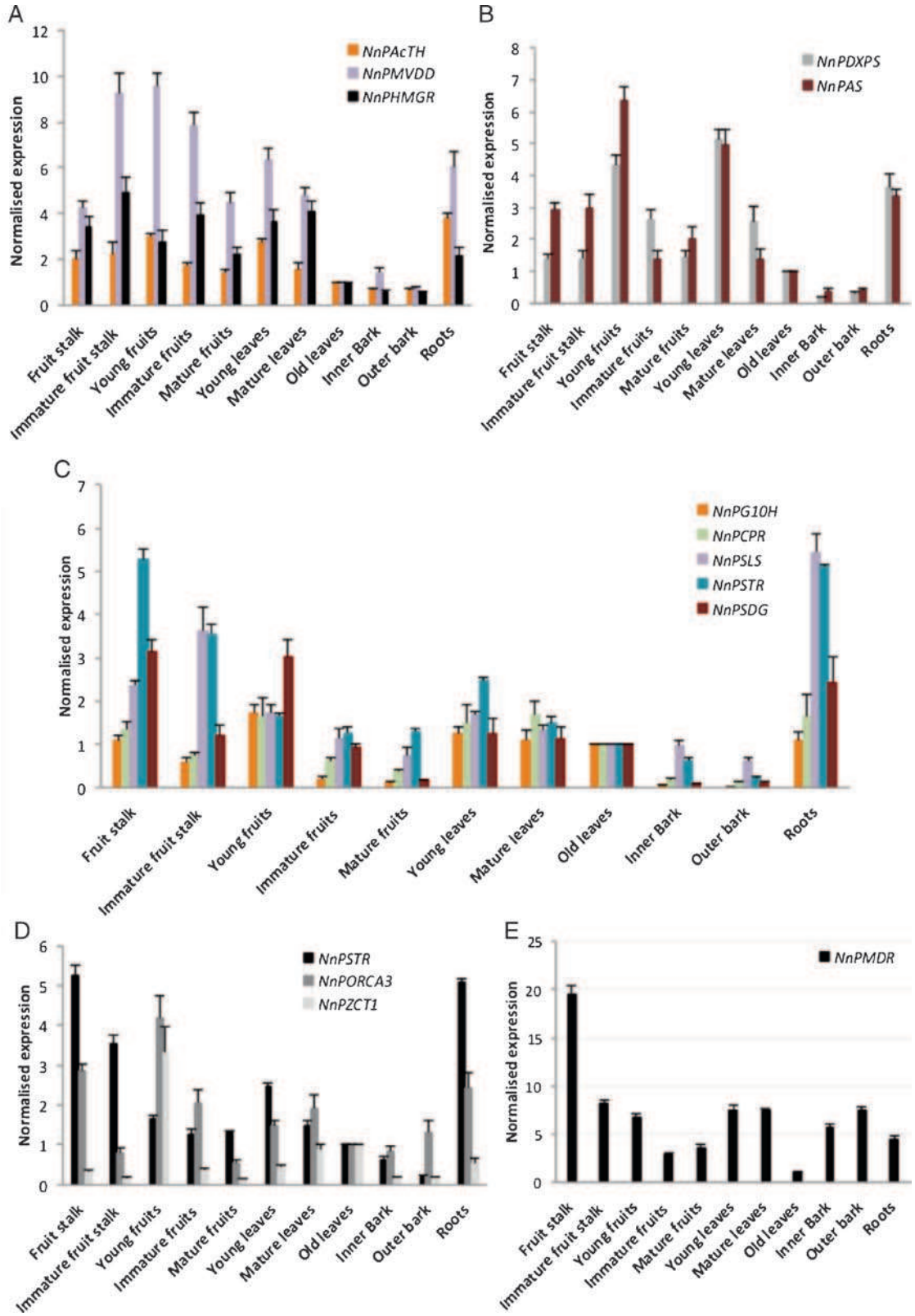

Figure 5. Expression of candidate unigenes of upstream and downstream CPT biosynthetic pathway: (**A**) Genes of MVA pathway; (**B**) DXP and shikimate pathway; (**C**) Upstream pathway including *NnPSDG*; (**D**) transcription factors of *NnPSTR*; (**E**) *MDR transporter*. Transcript levels in old leaves served as control for the calculation of normalized expression. *26s* rRNA served as internal control to normalize the expression across different plant parts.

