**Title: Transcriptomic analysis for differential expression of genes involved in secondary metabolite production in *Narcissus pseudonarcissus* field derived bulb and *in vitro* callus**

**a Aleya Ferdausi, b Xianmin Chang, c Meriel Jones**

**Given names: Aleya; Xianmin; Meriel**

**Family names: Ferdausi; Chang; Jones**

a PhD, Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom.

Present address: Associate Professor, Department of Genetics and Plant Breeding, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh.

Email: aferdausi.gpb@bau.edu.bd

Tel: +8801747173836

bAssociate Pro-Vice Chancellor and Senior Lecturer, Crop Production Technology, Royal Agricultural University, Cirencester, Gloucestershire, GL7 6JS, United Kingdom

Email: Xianmin.Chang@rau.ac.uk

Tel: 01285 652531

cHonorary Senior Lecturer, Functional and Comparative Genomics, Institute of Integrative Biology, Biosciences Building, Crown Street, University of Liverpool, Liverpool, L69 7ZB

Email: m.g.jones@liverpool.ac.uk

**Corresponding author: Aleya Ferdausi a**

Declarations of interest: none

**ABSTRACT**

The Amaryllidaceae genus *Narcissus* contains diversified secondary metabolites, which are important sources of biologically active alkaloids. The biochemical pathways for secondary metabolite production require primary metabolites to undergo a series of modifications catalyzed by several enzymes but the knowledge on their biosynthesis is very limited. This study is undertaken to reveal the differential expression of transcripts related to secondary metabolites biosynthesis in *Narcissus* basal plate and *in vitro* callus.

RNA-seq analysis of *Narcissus* basal plate (field) and *in vitro* callus ensued total 2153 differentially expressed transcripts and 83.46% of those were annotated. The largest gene group was designated as uncharacterized proteins (10.24%) and the second largest group was genes responsible for secondary metabolite production (8.88%). The genes involved in the biosynthesis of alkaloids; cytochrome *P450s*, O-methyltransferases (*OMTs)*, *NADP/NADPH* dehydrogenases or reductases, *SAM-*synthetases or decarboxylases, 3-ketoacyl-CoA, acyl-CoA, cinnamoyl-CoA, cinnamate 4-hydroxylase, alcohol dehydrogenase, caffeic acid, *N*-methyltransferase, and *NADPH*-cytochrome *P450s* were present in both basal plate and callus. However, cytochrome P450s, and *OMTs* which are responsible for regulating the later stage of alkaloids biosynthesis were mainly up-regulated in field samples. Whereas, the enzymes involved in initial biosynthetic pathways, fructose biphosphate adolase (FBA), aminotransferases, dehydrogenases, hydroxyl methyl glutarate (HMT), and glutamate synthase (GS) leading to the biosynthesis of precursors of secondary metabolites; tyrosine, phenylalanine, and tryptophan were up-regulated in callus. Furthermore, Gene Ontology (GO) annotation of the transcripts and associated Kyoto Encyclopedia of Genes and Genomes (KEGG) and Plant Reactome pathway maps were postulated. This knowledge on probable genes or enzymes involved in secondary metabolism, their networking pathway and molecular regulation in field and *in vitro* callus would provide a deep insight into the *Narcissus* plant biology related to valuable metabolite production.

