FEBS Letters 580 (2006) 4501–4507

Expressed sequence tags from Madagascar periwinkle (*Catharanthus roseus*)

Jun Murata[a], Dorothee Bienzle[b], Jim E. Brandle[c], Christoph W. Sensen[d], Vincenzo De Luca[a,*]

[a] *Department of Biological Sciences, Brock University, 500 Glenridge Avenue, St. Catharines, Ont., Canada L2S3A1*
[b] *Ontario Veterinary College, University of Guelph, Guelph, Ont., Canada N1G2W1*
[c] *Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, 1391 Sandford Street, London, Ont., Canada N5V4T3*
[d] *University of Calgary, Faculty of Medicine, Department of Biochemistry and Molecular Biology, 3330 Hospital Drive NW, Calgary, Alta., Canada T2N4N1*

Received 2 June 2006; revised 29 June 2006; accepted 7 July 2006

Edited by Michael R. Sussman

Abstract The Madagascar periwinkle (*Catharanthus roseus*) is well known to produce the chemotherapeutic anticancer agents, vinblastine and vincristine. In spite of its importance, no expressed sequence tag (EST) analysis of this plant has been reported. Two cDNA libraries were generated from RNA isolated from the base part of young leaves and from root tips to select 9824 random clones for unidirectional sequencing, to yield 3327 related sequences and 1696 singletons by cluster analysis. Putative functions of 3663 clones were assigned, from 5023 non-redundant ESTs to establish a resource for transcriptome analysis and gene discovery in this medicinal plant.

Keywords: Expressed sequence tags; Secondary metabolism; *Catharanthus roseus*

1. Introduction

Catharanthus roseus (Madagascar periwinkle), which is a member of *Apocynaceae* (dogbane) family, produces various monoterpenoid indole alkaloids (MIAs), some of which are valuable for their medical applications. In particular, the bisindole MIAs, vinblastine and vincristine that are derived by the oxidative coupling of vindoline and catharanthine, have been used for cancer chemotherapy. This subtropical plant is also known for its drought tolerance, and many horticultural varieties with various petal colors have been developed. The estimated haploid genome size of *Catharanthus* varies between 696 Mbp and 2377 Mbp, depending on the reference DNA used for analyses [1,2], and this is typical among other *Apocynaceae* plants (average 1633 Mbp among 8 species) (also see Plant DNA C-values Database, http://www.rbgkew.org.uk/cval/homepage.html). These numbers are larger than those of *Arabidopsis* (125 Mbp) and rice (497 Mbp), but relatively smaller than maize (2671 Mbp) or most of the angiosperms including those from members of the *Solanaceae* (average 2779 Mbp among 175 species) and *Papaveraceae* families (average 2929 Mbp among 30 species).

Recent biochemical and molecular approaches to study MIA biosynthesis in *Catharanthus* have led to the discovery of more than ten pathway enzymes and their corresponding genes [3]. Efforts have also focused on overexpression of MIA biosynthetic genes in cell culture systems in order to produce a stable and economical supply of commercially relevant MIAs, but no consistent reports of vindoline accumulating cell cultures have appeared [4]. Vindoline negative cell cultures have partly been attributed to the multiple cell types that appear to be involved in MIA biosynthesis in the leaf [5]. In order to completely characterize this pathway, more than 20 genes remain to be cloned, including those encoding the enzymes that catalyze three out of the last six enzyme steps of vindoline biosynthesis [4]. Clarification of the biochemistry and molecular biology of these remaining steps will provide the necessary information for metabolic engineering of vindoline in tabersonine accumulating plant cell cultures. Since MIA biosynthesis involves many unique enzymatic steps, it has been difficult to isolate the corresponding genes by sequence identity-based PCR cloning or by biochemical approaches, as illustrated by the problems associated with the cloning of *16-hydroxytabersonine-16-O-methyltransferase* (*16OMT*) that converts 16-hydroxytabersonine to 16-methoxytabersonine [6]. In addition, only 3111 nucleotide sequences have been annotated from the whole *Apocynaceae* family, including 372 nucleotide sequences from *Catharanthus*, in the publicly accessible GenBank database as of April 2006 (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi). Those numbers are much smaller compared to the data available for families composed of commercially more important food crops (911 192 from *Solanaceae*, 882 210 from *Brassicaceae*, excluding *Arabidopsis*, and 250 714 from *Vitaceae*). This information and our interest in *Catharanthus* biochemistry triggered efforts to construct cDNA libraries from selected *Catharanthus* tissues in order to generate a sufficient number of expressed sequence tags (ESTs) for the development of microarray based transcriptome analysis and gene discovery in this medicinal plant. This report describes the first example of structural EST analysis from *Catharanthus*, a representative member of the *Apocynaceae* family.

---

[*] Corresponding author.
E-mail addresses: jmurata@brocku.ca (J. Murata), dbienzle@uoguelph.ca (D. Bienzle), brandleje@agr.gc.ca (J.E. Brandle), csensen@ucalgary.ca (C.W. Sensen), vdeluca@brocku.ca (V. De Luca).

Abbreviations: EST, expressed sequence tag; BLAST, basic local alignment search tool; MIA, monoterpenoid indole alkaloid

0014-5793/$32.00 © 2006 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2006.07.020

2. Materials and methods

2.1. cDNA library construction

Catharanthus roseus (cv. *Little Delicata*) plants were grown in a greenhouse using a 16/8 h photoperiod. Two grams from the base part of young 1.5 cm long leaves (composed of the petiole, including 1/3 of the leaf) and 2 g of root tips (3–10 mm in length) from a *Catharanthus* hairy root culture [7] were harvested and extracted for total RNA, followed by poly(A) + RNA purification as described previously [8]. The poly(A) + RNA from each tissue was converted to cDNA using the ZAP cDNA synthesis kit (Stratagene), and the resulting cDNA was unidirectionally subcloned into *Eco*RI and *Xho*I sites within the multiple cloning site in the phage vector, and packaged by Gigapack III Gold packaging extract (Stratagene). Bacterial clones were obtained for random sequencing by *in vivo* excision according to the manufacturer's protocol, and single colonies were transferred to 96-well microtiter plates for automated plasmid preparation. Twenty randomly chosen plasmid clones from each library were digested with *Eco*RI and *Xho*I restriction enzymes for agarose gel electrophoresis analysis to check the insertion rates and the average insert length. Sequencing of cDNA inserts was conducted using the ABI Prism Big Dye terminator sequencing kits (Applied Biosystems) and an AB 3730 genetic analyzer (Applied Biosystems).