D-glucosidase involved in the deglycosylation in TIA biosynthesis in plants such as *C. roseus* (Hutchinson *et al.* 1979), *Psychotria ipecacuanha* (Nomura *et al.* 2008) and *R. serpentina* (Warzechaa *et al*. 2000; Barleben *et al.* 2007). This gene is likely to be participating in the last step of CPT biosynthesis (figure 1). Three transcripts (C50327_30.0, scaffold130_36.8 and C55545_39.0) representing the superfamily glucosidases (*NnPSDG*) were identified from the assembled datasets and its expression was examined in different tissues. Higher expression was obtained in fruit stalk (3.13-fold), young fruits (3.0-fold), and roots (2.43-fold) as compared to mature tissues of the trees (figure 5D). These results further supported the deduction that actively growing tissues could be the sites of CPT synthesis.

3.2.2 *Genes involved indirectly in CPT biosynthesis and/or translocation:* Transcriptional regulators are well known to regulate the TIA pathway by either activating or repressing the genes involved. *ORCA3* (octadecanoid-responsive *Catharanthus* AP2-domain) reported to be an activator, increases the transcripts of the TIA biosynthesis genes, *CPR, TDC, SLS* and *STR* when overexpressed in *C. roseus* cultured cells (Menke *et al*. 1999; Van der fit and Memelink 2000). Whereas *ZCT1* (zinc-finger *catharanthus* transcription factor) had been shown to repress the activity of *STR* and *TDC* promoters in transient trans-activation assays (Ohta *et al.* 2001; Hiratsu *et al.* 2003; Pauw et al. 2004). Putative transcripts representing *ORCA3* (C60649_110.0, C49485_40.0 and scaffold 2270_63.2) and *ZCT1* (scaffold2466_37.1 and scaffold2577_157.4) were identified from the data set and analysed for their transcript levels in various tissues (figure 5D). Expression of *NnPORCA* was uniformly high in all actively growing tissues (fruit stalk, 2.8-fold; young fruits, 4.18-fold; young leaves, 2.41-fold; and roots, 2.42-fold) as compared to *NnPZCT,* which showed higher expression in mature and old leaves. These results also strengthen the proposition that the younger tissues may be the sites of synthesis of CPT in *N. nimmoniana*.

Multidrug resistance protein (MDR) transporters belonging to the ATP-binding cassette (ABC) transporter family has been reported to be involved in the transport of alkaloids in plants such as *Coptis japonica* (Sakai *et al.* 2002; Shitan *et al.* 2003; Yu and De Luca 2013). From the transcriptome data sets of *N. nimmoniana*, two scaffolds (scaffold1029_35.5 and scaffold2323_92.4) corresponding to MDR transporters were identified. Expression analysis of this transporter in different parts of *N. nimmoniana* indicated about 20-fold higher transcripts levels in fruit stalks and about 5-fold higher expression in young fruits, young and mature leaves as well as in inner and outer bark tissues; old leaves had by and large the poorest expression of this gene (figure 5E). These results indicate that after the synthesis of CPT in the younger tissues such as young leaves, fruit stalk and roots, it is probably translocated to older tissues such as the inner and outer bark (for accumulation) through the MDR type ABC transporters. Further studies on functional characterization and validation of these transporters in CPT translocation are being undertaken.

3.3 *Correlation between the tissue gene expression levels and CPT accumulation*

CPT content of tissues was significantly correlated to expression of *NnPAS*, *NnPG10H*, *NnPORCA* and *NnPZCT*. Upon excluding the fruit stalk and roots whose *NnPSTR* expression levels were very high, tissue CPT was significantly correlated with *NnPAS* (r=0.8265), *NnPSTR* (r=0.6273), *NnPSDG* (r=0.7707), *NnPORCA* (r=0.8265) and *NnPZCT* (r=0.7337) (figure 4B; supplementary table 5). The low CPT content in fruit stalk and roots despite their very high *STR* expression could be in part explained by the possible transport of CPT synthesized in these tissues to other parts of the plant. This is partly supported by the fact that the expression level of the *NnPMDR*-type ABC transporter gene of the fruit stalk is the highest (nearly 20-fold) among all tissues. It is likely that the fruit stalk is actively engaged in the transport of CPT to the developing fruits. The roots may be involved in the active synthesis and transport of CPT to the bark tissues, especially evident in later stages of the plant growth. These results suggested that younger tissues could be the sites of synthesis of CPT though not necessarily the major sites of accumulation. The findings are consistent with those reported earlier in *C. acuminata* (Samuelsson 1999; Coon 2005). Further, significantly higher expression of many of the genes including *NnPSTR* in the roots suggested a role of root tissues in synthesizing CPT.