**Keywords:** *Narcissus*, RNA-Seq, secondary metabolism, alkaloids, gene expression

1. **Introduction**

A massive diversity of natural products or metabolites is produced by plants which possess wide-ranging biological and metabolic activities (Hotchandani et al., 2019). The chemical compounds extracted from plants are used for the semi-synthetic synthesis of pharmaceuticals where the starting materials, extracted from plants, are modified chemically to obtain better stability, specificity or action (Oksman-Caldentey and Inze, 2004). The Amaryllidaceae genus *Narcissus* owns a vast range of alkaloids which are known as Amaryllidaceae alkaloids such as galanthamine, lycorine, narciclasine, haemanthamine. Among these only galanthamine is produced on a commercial scale for the pharmaceutical industry to treat early to mid-stage Alzheimer’s diseases (Takos and Rook, 2013). Lycorine and lycoramine from *Lycoris radiata* (red spider lily) are used in Chinese traditional medicine. Since the ancient past, Amaryllidaceae species have been used in traditional herbal medicine and the initial evidence of their therapeutic application was recorded in the fourth century by the use of oil extracted from *N. poeticus* L. for the treatment of uterine tumors (He et al., 2015). Amaryllidaceae alkaloids possess a number of medicinal properties including inhibition of acetylcholinesterase and ascorbic acid biosynthesis, cytotoxicity, anti-cancer, anti-tumor and anti-viral activities (Berkov et al., 2009; Tahchy et al., 2010; Osorio et al., 2010; Berkov et al., 2014).

It is mostly expensive and challenging to extract valuable metabolites from naturally grown plant due to the major insect pest infestation as well as environmental factors (Oksman-Caldentey and Inze, 2004). Therefore, the biotechnological production of secondary metabolites in plant *in vitro* cultures has become an attractive alternative to the extraction from whole plant than the conventional method (Oksman-Caldentey and Inze, 2004; Ferdausi et al., 2021a). In addition, tissue culture materials are uniform and free from all natural contaminants (Nalawade and Tsay, 2004); which may allow the easy extraction and purification of metabolites, and a cheaper method for the production of stereo-chemically complex compounds than chemical synthesis-based production (Oksman-Caldentey and Inze, 2004; Mulabagal and Tsay, 2004).

Due to the complex chemical structures, multifaceted metabolic networking, and limited knowledge on the biosynthesis of plant secondary metabolites, their nature of production and accumulation in plants is undistinguishable (Verpoorte et al., 2002). However, new technologies such as biotechnology, metabolomics and transcriptomics are starting to give new insights on plant secondary metabolism (Lubbe et al., 2013; Berkov et al., 2014). Secondary metabolites production often requires the primary metabolites to undergo a series of modifications catalyzed by enzymes such as cytochrome P450s, methyltransferases, glycosyltransferases and acyltransferases. The putative genes involved in secondary metabolite production exist in families with multiple members because of recurrent gene duplication (Ober, 2005). Some of those putative genes are belong to small families (<10 members) involved in the production of a specific secondary product such as cinnamate 4-hydroxylase (C4H); while others present in mid-size families (10-100 members) such as cinnamoyl-CoA reductase (C3H) and terpene synthase. Moreover, some genes are belong to very large families (>100 members) involved in the production of a plethora of plant secondary metabolites which include the cytochrome P450s, UDP-dependent glycosyltransferases, and methyltransferases (Zhao et al., 2014). Putative genes including phenylalanine ammonia lyase (*PAL*), *N*-methyltransferase (*NMT*), cytochrome P450s (*CYPs*), tyrosine decarboxylase (*TYDC*), *O*-methyltransferase (*OMT*), and other potentially important candidate genes involved in Amaryllidaceae alkaloids synthesis have been identified in *Lycoris* (Wang et al., 2013) and *Narcissus* (Kilgore et al., 2014; Kilgore et al., 2016a; Hotchandani et al., 2019; Ferdausi et al., 2021b). One candidate gene for the methylation of norbelladine to 4'-*O*-methylnorbelladine (norbelladine 4'-*O*-methyltransferase, *Np*N4OMT) has been identified in *N. pseudonarcissus* and genes that co-express with it could be identified and used as candidates for the other steps in the proposed Amaryllidaceae alkaloids biosynthetic pathway (Kilgore et al., 2014). Furthermore, a cytochrome P450, *CYP96T1* has been identified through comparative transcriptomics of *N. pseudonarcissus* and *Galanthus spp*. (Kilgore et al., 2016a) which is capable of forming the products (10b*R*,4a*S*)-noroxomaritidine, and (10b*S*,4a*R*)-noroxomaritidine from 4′-*O* methylnorbelladine (Kilgore et al., 2016b). Putative genes such as *PAL*, *TYDC*, cytochrome P450, and *Np*N4OMT were also reported to be expressed in diverse pattern in *in vitro* tissues of *N. pseudonarcissus* (Ferdausi et al., 2021b).