2.2. Sequence analysis

A total number of 9824 random clones were chosen from the libraries and submitted to unidirectional sequencing from the 5′ end using T7 primer. The MAGPIE automated analysis and annotation system [9] was used to cluster and analyze the sequences, while the automated putative functions of genes were verified manually. The results of the analysis are available via http://magpie.ucalgary.ca/ [9]. The list of sequences is also available as an Excel-formatted table, with hyperlinks inserted, as supplementary material.

2.3. Tissue fixation and embedding, laser capture microdissection, RNA extraction and processing

Catharanthus roseus tissues were prepared, processed for laser capture microdissection (LCM) and extracted for RNA from the entire young leaf section (W, whole leaf) epidermal (E) leaf cells, mesophyll (M) cells, idioblast (I) cells, laticifer (L) cells and vasculature (V) cells as described previously [10].

2.4. RT-PCR analysis

The mRNA levels of selected *Catharanthus O-methyltransferase* genes in different *Catharanthus* leaf cell types were analyzed by RT-PCR using gene specific oligonucleotides as follows; CrOMT2 left primer 5′-GGCGGAAGATAAGTATGGTA-3′, CrOMT2 right primer 5′-CCAGTCATGAAGAATCCACT-3′, CrOMT4 left primer 5′-ACTGCTGAGATTCGTAAAGC-3′, CrOMT4 right primer 5′-TACGCATTAGTCGGTGTATG-3′, CrOMT6 left primer 5′-CCATGGCTAATGACTCTG-3′, CrOMT6 right primer 5′-GTCTCCTCCACAAACTC-3′, OOMT-like RT01 5′-ATGATTGGTCCGACAAGGAG-3′, OOMT-like RT02 5′-TCCAGTCACCAAAACCATCA-3′, CrActin-RT01 5′-GGCTGGATTTGCTGGAGATGAT-3′, CrActin-RT02 5′-TAGATCCTCCGATCCAGACACTG-3′. Reverse transcription was performed using RNA PCR Kit (AMV) ver. 3.0 (Takara, Otsu, Japan), according to the manufacturer's protocol. The PCRs were carried out for 35 cycles of for 15 s at 94 °C, 20 s at 57 °C and 30 s at 72 °C. By testing every three cycles between 26 and 38 cycles, the program for PCR was optimized considering the linearity of the amount of PCR products. Amplified cDNA fragments were run on 2.0% agarose gel, and visualized by ethidium bromide staining. For the verification of the existence of known MIA biosynthesis-related genes in the cDNA libraries, the phage libraries from leaf base and root tip were directly used as templates for the PCR analysis using oligonucleotides as follows; CrG10H-RT01 5′-GGTAGCCTCACGATGGAGAA-3′, CrG10H-RT02 5′-CCTTGGCAGAATCCGAATAA-3′, CrSLS-RT01 5′-CTTTGAGGGTGCAAAATGGT-3′, CrSLS-RT02 5′-TGGGATCCTTGTTTTTCAGC-3′, CrTDC-RT01 5′-CGCCTGTATATGTCCCGAGT-3′, CrTDC-RT02 5′-GTTGCGATTTGCCAATTTTT-3′, CrSTR-RT01 5′-ACCATTGTGTGGGAGGACAT-3′, CrSTR-RT02 5′-CCATTTGAATGGCACTCCTT-3′, CrSGD-RT01 5′-ATTTGCACCAGGAAGAGGTG-3′, CrSGD-RT02 5′-TATGAACCATCCGAGCATGA-3′, CrT16H-RT01 5′-GCTTCATCCACCAGTTCCAT-3′, CrT16H-RT02 5′-CCGGACATATCCTTCTTCCA-3′, CrD4H-RT01 5′-TTGACATTTGGGACAAGCAA-3′, CrD4H-RT02 5′-CCAAAAGCAACAGCAACAGA-3′, CrDAT-RT01 5′-GTGCGTATCCGTTGGTTTCT-3′, CrDAT-RT02 5′-CGAACTCAATTCCATCGTCA-3′, CrMAT-RT01 5′-XXXXXX-3′, CrMAT-RT02 5′-XXXXXXXX-3′, CrActin-RT01 5′-GGCTGGATTTGCTGGAGATGAT-3′, CrActin-RT02 5′-TAGATCCTCCGATCCAGACACTG-3′. The PCRs were carried out for 35 cycles of for 15 s at 94 °C, 20 s at 57 °C and 30 s at 72 °C.