3.4 *Conclusions*

Transcriptome analysis of stem wood tissue of *N. nimmoniana* unraveled several key genes involved in the biosynthesis of CPT. Detailed gene expression studies of various tissues coupled with CPT content analysis suggested that younger tissues and roots could be the major sites of synthesis of CPT. The CPT thus synthesised is possibly translocated to the stem wood tissue during the course of the growth and development of the trees. Currently studies are underway to study the role of transporters that might be involved in the translocation of CPT from sites of synthesis to the stem wood and other storage tissues like seeds. These studies are the first to unravel the CPT biosynthetic genes in *N. nimmoniana*.

Acknowledgements

The work reported here has been facilitated by grants from the Department of Biotechnology, New Delhi, India, and

collaboration between the CSIR–Institute of Himalayan Bioresource Technology (IHBT), Palampur, HP, India, and the University of Agricultural Sciences, Bangalore, India. The authors thank the Director CSIR-IHBT for providing necessary facilities for carrying out transcriptome sequencing, bioinformatics analyses, and gene cloning work.

References

Aimi N, Masashi N, Akiko M, Hiroyuki H, Shin-ichiro S and Joju H 1989 Pumiloside and deoxypumiloside: plausible intermediates of camptothecin biosynthesis. *Tetrahedron Lett.* **30** 4991–4994

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ 1990 Basic local alignment search tool. *J. Mol. Biol.* **215** 403–410

Argelia L and Craig LN 2004 Camptothecin, over four decades of surprising findings. *Phytochemistry* **65** 2735–2749

Arisawa M, Gunasekera SP, Cordell GA and Farnsworth NR 1981 Plant anticancer agents XXI. Constituents of *Merrilliodendron megacarpum. Planta Med.* **43** 404–407

Aron MB, Shennan L, John B, Anderson FC, Myra KD, Carol DS, Jessica HF, *et al*. 2011 CDD: a Conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **39** 225–229

Barleben L, Panjikar S, Ruppert M, Koepke J and Stöckigt J 2007 Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *Plant Cell.* **19** 2886–2897

Burnett RJ, Maldonado-Mendoza IE, McKnight TD and Nessler CL 1994 Expression of a 3-Hydroxy-3-Methylglutaryl coenzyme a reductase gene from *Camptotheca acuminata* is differentially regulated by wounding and methyl jasmonate. *Plant Physiol.* **103** 41–48

Coon MJ 2005 Cytochrome P450: nature's most versatile biological catalyst. *Annu. Rev. Pharmacol.* **45** 1–25

Cordell GA 1974 The biosynthesis of indole alkaloids. *Lloydia.* **37** 219–298

Eisenreich W, Bacher A, Arigoni D and Rohdich F 2004 Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cell. Mol. Life Sci.* **61** 1401–1426

Gahlan P, Singh HR, Shankar R, Sharma N, Kumari A, Vandna C, Paramvir SA and Sanjay K 2012 *De novo* sequencing and characterization of *Picrorhiza kurrooa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics* **13** 126

Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Pardeep KB, Arti R, Ravi SS, *et al*. 2011 An RNA isolation system for plant tissues rich in secondary metabolites. *BMC. Res. Notes* **4** 85

Gunasekera SP, Badawi MM, Cordell GA, Farnsworth NR and Chitnis M 1979 Plant anticancer agents X. Isolation of camptothecin and 9-methoxycamptothecin from *Ervatamia heyneana. J. Nat. Prod.* **42** 475–477

Hardie DG 1999 Plant protein serine/threonine kinases: classification and functions. *Annu. Rev. Plant Physiol.* **50** 97–131

Hiratsu K, Matsui K, Koyama T and Ohme-Takagi M 2003 Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis. Plant J.* **5** 733–739

Hsiang YH, Hertzberg R, Hecht S and Liu LF 1985 Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* **260** 14873–14878

Huang FC, Sung PH, Do YY and Huang PL 2012 Differential expression and functional characterization of the NADPH cytochrome P450 reductase genes from *Nothapodytes foetida. Plant Sci.* **190** 16–23

Hutchinson CR, Heckendorf AH, Straughn JL, Daddona PE and Cane DE 1979 Biosynthesis of camptothecin: III. Definition of strictosamide as the penultimate biosynthetic precursor assisted by carbon^{-13} and deuterium NMR spectroscopy. *J. Am. Chem. Soc.* **101** 3358–3369