Next generation sequencing offers two approaches to transcriptome analysis in non-model plants, the *de novo* which involves assembly then alignment techniques, in the case of limited or absence of genomic or transcriptomic resources (Ward et al., 2012). And the mapping which is based on aligning then assembling techniques that are completely dependent upon the existence of a reference genome or transcriptome (Haas and Zody, 2010). RNA-seq makes the transcriptomic experiments applicable for non-model plants where genomic resources are limited or completely unavailable (Xiao et al., 2013). It offers a comprehensive understanding of transcriptome complexity (Kumar et al., 2012) and allows the investigation of a dynamic range of expression levels in non-model plant like *Narcissus* (Wang et al., 2009). It can detect the expressed sequences in specific tissues at a specific time, as it does not require a reference genome to gain specific and meaningful transcriptomic information (Strickler et al., 2012; Egan et al., 2012). Otherwise, the reads obtained can be directly aligned with the reference, if genome or transcriptome information is available (Weber et al., 2007; Mortazavi et al., 2008) which also has been extensively used for non-model plants in the absence of genome or transcriptome information by transcript reconstruction using *de novo* assembly (Wang et al., 2013, Kilgore et al., 2014). It provides the capability to discover new genes and transcripts, and allows measurement of transcript expression in a single assay (Trapnell et al., 2012; Nagalakshmi et al., 2008). Implementation of RNA-seq can sometimes be challenging in plants because of their complex structures, complicated genetic make-up, and lack of genomic resources (Strickler et al., 2012). The biosynthetic genes involved in plant secondary metabolism are mainly structured in complex enzymatic networks yielding several compounds rather than simple linear schemes leading to a single compound. Thus, the discovery of the biosynthetic genes involved in the production of secondary compounds becomes more challenging ([Hall et al., 2013](http://www.sciencedirect.com/science/article/pii/S0168165613001843#bib0105)). Furthermore, limited genomic resources are available for the most valuable specialized metabolites, and generally in non-model plant species such as *Narcissus* (Strickler et al., 2012).

In our previous finding, the *N. pseudonarcissus* cv. Carlton bulb (basal plate) tissue exhibited the high galanthamine content while *in vitro* callus was the lowest galanthamine producer (Ferdausi at al., 2020). Furthermore, it was observed that the putative genes for the production of alkaloids were highly expressed in field samples than the *in vitro* samples (Ferdausi et al., 2021b). Therefore, a comprehensive transcriptomic analysis of tissues derived from natural (basal plate) and *in vitro* (callus) conditions would provide a better understanding on the differential expressions of genes involved in the biosynthesis of secondary metabolites in *Narcissus*.

1. **Materials and Methods**

*2. 1 Plant tissues and chemicals*

Basal plate tissues (Ferdausi et al., 2020) of *Narcissus pseudonarcissus* cv. Carlton bulbs supplied by New Generation Daffodil Ltd, UK and callus (Ferdausi et al., 2020) from the same bulbs (twin-scale explants) grown on MS (Murashige and Skoog, 1962) basal medium supplemented with 100 mg/l yeast extract, 50 mg/l ascorbic acid, 30 mg/l polyvinylpyrrolidone, 0.5 mg/l kinetin (KN), 1.5 mg/l benzyl amino purine (BAP) and 20 mg/l naphthalene acetic acid (NAA) were used for RNA extraction. Basal plates and calli were cut into small pieces (1 to 2 mm thick), weighed (~150 mg) and stored in -80oC until required. The frozen tissues were ground in a pestle and mortar under liquid nitrogen to achieve fine powder of tissues. After grinding, ~100 mg of sample was transferred into 1.5 ml microfuge tubes and kept in liquid nitrogen to avoid thawing. MS media, ascorbic acid, KN, BAP and NAA were purchased from Duchefa Biochemie (Netherlands), polyvinylpyrrolidone from Sigma-Aldrich (UK) and yeast extract from Thermo Scientific (Germany).