3. Results and discussion

3.1. Sequencing and establishing ESTs of C. roseus

Two cDNA libraries were constructed from the base part (composed of the petiole, including 1/3 of the leaf) of young leaves (1.5 cm), and of root tips (3–10 mm in length), respectively, and 9824 randomly chosen clones were partially sequenced. Cluster analysis produced 3327 clustered sequences that were assembled from more than one EST, and 1696 singletons that appeared only once in the random sequencing to yield a total of 5023 unique genes (44.7% from leaf base library and 55.3% from root tip library) from *Catharanthus* (CrUniGene). The average sequence length of clustered sequences and singletons was 592.73 bp. Genes in the CrUniGene set were annotated according to their sequence identity to sequences in the GenBank database using the BLASTX program [11] through the MAGPIE gene annotation pages system (http://magpie.ucalgary.ca/). The results showed that 67.1% (3363 ESTs) had significant sequence similarity to genes with assigned functions, 18.7% (940 ESTs) showed sequence similarity to genes with unknown function and 14.2% (715 ESTs, average length 424 bp) clearly encoded Open Reading Frames of putative *Catharanthus* proteins with no sequence similarity to any genes in the GenBank database (Fig. 1). After the automated annotation, their validity was checked by visual inspection and the 3363 ESTs were further classified into 13 categories according to their putative functions as shown in Fig. 2. Of those 3363 ESTs, very few (approximately 0.5%, data not shown) have identical sequences to genes from *Catharanthus* in the PubMed database, and confirm that the vast majority of the ESTs in our database are quite unique. The largest gene category in CrUniGene was 'Enzyme' (798 genes), which contains enzymes involved in primary and secondary metabolism, representing approximately 23.7% of functionally categorized ESTs. This number is similar to those found in other plant EST analyses including those of *Arabidopsis* (25%) [12], *Stevia rebaudiana* (21%) [13], *Lotus japonicus* (21%) [14] and oil palm (20%) [15]. The second largest category was 'Translation' (19.3%), followed by Miscellaneous (14.0%), Chaperone (10.0%), Transport (7.7%), Structure (6.8%) and Transcription (6.3%). These results illustrated that the global gene expression profiles described for *Catharanthus* were comparable to the ones from other plants in terms of categories of gene functions.

3.2. CrUniGene as a source for gene discovery

3.2.1. CrUniGene set contains various novel sequences. Initially, we were interested in obtaining ESTs encoding novel enzymes involved in MIA biosynthesis including the three late stage vindoline biosynthesis genes encoding 16OMT, a biochemically uncharacterized hydration and a 16-methoxy 2,3-dihydro 3-hydroxy tabersonine *N*-methyltransferase (NMT).

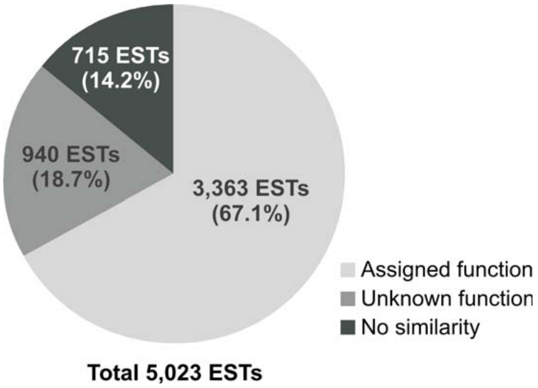

Fig. 1. Approximately 67.1% of the 5023 non-redundant *Catharanthus* ESTs (CrUniGenes) show sequence similarity to genes that have putative biochemical functions, whereas 18.7% of ESTs have sequence similarity to genes with unknown function and 14.2% of ESTs did not show significant sequence similarity to any genes in GenBank.

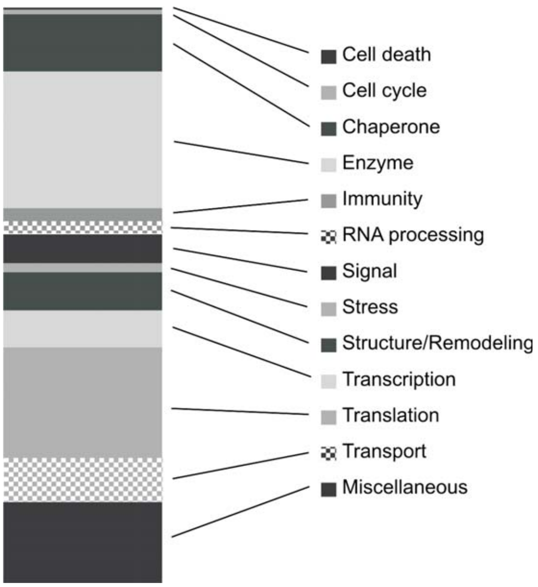

Fig. 2. CrUniGenes were functionally categorized into 13 putative categories. Percentages are given compared to the CrUniGenes which have sequence similarity to genes with putative biochemical functions.

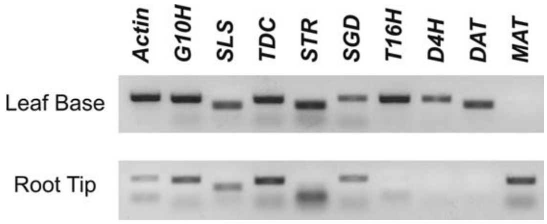

Fig. 3. Comparative representation of 8 functionally characterized MIA biosynthesis genes in leaf base and root cDNA libraries. The PCRs were carried out using specific primers to various MIA biosynthesis genes. G10H, geraniol 10-hydroxylase; SLS, secologanin synthase; TDC, tryptophan decarboxylase; STR, strictosidine synthase; SGD, strictosidine β-glucosidase; T16H, tabersonine 16-hydroxylase; D4H, desacetoxyvindoline 4-hydroxylase; DAT, deacetylvindoline *O*-acetyltransferase; MAT, minovincinine 19-*O*-acetyltransferase. Actin was used as an internal control.

In spite of the high MIA biosynthetic activity of the selected *Catharanthus* tissues [5] and verification by PCR analysis that functionally characterized MIA biosynthesis genes were represented in the cDNA libraries from young leaf base (Fig. 3: *G10H*, *SLS*, *TDC*, *STR*, *SGD*, *T16H*, *D4H* and *DAT*) and from root tip (Fig. 3: *G10H*, *SLS*, *TDC*, *STR*, *SGD* and *MAT*), the CrUniGene set did not include any known genes involved in MIA biosynthesis except for *geraniol 10-hydroxylase* (*G10H*) (Table 1) and *strictosidine* β*-glucosidase* (*SGD*). Also, the specificity of the two cDNA libraries was maintained in the cDNA libraries because *deacetylvindoline O-acetyltransferase* (*DAT*) and *minovincinine 19-O-acetyltransferase* (*MAT*) were detected only from leaf base and root tip in the cDNA library, respectively (Fig. 3). The absence of MIA biosynthesis ESTs in the CrUniGene set suggests that expressed MIA biosynthetic genes are rare compared to other non-MIA-related genes. The results obtained were still quite remarkable since obtained 5023 CrUniGene sequences represent roughly 20% of the estimated 25 500 genes from *Arabidopsis* [16]. While the genome size of *Arabidopsis* is speculated to be 6–19 times smaller than that of *Catharanthus*, the CrUniGene set described here represents a considerable new resource for gene discovery in this important medicinal plant.