Keat H T, Elizabeth G and McKnight T D 2000 Characterization and cloning of 10- hydroxyl geraniol oxidoreductase. *Plant Biol.* (rockville) **71**

Ketudat CJR and Esen A 2010 β-Glucosidases. *Cell. Mol. Life. Sci.* **67** 3389–3405

Kumari A, Heikham RS, Ashwani J, Mohit KS, Shankar R and Kumar S 2014 Transcriptome sequencing of rhizome tissue of *Sinopodophyllum hexandrum* at two temperatures. *BMC Genomics* **15** 871

Kutchan TM, Hampp N, Lottspeich F and Beyreuther K 1988 The cDNA clone for strictosidine synthase from *Rauvolfia serpentina*: DNA sequence determination and expression in *Escherichia coli. FEBS Lett.* **237** 40–44

Li W and Godzik A 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22** 1658–1659

Li W, Jaroszewsk L and Godzik A 2002 Sequence clustering strategies improve remote homology recognitions while reducing search times. *Protein Eng.* **15** 643–649

Li R, Zhu H, Ruan J, Qian W, Fang X, Zhongbin S, Yingrui L, Shengting L, *et al*. 2010 *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20** 265–272

Lopez-M M and Nessler CL 1997 Tryptophan decarboxylase is encoded by two autonomously regulated genes in *Camptotheca acuminata* which are differentially expressed during development and stress. *Plant J.* **11** 1167–1175

Lu H and McKnight TD 1999 Tissue-specific expression of the β-subunit of tryptophan synthase in *Camptotheca acuminata*, an indole alkaloid producing plant. *Plant Physiol.* **120** 43–52

McKnight TD, Roessner CA, Devagupta R, Scott A and Nessler CL 1990 Nucleotide sequence of a cDNA encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus. Nucleic Acids Res.* **16** 4939

Menke FLH, Champion A, Kijne JW and Memelink J 1999 A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor- inducible AP2-domain transcription factor, ORCA2. *EMBO J.* **18** 4455–4463

Michael W, Graham W and Dempfle L 2002 Relative Expression Software Tool (REST©) for group wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30** 36

Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, *et al*. 2014 The seco-iridoid pathway from Catharanthus roseus. *Nat. Commun.* **5** 3606

Morant M, Bak S, Moller BL and Werck-Reichhart D 2003 Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Curr. Opin. Biotechnol.* **14** 151–162

Nomura T, Quesada AL and Kutchan TM 2008 The new β-D-glucosidase in terpenoid-isoquinoline alkaloid biosynthesis in *Psychotria ipecacuanha*. *J. Biol. Chem.* **283** 34650–34659

O'Connor SE and Maresh JJ 2006 Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23** 532–547

Oberlies NH and Kroll DJ 2004 Camptothecin and taxol: historic achievements in natural products research. *J. Nat. Prod.* **67** 129–135

Ohta M, Matsui K, Hiratsu K, Shinshi H and Ohme-Takagi M 2001 Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell.* **13** 1959–1968

Pauw B, Hilliou FA, Martin VS, Chatel G, Wolf CJ, Champion A, Pré M, Van DB, *et al*. 2004 Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in *Catharanthus roseus*. *J. Biol. Chem.* **279** 52940–52948

Pertea G, Huang X, Lian F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, *et al*. 2003 TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics.* **19** 651–652

Ramesha BT, Amna T, Ravikanth G, Gunaga RP, Vasudeva R, Ganeshaiah KN, Uma Shaanker R, Khajuria RK, *et al*. 2008 Prospecting for Camptothecines from *Nothapodytes nimmoniana* in the Western Ghats, South India: identification of high-yielding sources of camptothecin and new families of camptothecines. *J. Chromatogr. Sci.* **46** 362–368

Ramesha BT, Suma HK, Senthilkumar U, Priti V, Ravikanth G, Vasudeva R, Kumar TR, Ganeshaiah KN, *et al*. 2013 New plant sources of the anti-cancer alkaloid, camptothecine from the Icacinaceae taxa, India. *Phytomedicin.* **20** 521–527

Riki K and Julia BS 2005 mRNA sequence features that contribute to translational regulation in *Arabidopsis*. *Nucleic Acids Res.* **33** 955–965

Sakai K, Shitan N, Sato F, Ueda K and Yazaki K 2002 Characterization of berberine transport into *Coptis japonica* cells and the involvement of ABC protein. *J. Exp. Bot.* **53** 1879–1886