*2.2 Library preparation and Illumina sequencing*

Total RNA was extracted using the innuPREP Plant RNA Kit (Analytic Jena, Germany) according to manufacturer’s protocol with slight modification before wash-up stage by adding DNAse treatment (RNase-Free DNase Set, Qiagen) which was a column DNAse treatment, performed following RNeasy Plant mini kit (Qiagen) DNA digestion steps. NanoDrop Spectrometer (Thermo, USA) and Qubit Quant-iTTM RNA Assay Kit (Thermo Fisher Scientific, USA) were used to determine the initial quality and quantity of extracted RNA samples. In total six samples, three basal plate tissues from three different Carlton bulbs and three callus samples induced from the same bulbs were sent to the Centre of Genomic Research (CGR), University of Liverpool, for RNA sequencing. Total RNA quality and RNA integrity (RIN) were tested again by CGR using Agilent 2100 Bioanalyser (Agilent Technologies), which fulfilled the quality requirement for further sample processing. rRNA depletion was performed using the Ribo-Zero rRNA removal kit (Plant seed/ root) (Illumina, USA) following manufacturer's protocol. The ScriptSeq v2 (Epicentre, USA) RNA-seq library preparation kit was used for library preparation, which produces directional sequencing reads using a random-primed cDNA synthesis reaction. The HiSeq 2500 System (Illumina, UK) was used for RNA sequencing.

*2.3 Data processing*

The raw FASTQ files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 ([Martin, 2011](#_ENREF_3)). The option -O3 was used, so the 3' end of any reads, which matched the adapter, sequence for 3 base pairs or more, was trimmed. The reads were further trimmed using Sickle version 1.200 (GitHub) with a minimum window quality score of 20. After trimming, reads shorter that 10 base pairs were removed. Statistics were generated using FASTQ-stats from EAUtils. After that the RNA-seq data in FASTQ format was released by CGR and was available to download for further downstream analyses.

*2.4 RNA-seq analyses*

There was no reference genome available for *Narcissus* in databases. Kilgore et al., (2014) reported a *N. pseudonarcissus* transcriptome assembly and the raw reads are available at the MedPlant RNA Seq Database (<https://medplantrnaseq.org/>) which was downloaded and used as a reference transcriptome for further RNA-seq analyses.

The paired-end RNA-seq FASTQ files containing forward and reverse read files of three replicates for basal plate and callus were used for characterizing differential expression. The analyses were performed on Discovery Environment of CyVerse, following the tutorial: DE 002: Characterizing Differential Expression with RNA-Seq (Tuxedo Method) (www.cyverse.org). The method is based on the Tuxedo Suite; TopHat software was used for read alignment while Cufflinks was used to assemble the aligned reads for the reconstruction of transcriptome. Cuffmerge was applied to merge multiple conditions and Cuffdiff to calculate the differential gene expressions. Finally, CummeRbund was used for the visualization of data (Trapnell et al., 2012).

# *2.5 Differentially expressed genes/ transcripts annotation*

# The differentially expressed genes/ transcripts (Cuffdiff output) were annotated following a BLAST search. A BLAST database containing proteins in the non-redundant protein database related to plants was used to BLAST the transcriptome. The BLAST result of annotated genes/ transcripts provides an indication of the identity of transcripts that might be involved in differential expression between the two conditions under study. The data obtained was completely un-curated and based on top hits from the BLAST output.