Out of 798 genes in the Enzyme category, at least 71 sequences showed significant sequence identity (E values lower than 1×10^{-6}) to known enzymes involved in secondary metabolism. This does not necessarily mean that the number of enzymes related to secondary metabolism is limited to 8.9% out of the genes in this category, since a similar number of additional genes also have limited similarity to secondary metabolism enzymes. This analysis suggested that while various genes could be candidates for MIA biosynthesis, this is not likely since only *G10H* and *SGD* in this pathway could be identified. Some novel features of these genes are described in the following sections.

3.2.2. Cytochrome P450-dependent monooxygenases. Cytochrome P450-dependent monooxygenases (CYPs) are a typically large gene family in higher plants, with up to 286 genes listed in this class of enzyme in the *Arabidopsis* genome [16]. Although several *Arabidopsis* CYPs have been characterized and were shown to require NADPH- and O_2 to catalyze hydroxylations of various metabolites, the biochemical function of most of these genes remains unknown. The CrUniGene set contains three genes with substantial sequence identity to known plant CYPs; *N-methylcoclaurine 3′ hydroxylase* from *Coptis japonica*, *p-coumaroyl 5-O-shikimic acid 3′ hydroxylase* from *Ocimum basilicum* and *flavonoid 3′5′ hydroxylase* (*F3′5′H*) from *Arabidopsis* [17–19] (Table 1). Interestingly, *F3′5′H* of *Catharanthus* [20] had lower sequence identity to this new CYP than the *Arabidopsis* gene, suggesting it may be involved in related hydroxylase activity. Another 12 ESTs showed some sequence identity to uncharacterized putative CYPs, but none of the 13 CYPs in CrUniGene had exact sequences matches with *Catharanthus* CYPs in the publicly accessible database. While most of the CYPs in CrUniGene set are quite novel they probably are not candidates for several hydroxylases in MIA biosynthesis such as *7-deoxyloganic acid 7-hydroxylase* that has yet to be cloned [21].

3.2.3. S-adenosyl L*-methionine-dependent methyltransferases.* Methylation is one of the key modifications to change

Table 1
CrUniGenes with significant sequence identity to cytochrome P450-dependent monooxygenases

CrUniGene ID	GI	BLAST homology search	Score and E value
CrUniGene03754	AAL99201	*p*-Coumaroyl shikimate 3′-hydroxylase isoform 2 [*Ocimum basilicum*]	551 bits (1421), Expect = e − 156
CrUniGene01144	AAD48912	Aldehyde 5-hydroxylase [*Liquidambar styraciflua*]	276 bits (706), Expect = 3e − 73
CrUniGene03002	CAB85635	Putative ripening-related P-450 enzyme [*Vitis vinifera*]	256 bits (655), Expect = 2e − 67
CrUniGene00979	CAC27827	Geraniol 10-hydroxylase [*Catharanthus roseus*]	252 bits (643), Expect = 8e − 66
CrUniGene04808	NP_196623	Cytochrome P450 family protein [*Arabidopsis thaliana*]	247 bits (631), Expect = 2e − 64
CrUniGene04102	NP_568463	Cytochrome P450 family protein [*Arabidopsis thaliana*]	237 bits (604), Expect = 1e − 61
CrUniGene02444	NP_917794	Putative cytochrome P450 [*Oryza sativa* (japonica cultivar-group)]	231 bits (588), Expect = 2e − 59
CrUniGene04738	NP_192970	Cytochrome P450 family protein [*Arabidopsis thaliana*] F3′5′H	219 bits (559), Expect = 1e − 56
CrUniGene04086	P37121	Cytochrome P450 76A1 (CYPLXXVIA1) (P-450EG8)	218 bits (556), Expect = 5e − 56
CrUniGene00330	BAB12433	(*S*)-*N*-methylcoclaurine-3′-hydroxylase [*Coptis japonica*]	216 bits (549), Expect = 4e − 55
CrUniGene02799	CAC27827	Cytochrome P450 [*Catharanthus roseus*]	161 bits (408), Expect = 6e − 39
CrUniGene04438	NP_176882	Cytochrome P450 [*Catharanthus roseus*]	116 bits (291), Expect = 2e − 25
CrUniGene04276	AAO32822	Cytochrome P450 71D1 [*Catharanthus roseus*]	101 bits (252), Expect = 5e − 21

the properties of metabolites by attaching *S*-adenosyl L-methionine (SAM)-derived methyl group to various functional groups, such as hydroxyl and carboxyl, that would be protected from unwanted further modification at the particular position. In other cases, methylation helps the metabolite to be volatilized simply by decreasing its polarity. There are at least three methyltransferases (MTs) involved in MIA biosynthesis, namely loganic acid *O*-methyltransferase and 16OMT and NMT [22,23]. There are 12 putative MTs in CrUniGene set that do not have exact matches with known MTs (Table 2). One interesting example is an EST with high sequence identity to both orcinol *O*-methyltransferase from roses [24,25] and the recently characterized 3,5-dimethoxyphenol *O*-methyltransferase from *Ruta graveolens* [26]. The different substrate specificities observed with these biochemically characterized clones [24–26] suggests that the *Catharanthus* EST could also accept a different substrate than those previously described [24–26] and related to its particular have biochemical function.