Samira M, Vincent B and Benoit SP 2007 Cellular and sub-cellular organisation of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Phytochem. Rev.* **6** 363–381

Samuelsson G 1999 *Drugs of natural origin: a textbook of pharmacognosy* 4th ed. (Stockholm: Pharmaceutical Press)

Schmid R and Blaxter ML 2008 annot8r: GO, EC and KEGG annotation of EST datasets. *BMC Bioinf.* **9** 180

Shitan N, Bazin I, Dan K, Obata K, Kigawa K, Ueda K, Sato F, Forestier C, *et al*. 2003 Involvement of CjMDR1, a plant MDR-type ABC protein, in alkaloid transport in *Coptis japonica*. *Proc. Natl. Acad. Sci. USA* **100** 751–756

Singh K, Raizada J, Bhardwaj P, Ghawana S, Rani A, Singh H, Kaul K and Kumar S 2004 26S rRNA-based internal control gene primer pair for reverse transcription-polymerase chain reaction-based quantitative expression studies in diverse plant species. *Anal. Biochem.* **335** 330–333

Sirikantaramas S, Arthitaya M, Thanyada R, Hideyoshi F, Tyuji H, Hiroshi S, Mami Y and Kazuki Saito 2015 Structural insight of DNA topoisomerases I from camptothecinproducing plants revealed by molecular dynamics simulations. *Phytochemistry* (article in press)

Stockigt J and Ruppert M 1999 Strictosidine: the biosynthetic key to monoterpenoid indole alkaloids; in: *Comprehensive natural products chemistry: amino acids, peptides, porphyrins and alkaloid* (eds) DHR Barton, K Nakanishi, O Meth-Cohn and JW Kelly (Amsterdam: Elsevier) **4** pp 109–138

Suhas S, Ramesha BT, Ravikanth G, Gunaga Rajesh P, Vasudeva R, Ganeshaiah KN and Uma Shaanker R 2007 Chemical profiling of *Nothapodytes nimmoniana* populations in the Western Ghats, India for anti-cancer compound, camptothecin. *Curr. Sci.* **92** 1142–1147

Tafur S, Nelson JD, DeLong DC and Svoboda GH 1976 Antiviral components of *Ophiorrhiza mungos*. Isolation of camptothecin and 10-methoxycamptothecin. *Lloydia.* **39** 261–262

Terasaka K, Sakai K, Sato F, Yamamoto H and Yazaki K 2003 *Thalictrum minus* cell cultures and ABC-transporter. *Phytochemistry.* **62** 483–489

Uma Shaanker R, Ramesha BT, Ravikanth G, Gunaga R, Vasudeva R and Ganeshaiah KN 2008 Chemical profiling of *Nothapodytes nimmoniana* for camptothecin, an important anticancer alkaloid: toward the development of a sustainable production system; in *Bioactive molecules and medicinal plants* (eds) KG Ramawat and JM Merillon (UK: Springer Publishing) pp 197–213

Van der Fit L and Memelink J 2000 ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science* **289** 95–297

Vinogradov AE 2003 DNA helix: the importance of being GC-rich. *Nucleic Acids Res.* **31** 838–1844

Wall ME, Wani MC, Cook CE, Palmer KH, Mcphail AT and Sim GA 1966 Plant antitumor agents I. The isolation and structure of camptothecin, a novel alkaloidal leukemia and tumor inhibition from *Camptotheca acuminate*. *J. Am. Chem. Soc.* **88** 3888–3890

Warzechaa H, Gerasimenkoa I, Kutchan TM and Stöckigt J 2000 Molecular cloning and functional bacterial expression of a plant glucosidase specifically involved in alkaloid biosynthesis. *Phytochemistry* **54** 657–666

Yamazaki Y, Urano A, Sudo H, Kitajima M, Takayama H, Yamazaki M, Aimi N and Saito K 2003 Metabolite profiling of alkaloids and strictosidine synthase activity in camptothecin producing plants. *Phytochemistry.* **62** 461–470

Yao H, Gong Y, Zuo K, Ling H, Qiu C, Zhang F, Wang Y, Pi Y, *et al*. 2008 Molecular cloning, expression profiling and functional analysis of a DXR gene encoding 1-deoxy-d-xylulose 5-phosphate reductoisomerase from *Camptotheca acuminate*. *J. Plant Physiol.* **165** 203–213

Yu F and De Luca V 2013 An ABC transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proc. Natl. Acad. Sci.* **110** 15830–15835

MS received 14 July 2015; accepted 16 December 2015

Corresponding editor: Utpal Nath