*2.6 Functional annotations and pathway analysis*

The differentially expressed transcripts, significant at q-value (corrected p-values) ≤ 0.05, in basal plate and callus were then annotated to their Gene Ontology categories using UniProt Gene Ontology (www.uniprot.org) web tools. Further, GO terms for UniProt Gene Ontology annotated transcripts were obtained using EMBL-EBI Quick GO-Beta Gene Ontology and GO Annotations ([www.ebi.ac.uk/QuickGO-Beta](http://www.ebi.ac.uk/QuickGO-Beta)). The transcripts detected in basal plate and callus were mapped to the KEGG pathway mapper (www.genome.jp/kegg/mapper.html) and also mapped with the curated Plant Reactome (plantreactome.gramene.org) for *Oryza sativa* to obtain a postulated pathway for *Narcissus* basal plate and callus transcripts.

Figure 1 represented an overview of the methods used for Illumina sequencing and of the subsequent data analysis.

1. **Results**

*3.1 Raw reads trimming and filtering*

Among the six sequenced samples one basal plate sample (CBS2) showed poor statistics in post-run, with mapping to rRNA at a level of about 74% whereas the other samples showed mapping at about 1 to 12%. The variation of 1 to 12% is not unusual and acceptable, but the mapping variation in CBS2 was questionable and unpredicted even though it was collected and processed using the same methods used for other two replicates of basal plate (CBS1 and CBS2). The average base quality score for all samples and replicates were >30, which indicated base call accuracy of 99.9%. More than 95% of total raw reads were recovered after trimming and filtering i.e. 97.7% in callus (CAL1), 97.5% in CAL2, 96.05% in CAL3, 96.6% in CBS1 and 97.2% in CBS3, except for CBS2 (92.2%). The results of the adapter trimming and quality filtering for Carlton basal plate and callus samples are available in Supplementary 1, Table S1 (Carlton basal plate) and Table S2 (callus).

*3.2 Differential gene expression*

A *Narcissus* transcriptome assembly was obtained from MedPlant RNA Seq Databases (<https://medplantrnaseq.org/>) and used for *Narcissus pseudonarcissus* cv. Carlton basal plate and *in vitro* callus RNA-seq read alignment. The TopHat read mapping showed a moderate mapping coverage ranging from about 35% to 48% of total input in all samples under study except one replicate of basal plate, CBS2 which was about 19% (Table 1).

The TopHat analysis parameters were set as default to an anchor length of 8, maximum number of mismatches that can appear in the anchor region of spliced alignment to 0 (zero), minimum intron length, 70 bp and maximum intron length, 50000 bp. Maximum number of alignments to be allowed (alignment score) was of 20, number of mismatches allowed in each segment for reads mapped independently was 2, minimum length of read segment was 20 bp and meta-pair inner distance was 50bp. These parameters were set for reporting the best possible alignment which may greatly increase the mapping accuracy at the expense of an increase in running time and rate of alignment.

The transcript assembly output obtained from Cufflinks detected a total of 2153 differentially expressed and regulated transcripts in two conditions under study (field derived basal plate and *in vitro* grown callus). About 83.46 % transcripts were annotated among the 2153 differentially expressed genes using a BLAST search against UniProt, SwissProt, RefSeq and TAIR reported in Supplementary 2 (Appendix 1). About 64.21% genes were detected as up-regulated in callus, whereas 35.79% were up-regulated in the basal plate. However, many of these transcripts were not significantly different between the two conditions (Figure 2). The distribution of all significant differentially expressed genes (-log10 (p-value) ≥ 2.0) in Figure 2 showed that the genes on positive log2 axis were up-regulated in callus, hence were down regulated in basal plate, while genes on negative log2 axis were up-regulated in basal plate and down regulated in callus. A total of 206 genes (9.57%) were detected as significantly up or down regulated in basal plate and callus with q-values (corrected p-values) of ≤ 0.05. The mapping details of all 206 genes (annotated) with their corresponding log2 (fold change), total FPKM, regulation pattern, and q-values are reported in Supplementary 2 (Appendix 2). Among the 206 significant differentially expressed genes, 95 genes were found to be up-regulated in callus and down regulated in basal plate and 111 genes were up-regulated in basal plate, hence down regulated in callus.