3.2.4. UDP-glucose-dependent glycosyltransferases. Glycosylation attaches a bulky and polar sugar moiety to the substrate that often leads to changes in its stability and solubility. The terpenoid part of MIAs is derived from a glucoside, secologanin that condenses with tryptamine in the presence of strictosidine synthase to produce strictosidine. The sugar moiety of strictosidine is removed by strictosidine β-glucosidase to yield an unstable aglycone that is a central precursor to more than 200 MIAs in *Catharanthus* [27]. Originally, the sugar moiety came from 7-deoxyloganic acid, which is glucosylated by an uncharacterized glucosyltransferase (GT). While none of the five putative GTs in the CrUniGene set show any striking sequence identity to known GTs (Table 3), it remains to be determined what biochemical pathways they are involved in.

3.2.5. Transporters. The fact that multiple cell types are engaged in vindoline biosynthesis in *Catharanthus* implies that one or more intermediate(s) must be transported in a controlled manner from one cell type to another. ATP-binding

Table 2
CrUniGenes with significant sequence identity to alkaloid/flavonoid methyltransferases

CrUniGene ID	GI	BLAST homology search	Score and E value
CrUniGene01389	NP_195131	Caffeoyl-CoA 3-*O*-methyltransferase, putative [*Arabidopsis thaliana*]	473 bits (1217), Expect = e − 132
CrUniGene00837	AAM97497	Flavonoid *O*-methyltransferase (CrOMT2) [*Catharanthus roseus*]	459 bits (1181) Expect = 5e − 128
CrUniGene01076	AAM23004	Orcinol *O*-methyltransferase (OOMT) [Rosa hybrid cultivar]	456 bits (1174), Expect = e − 127
CrUniGene00835	AAM97498	*O*-methyltransferase (CrOMT4) [*Catharanthus roseus*]	373 bits (958) Expect = 1e − 107
CrUniGene00836	BAD07529	Putative gamma-tocopherol methyltransferase [*Oryza sativa*]	296 bits (758), Expect = 4e − 79
CrUniGene03706	AAN61072	*O*-methyltransferase [*Mesembryanthemum crystallinum*]	293 bits (749), Expect = 5e − 78
CrUniGene02879	BAD07529	Putative gamma-tocopherol methyltransferase [*Oryza sativa*]	291 bits (745), Expect = 2e − 77
CrUniGene03423	NP_181669	Embryo-abundant protein-related [*Arabidopsis thaliana*]	280 bits (716), Expect = 4e − 74
CrUniGene04591	A86285	Protein F9L1.6 [imported] – *Arabidopsis thaliana*	181 bits (460), Expect = 1e − 44
CrUniGene02171	BAC57960	Phosphoethanolamine *N*-methyltransferase [*Aster tripolium*]	179 bits (453), Expect = 3e − 44
CrUniGene03704	AAN61072	*O*-methyltransferase [*Mesembryanthemum crystallinum*]	167 bits (423), Expect = 9e − 41
CrUniGene03705	Q43161	Caffeoyl-CoA *O*-methyltransferase (CCoAMT) (CCoAOMT)	149 bits (375), Expect = 4e − 35

Table 3
CrUniGenes with significant sequence identity to UDP-glycosyltransferases

CrUniGene ID	GI	BLAST homology search	Score and E value
CrUniGene04543	BAB86920	Glucosyltransferase-2 [*Vigna angularis*]	355 bits (910), Expect = 8e − 97
CrUniGene01173	NP_173653	UDP-glucoronosyl/UDP-glucosyl transferase family protein	327 bits (839), Expect = 1e − 88
CrUniGene04499	BAB86922	Glucosyltransferase like protein [*Vigna angularis*]	187 bits (476), Expect = 1e − 46
CrUniGene00503	XP_469832	Putative glucosyltransferase [*Oryza sativa*]	165 bits (418), Expect = 8e − 40
CrUniGene03954	NP_174569	Galactosyltransferase family protein [*Arabidopsis thaliana*]	97.4 bits (241), Expect = 5e − 20

Table 4
CrUniGenes with significant sequence identity to ABC-transporters

CrUniGene ID	GI	BLAST homology search	Score and E value
CrUniGene03068	CAA05625	AtMRP4 [*Arabidopsis thaliana*]	362 bits (930), Expect = 6e − 99
CrUniGene00174	NP_564404	ATP-binding-cassette transporter, putative [*Arabidopsis thaliana*]	229 bits (584), Expect = 3e − 59
CrUniGene00554	AAT85292	ABC transporter ATP-binding protein, putative [*Oryza sativa*]	191 bits (486), Expect = 9e − 48
CrUniGene04406	XP_469804	Putative ABC (ATP-binding cassette) transporter transmembrane	83.6 bits (205), Expect = 2e − 15

cassette (ABC)-type transporter proteins are the possible key players for this transport system as shown by their involvement in the development of multidrug resistance against vinblastine in certain types of mammalian cancer cells. In plants, several ABC-type transporters have recently been characterized, including the CjMDR1 transporter from *C. japonica* that functions as an influx pump for berberine transport through the plasma membrane [28], and the AtMDR1 transporter from *Arabidopsis* that is involved in polar auxin transport [29]. A total of 225 ESTs were grouped into the Transport category including various ion channels and vesicle transport-related proteins as well as four novel ABC-type transporters (Table 4).

3.2.6. Other secondary metabolism genes. There are two dioxygenase-like genes and six acyltransferase-like genes in CrUniGene set, both of which are important classes of enzymes that might be involved in MIA biosynthesis.

3.2.7. Transcription factors. Among 213 sequences classified in the Transcription category, 106 showed significant sequence similarity to transcription factors that included zinc finger proteins, bHLH- and MYB-type proteins (Table 5). Three transcription factors from *Catharanthus* have been cloned and characterized, including AP2/ERF-type factor ORCA3 that activates mRNA expression of multiple MIA biosynthetic pathway genes in ORCA3 overexpressing cell cultures [30], CrMYC1 that may be involved in *STR* expression [31] and Zn finger ZCT protein may repress *tryptophan decarboxylase* (*TDC*) and *strictosidine synthase* (*STR*) gene expression [32]. The sequences in the transcription category, while novel to *Catharanthus*, did not include *ORCA*, *ZCT* or *MYC* genes known to be involved in MIA biosynthesis [30,31].

3.3. Laser capture microdissection-guided expression screening

As part of the effort to use the CrUniGene set to develop a rapid screening method for candidate genes involved in MIA biosynthesis, laser capture microdissection (LCM) was used to obtain mRNA from epidermal (E) cells, mesophyll (M) cells, idioblast (I) cells, laticifer (L) cells and vasculature (V) cells, as well as from whole young leaf (W) for RT-PCR analysis of genes of interest at the cellular level. Several lines of evidence indicate that leaf epidermal cells are the major site for the biosynthesis of MIA alkaloids leading to the formation of 16-methoxytabersonine, an intermediate that is eventually converted into vindoline via 4 subsequent enzyme reactions [5,9,33,34]. A candidate gene for *16OMT*, for example, is expected to be expressed preferentially in leaf epidermal cells, since 16OMT enzyme activity was predominantly detected from leaf epidermis [10]. LCM-based mRNA isolation coupled to RT-PCR analysis was used here to show that the *orcinol-O-methyltransferase* (*OOMT*)-like gene found in CrUniGene set was mostly expressed in epidermal cells compared to other *O-methyltransferases* that appear to be involved in flavonoid/anthocyanin reactions (CrOMT2; CrOMT6) or that have no known function (CrOMT4) (Fig. 4). Although there is no evidence that *Catharanthus* produces orcinol, it is noteworthy that functional *OOMT* in rose was recently shown to be expressed in epidermal cells of petals as expected from its function for the production of 3,5-dimethoxytoluene, a major scent compound of rose [35].

Table 5
List of putative transcription factors found in CrUniGene set

AP2/ERF	9
AT	2
bHLH	11
bZIP	8
CCAAT	5
HMG	3
Homeodomain	2
MADS box	1
MYB	7
NAC	6
Other	17
WRKY	3
Zn finger	32
Total	106

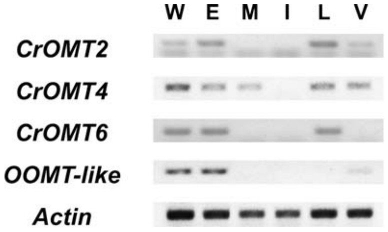

Fig. 4. Epidermis localized expression of *Orcinol-O-methyltransferase* (*OOMT*)-like gene identified in the CrUniGene set from the leaf base library. The relative expression of selected genes encoding *O*-methyltransferases (CrOMT2, flavonol/anthocyanidin-3′,5′-*O*-methyltransferase ([6] and identified in CrUniGene set from root tip library); CrOMT4, *O*-methyltransferase with unknown function ([6] and identified in CrUniGene set from root tip library) and CrOMT6, flavonoid 4′-*O*-methyltransferase [39] were compared to the expression of the *OOMT*-like gene in different LCM captured cell types [epidermal (E) cells, mesophyll (M) cells, idioblast (I) cells, laticifer (L) cells and vasculature (V) cells, together with the RNA from the entire leaf section (W, whole leaf)]) by RT-PCR analysis as described in [10]. The data shown is representative of the results obtained in 3 separate experiments. *Actin* was used as an internal control.

4. Conclusions

This report describes the first EST analysis of sequences obtained from *C. roseus*, a member representative of the *Apocynaceae* family. The 5023 non-redundant ESTs from the CrUniGene set increased by 15-fold the number of nucleotide sequences from *Catharanthus* available in the GenBank database. While functional analyses are essential to confirm the validity of the annotations, the present study has identified several candidate genes that should be further characterized for their involvement in MIA biosynthesis but more likely in other secondary metabolism pathways. The expanded EST sequencing from *Catharanthus* adds a species from *Apocynaceae* family to perform comparative genomics between various plant species. The data produced here represents an initial step for future cDNA microarray experiments and for broader transcriptome analysis in order to take advantage of *Catharanthus* as a model plant to study secondary metabolism.

One alternative to obtain information of expressed sequences is cDNA-AFLP analysis. A recent report on combined transcriptome and metabolome analysis of *Catharanthus* cell cultures showed the potential of this type of approach in studying MIA biosynthesis [36]. However, the 417 sequences that were produced by a PCR-based differential display technique tended to be rather short (166 bp in average) and may explain why the majority of the sequences showed relatively low similarities to known genes. The results obtained in this study [36] suggest that, while this approach is valid, higher quality cDNA libraries need be established to produce an EST database that would produce more detailed transcriptome analysis. It is noteworthy that the percentage of the overlapping genes between CrUniGene in this study and CR tags [36] is approximately 1.7% to CrUniGene (86/5023 unique sequences) and 22% to CR tags (86/388 unique sequences). This is not surprising, since the cell culture does not seem to express genes that are related to cell differentiation or photosynthesis at high levels and cell cultures do not appear to express the late steps of vindoline biosynthesis [27]. While the CR tag sequences contained *DXS*, *MECS*, *HDS*, *GPPS*, *G10H*, *SGD* and *ORCA3* gene fragments, it is important to note that virtually no MIA-specific biosynthetic genes including *strictosidine synthase* (*STR*), *secologanin synthase* (*SLS*), *tryptophan decarboxylase* (*TDC*) and *tabersonine 16-hydroxylase* (*T16H*) as well as *desacetoxyvindoline 4-hydroxylase* (*D4H*) and *deacetylvindoline O-acetyltransferase* (*DAT*) were present in this small set of sequences. The data obtained with the CrUniGene and CR tag set expose the difficulties in identifying genes that might be constitutively expressed at relatively low levels through these approaches.