The eleven most abundant gene groups that contributed to 41.7% of the total annotated genes were detected in *Narcissus* transcriptome data (Figure 3). The largest gene group was uncharacterized proteins (10.24%) followed by the genes responsible for secondary metabolite production (8.88%) which included cytochrome P450s, methyltransferases, CoA-reductases, NADP/ NADH dependent reductases, cinnamate-4-hydroxylases (C4H), hydrolases, aldo-keto reductases, oxidoreductases, and others (Supplementary 2; Appendix 3). The third largest group was denoted for tissue differentiation (4.06%) and stress-related proteins (4.06%) such as pathogenesis related, universal stress, zinc finger, and chitin or chitinase like proteins (Supplementary 2; Appendix 4). Other ample groups detected were tissue differentiation related proteins such as ERF, AP2-domain containing, heat-shock, homeobox-containing and glutathione *S*-transferases, other transcription factors, ATP/GTP binding, ubiquitin ligases/hydrolases and sugar, and amino acid synthase related proteins (Figure 3; Supplementary 2; Appendix 4).

The Cuffdiff output showed the differential expressions (Figure 4) of top fifty transcripts with q-values of ≤ 0.01 based on total FPKM-values and log2 (fold change). The transcripts in black box are present in both basal plate and callus tissues in different levels, whereas, the red box represents the transcripts expressed in basal plate but not in callus and the transcripts expressed in callus are marked in blue box (Figure 4).

*3.3 Genes or transcripts related to secondary metabolism*

The probable genes/ transcripts involved in the secondary metabolite production detected in the *Narcissus* transcriptome data among the 206 significantly differential expressed genes based on their BLAST annotations were grouped into the five major categories such as phenylpropanoid pathway related enzymes (Cytochrome P450s), methyltransferases, *NADP/NADPH* dependent reductases, oxidoreductases, and hydrolases (Table 2). Moreover, other notable genes related to secondary metabolism in plants i.e. phenolic compounds, lectins, polypeptides etc. were also observed in *Narcissus* basal plate and callus (Table 2). The significant genes with q-value of ≤ 0.05 are shown here but other transcripts (q-value greater than 0.05) related to Amaryllidaceae alkaloid biosynthesis and other secondary metabolite productions were also detected in *Narcissus* transcriptome data. The data for those transcripts i.e. Cytochrome P450s, methyltransferases, *S*-adenosyl-*L*-methionine decarboxylases, *NADP/NADPH* reductases, CoA reductase/ ligases, hydrolases and others, with their corresponding q-values has been listed in Supplementary 2; Appendix 3.

*3.4 Other notable gene/ transcript groups of interest*

Beside secondary metabolism related genes, important genes related to tissue growth and development such as tissue differentiation (callus and shoot formation), stress or defense related protein, and ATP synthase were detected in *Narcissus* transcriptome data (Figure 5, Supplementary 2; Appendix 4). The stress related transcripts were highly present in callus tissues along with ATP binding, glutathione S-transferases, ubiquitin, zinc figure, and splicing factors related transcripts than basal plate. Those for heat-shock proteins, which are also responsible for stress factors and callus formation, were detected in greater numbers in callus. Conversely, maximum transcripts of genes responsible for tissue differentiation such as ethylene response factors, AP2-domain containing, and homeobox DNA binding proteins were detected in basal plate (Figure 5).

*3.5 Functional annotation of Narcissus transcripts*

The Gene Ontology (GO) annotations of 100 out of 111 transcripts were obtained via UniProt which were detected as significantly (q-value ≤ 0.05) up-regulated in basal plate. Few of the transcripts were assigned to more than one GO category. It was observed that 55 transcripts were assigned to molecular function with the most enriched terms being binding and catalytic activity, 35 transcripts were assigned to cellular component with cell part, organelle, and membrane/membrane part being the most enriched terms. Moreover, 49 transcripts were assigned to the biological process category showing metabolic process, cellular process, and single-organism process as the most enriched terms (Figure 6a).

The functional annotations of 90 among the 95 significantly (q-value ≤ 0.05) up-regulated transcripts were acquired in callus. The UniProt GO analysis in callus revealed 62 transcripts assigned to the molecular function with catalytic activity and binding as the most enriched terms. Besides, 40 transcripts were assigned to cellular component indicating cell, cell part, and membrane as the top three categories and biological process category showed the assignment of 55 transcripts with the most enriched terms being the cellular process, metabolic process, and single-organism process (Figure 6b).

Furthermore, the EMBL-EBI Quick GO-Beta annotation (www.ebi.ac.uk/QuickGO-Beta/annotations) was used to assign the detected transcripts to complement GO functionality predictions with GO terms. A total of 365 GO terms for 111 basal plate and 618 GO terms for 95 callus transcripts were acquired which was substantially more than that obtained from UniProt. As shown in Figure 7, in case of basal plate, 162 (44.38%) GO terms were assigned for molecular function, 112 (30.68%) for biological process, and 91 (24.93%) for cellular component. Similarly, 275 (44.5%), 196 (31.72%), and 147 (23.79%) GO terms were assigned for molecular function, biological process, and cellular component respectively in callus. The proportions assigned to each category were therefore essentially the same for both tissues but with several terms to each transcript. In addition, secondary metabolism related GO terms were possible to detect using Quick GO-Beta annotations, which were not found in UniProt GO annotation such as GO terms for *O*-methyltransferase activity (GO:0008171), methyltransferase activity (GO:0008168), and S-adenosylmethionine-dependent methyltransferase activity (GO:0008757) were found in basal plate that were not present in callus. The top three GO terms for basal plate, assigned to molecular functions were metal ion binding (GO:0046872), oxidoreductase activity (GO:0016491), and DNA binding (GO:0003677). On the other hand, ATP binding (GO:0005524), catalytic activity (GO:0003824), nucleotide binding (GO:0000166), oxidoreductase activity (GO:0016491), transferases activity (GO:0016740) were the top assigned molecular functions related GO terms detected in callus (Figure 7; Supplementary 2; Appendix 5). A full list of all annotated GO terms (Quick GO-Beta) detected in basal plate and callus with their corresponding gene descriptions are in Supplementary 2; Appendix 5.

*3.6 Pathway analyses of Narcissus transcripts*

The UniProt IDs of basal plate and callus transcripts were separately mapped to their corresponding KEGG Orthology (KO) using the UniProtKB Retrieve/ID mapping tool. A total of 28 KO for basal plate and 78 for callus were retrieved by mapping against the corresponding UniProt IDs. The KO IDs were then mapped to KEGG Pathway Mapper that yielded a total of 55 mapped pathways for basal plate and 78 KEGG pathways for callus. It was evident that the mapped pathways for basal plate with highest number of hits (enzyme classes) were mainly corresponded with the pathways related to secondary metabolite production (Supplementary 1; Table S3). Likewise, the KEGG pathways mapped in callus were also found to be related to secondary metabolism. Moreover, carbon metabolism and glycolysis related pathways were detected in callus that was not found in basal plate (Supplementary1; Table S4). Pathways related to aromatic amino acid metabolism were present in both basal plate and callus (Supplementary 1; Table S3 and Table S4).

The Plant Reactome database (plantreactome.gramene.org) for plant metabolomic and regulatory pathways is a freely accessible, curated and peer reviewed pathway database which provided an alternative way to map metabolic pathways of *Narcissus* transcripts. The KO IDs obtained from UniProtKB Retrieve/ID mapping tool were mapped back to UniProtKB IDs for both basal plate and callus that yielded 45 and 120 UniProt genes corresponding to *Oryza sativa* for basal plate and callus respectively. These UniProt genes were mapped to the Plant Reactome *Oryza sativa* pathway database which revealed 14 and 26 pathways for basal plate and callus respectively. The pathways related to secondary metabolism (according to database) and central metabolism are represented in Table 3 and the details of all detected pathways are reported in Supplementary 1; Table S5.