Nevertheless representation of 8 functionally characterized MIA biosynthesis genes in leaf base and root cDNA libraries (Fig. 3) suggests that these as well as the rest of the MIA biosynthesis pathway should appear as ESTs with sufficient sequencing of these two libraries. A more direct and economical approach involving the construction of cell type-specific libraries from *Catharanthus* would be a straight forward and quite useful approach to identify alkaloid specific pathways in particular cells [5,32–34]. Unfortunately, our effort to construct cell type-specific cDNA libraries from RNA samples obtained by laser capture microdissection (LCM) did not yield decent cDNAs with high quality in terms of average length and insertion rates. Indeed, to our knowledge, there has been only one report [37] on random sequencing of cDNA libraries constructed through such a procedure in rice that highlighted this promising technique, but only provided limited information due to technical difficulties. The rice cDNA libraries from LCM captured materials were randomly sequenced to obtain approximately 332 (44.6%) out of 745 randomly selected clones with inserts shorter than 150 bp [37]. In mammalian systems LCM has been used effectively together with microarray screening technologies to identify markers of various disease states, rather than as a tool for obtaining materials for random sequencing [38]. In our recent studies, an alternate novel technique for harvesting RNA from leaf epidermis was developed [10]. This technique (carborundum abrasion: CA) involves the use of carborundum to abrade the surface of fresh leaves and has been used successfully to harvest epidermis enriched metabolites for metabolic profiling, biologically active enzymes for biochemical assays and high quality mRNA for construction of an epidermis enriched cDNA library. Such a high quality EST database established form leaf epidermis, the major site of MIA biosynthesis in *Catharanthus*, should include most of the genes in MIA biosynthetic pathway and transcriptome analysis could identify genes that are involved in uncharacterized enzymatic steps in MIA biosynthesis. Sequences of all the ESTs in CrUniGene will be available through MAGPIE gene annotation pages (http://magpie.ucalgary.ca/).

Acknowledgements: V.D.L. holds a Canada Research Chair in Plant Biotechnology. This work was supported by Strategic (Peter Facchini and V.D.L.) and Discovery (V.D.L.) grants from the Natural Sciences and Engineering Research Council of Canada, respectively. The MAGPIE development is funded by Genome Canada and Genome Alberta, with support from the Province of Alberta. C.W.S. is the iCORE chair in Applied Bioinformatics. The assistance of Alex Richman with DNA preparation (Agriculture Canada), Ida van Grinsven (Agriculture Canada) with sequencing and Paul Gordon (University of Calgary) with organizing our Catharanthus sequences in the MAGPIE database is gratefully acknowledged. We thank the DNA technologies unit, Plant Biotechnology Institute, National Research Council of Canada for performing the sequencing of the ESTs.

References

[1] Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P. and Firoozabady, E. (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 220, 1049–1051.

[2] Zonneveld, B.J., Leitch, I.J. and Bennett, M.D. (2005) First nuclear DNA amounts in more than 300 angiosperms. Ann. Bot. (Lond.) 96, 229–244.

[3] Facchini, P.J. (2001) Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. Annu. Rev. Plant Physiol. Plant Mol. Biol. 52, 29–66.

[4] van der Heijden, R., Jacobs, D., Snoeijer, W. and Verpoorte, R. (2004) The *Catharanthus* alkaloids: pharmacognosy and biotechnology. Curr. Med. Chem. 11, 607–628.

[5] St-Pierre, B., Vazquez-Flota, F.A. and De Luca, V. (1999) Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. Plant Cell 11, 887–900.

[6] Cacace, S., Schröder, G., Wehinger, E., Strack, D., Schmidt, J. and Schröder, J. (2003) A flavonol *O*-methyltransferase from *Catharanthus roseus* performing two sequential methylations. Phytochemistry 62, 127–137.

[7] Vazquez-Flota, F.A., Moreno-Valenzuela, O., Miranda-Ham, M.L., Coello-Coello, J. and Loyola-Vargas, V.M. (1994) Catharanthine and ajimalicine synthesis in *Catharanthus roseus* hairy root cultures. Medium optimization and elicitation. Plant Cell, Tissue Organ Culture 38, 273–279.

[8] Sambrook, J. and Russell, D.W. (2001) Molecular Cloning, Cold Spring Harbor Laboratory Press, New York.

[9] Turinsky, A., Gordon, P., Xu, E., Stromer, J. and Sensen, C.W. (2005) Genomic data representation through images: the MAGPIE/Bluejay System in: Handbook of Genome Research (Sensen, C.W., Ed.), pp. 383–413, Wiley-VCH, Weinheim.

[10] Murata, J. and De Luca, V. (2005) The localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-*O*-methyltransferase to leaf epidermal cells defines them as a major site of precursor biosynthesis in the vindoline pathway in *Catharanthus roseus*. Plant J. 44, 581–594.

[11] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res. 25, 3389–3402.

[12] Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S. (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. DNA Res. 7, 175–180.

[13] Brandle, J.E., Richman, A., Swanson, A.K. and Chapman, B.P. (2002) Leaf ESTs from *Stevia rebaudiana*: a resource for gene discovery in diterpene synthesis. Plant Mol. Biol. 50, 613–622.

[14] Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S. (2000) Generation of non-redundant expressed sequence tags from a legume, *Lotus japonicus*. DNA Res. 7, 127–130.

[15] Jouannic, S., Argout, X., Lechauve, F., Fizames, C., Borgel, A., Morcillo, F., Aberlenc-Bertossi, F., Duval, Y. and Tregear, J. (2005) Analysis of expressed sequence tags from oil palm (*Elaeis guineensis*). FEBS Lett. 579, 2709–2714.

[16] The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

[17] Facchini, P.J. and Park, S.U. (2003) Developmental and inducible accumulation of gene transcripts involved in alkaloid biosynthesis in opium poppy. Phytochemistry 64, 177–186.

[18] Gang, D.R., Beuerle, T., Ullmann, P., Werck-Reichhart, D. and Pichersky, E. (2002) Differential production of *meta* hydroxylated phenylpropanoids in sweet basil peltate glandular trichomes and leaves is controlled by the activities of specific acyltransferases and hydroxylases. Plant Physiol. 130, 1536–1544.

[19] Meyer, K., Cusumano, J.C., Somerville, C. and Chapple, C.C. (1996) Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases. Proc. Natl. Acad. Sci. USA 93, 6869–6874.

[20] Kaltenbach, M., Schröder, G., Schmelzer, E., Lutz, V. and Schröder, J. (1999) Flavonoid hydroxylase from *Catharanthus roseus*: cDNA, heterologous expression, enzyme properties and cell-type specific expression in plants. Plant J. 19, 183–193.

[21] Irmler, S., Schröder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U. and Schröder, J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: new activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. Plant J. 24, 797–804.

[22] Madyastha, K.M., Guarnaccia, R., Baxter, C. and Coscia, C.J. (1973) *S*-adenosyl-L-methionine: loganic acid methyltransferase. A carboxyl-alkylating enzyme from Vinca rosea. J. Biol. Chem. 248, 2497–2501.

[23] De Luca, V., Balsevich, J., Tyler, R.T., Eilert, U., Panchuk, B.D. and Kurz, W.G.W. (1986) Biosynthesis of indole alkaloids: developmental regulation of the biosynthetic pathway from tabersonine to vindoline in *Catharanthus roseus*. J. Plant Physiol. 125, 147–156.

[24] Scalliet, G., Journot, N., Jullien, F., Baudino, S., Magnard, J.L., Channeliere, S., Vergne, P., Dumas, C., Bendahmane, M., Cock, J.M. and Hugueney, P. (2002) Biosynthesis of the major scent components 3,5-dimethoxytoluene and 1,3,5-trimethoxybenzene by novel rose *O*-methyltransferases. FEBS Lett. 523, 113–118.

[25] Lavid, N., Wang, J., Shalit, M., Guterman, I., Bar, E., Beuerle, T., Menda, N., Shafir, S., Zamir, D., Adam, Z., Vainstein, A., Weiss, D., Pichersky, E. and Lewinsohn, E. (2002) *O*-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. Plant Physiol. 129, 1899–1907.

[26] Burga, L., Wellmann, F., Lukacin, R., Schwab, W., Schröder, J. and Matern, U. (2005) Unusual pseudosubstrate specificity of a novel 3,5-dimethoxyphenol *O*-methyltransferase cloned from *Ruta graveolens* L. Arch. Biochem. Biophys. 440, 54–64.

[27] De Luca, V. and Laflamme, P. (2001) The expanding universe of alkaloid biosynthesis. Curr. Opin. Plant Biol. 4, 225–233.

[28] Shitan, N., Bazin, I., Dan, K., Obata, K., Kigawa, K., Ueda, K., Sato, F., Forestier, C. and Yazaki, K. (2003) Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. Proc. Natl. Acad. Sci. USA 100, 751–756.

[29] Noh, B., Bandyopadhyay, A., Peer, W.A., Spalding, E.P. and Murphy, A.S. (2003) Enhanced gravi- and phototropism in plant mdr mutants mislocalizing the auxin efflux protein PIN1. Nature 423, 999–1002.

[30] van der Fits, L. and Memelink, J. (2000) ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science 289, 295–297.

[31] Chatel, G., Montiel, G., Pre, M., Memelink, J., Thiersault, M., Saint-Pierre, B., Doireau, P. and Gantet, P. (2003) CrMYC1, a *Catharanthus roseus* elicitor- and jasmonate-responsive bHLH transcription factor that binds the G-box element of the strictosidine synthase gene promoter. J. Exp. Bot. 54, 2587–2588.

[32] Pauw, B., Hilliou, F.A., Martin, V.S., Chatel, G., de Wolf, C.J., Champion, A., Pre, M., van Duijn, B., Kijne, J.W., van der Fits, L. and Memelink (2004) Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in *Catharanthus roseus*. J. Biol. Chem. 279, 52940–52948.

[33] Burlat, V., Oudin, A., Courtois, M., Rideau, M. and St-Pierre, B. (2004) Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. Plant J. 38, 131–141.

[34] Mahroug, S., Courdavault, V., Thiersault, M., St-Pierre, B. and Burlat, V. (2005) Epidermis is a pivotal site of at least four secondary metabolic pathways in *Catharanthus roseus* aerial organs. Planta 223, 1191–1200.

[35] Scalliet, G., Lionnet, C., Le Bechec, M., Dutron, L., Magnard, J.L., Baudino, S., Bergougnoux, V., Jullien, F., Chambrier, P., Vergne, P., Dumas, C., Cock, J.M. and Hugueney, P. (2006) Role of petal-specific orcinol *O*-methyltransferases in the evolution of rose scent. Plant Physiol. 140, 18–29.

[36] Rischer, H., Oresic, M., Seppanen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., Montagu, M.C., Inze, D., Oksman-Caldentey, K.M. and Goossens, A. (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc. Natl. Acad. Sci. USA 103, 5614–5619.

[37] Asano, T., Masumura, T., Kusano, H., Kikuchi, S., Kurita, A., Shimada, H. and Kadowaki, K. (2002) Construction of a specialized cDNA library from plant cells isolated by laser capture microdissection: toward comprehensive analysis of the genes expressed in the rice phloem. Plant J. 32, 401–408.

[38] Nelson, T.S., Tausta, L., Gandotra, N. and Liu, T. (2006) Laser microdissection of plant tissue: What You See Is What You Get. Annu. Rev. Plant Biol. 57, 181–201.

[39] Schroeder, G., Wehinger, E., Lukacin, R., Wellmann, F., Seefelder, W., Schwab, W. and Schroeder, J. (2004) Flavonoid methylation: a novel 4′-*O*-methyltransferase from *Catharanthus roseus*, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases. Phytochemistry 65, 1085–1094.