The postulated pathway of secondary metabolite biosynthesis in *Narcissus*, based on KEGG and Plant Reactome showed differential enzyme expressions (basal plate, callus and both) yielding different secondary metabolites including alkaloids (isoquinoline, indole, quinoline, terpenoids, and Amaryllidaceae alkaloids), plant hormone synthesis, lignin biosynthesis as well as pathways related to amino acid (phenylalanine, tyrosine and tryptophan) metabolism (Figure 8).

1. **Discussion**

Secondary metabolites produced by plants are invaluable sources of medicines, food additives, and industrial products, being widely used for human benefits. However, a little is known about the genes or enzymes involved in the biosynthesis or accumulation of these metabolites in whole plant as well as in plant cell cultures (Wang et al., 2019; Karuppusamy, 2009). *Narcissus* is a well-reported genus producing a wide range of metabolites along with pharmaceutically important alkaloids such as galanthamine, lycorine, lycoramine, narciclasine, haemanthamine (Ferdausi et al., 2020; Lubbe et al., 2013, Takos and Rook, 2013). The bulbs, especially the basal plate of *Narcissus* showed the production of the highest amount of galanthamine, one of the pharmaceutically important alkaloids along with other related metabolites. While, the *in vitro* induced callus indicated the lowest galanthamine content using gas chromatography-mass spectrophotometry (Ferdausi et al., 2020). However, relative expression analysis using RT-PCR showed the diverse expression pattern of putative transcripts involved in the secondary metabolites production such as *PAL*, *TYDC*, *OMT*, and P450s in field as well as in *in vitro* tissues of *Narcissus* (Ferdausi et al., 2021b). Hence, further transcriptome and pathway analyses are required to have a deep insight on the genes involved and their metabolic networking and regulation in valuable secondary metabolites production in *Narcissus*. Transcriptome analysis (RNA-seq) is an effective platform for the identification and functional characterization of novel candidate genes as well as to identify genes encoding uncharacterized enzymes (Desgagné-Penix et al., 2010). Hence, a RNA-seq platform, which has been established as a valuable technology for transcriptome studies in non-model plants such as *Narcissus* for which genomic information in limited (Hotchandani et al., 2019; Park et al., 2019), was taken into account for transcriptome comparison of *Narcissus pseudonarcissus* cv. Carlton basal plate and callus in terms of secondary metabolites production.

Six (6) sequence libraries of basal plate (3) and callus (3) were generated but one replicate of basal plate (CBS2) showed very high mapping (74%) to rRNA due to inadequate rRNA depletion and poor statistics in post-run; although it was processed and treated as all others samples and showed good results in initial quality check on Bioanalyser (RIN > 7; clear 18S and 28S peaks and 28S/ 18S ratios 1.5-2.0) (Ward et al., 2012; Johnson et al., 2012). This result indicates that the measurements of RNA quality unfortunately cannot always predict successful sequencing as well as could affect the measurement of gene expression (Romero et al., 2014). Repeating the sequencing of one sample was not possible due to the expense and limitation of CGR sequencing protocol set-up.

In case of non-model plants, *de novo* assembly is often used (Kilgore et al., 2014, Kilgore et al., 2016a, Hotchandani et al., 2019), nevertheless, mapping strategy has also been reported as more applicable for differential expression analysis in non-model plants (Vijay et al., 2013). The *Narcissus* basal plate and callus transcriptome sequences of this study were mapped to *Narcissus spp*. transcriptome data published on MedPlant RNA Seq Database (<https://medplantrnaseq.org/>) (Kilgore et al., 2014). Differential gene expression analysis for *Narcissus* basal plate and callus transcriptome data was performed on Discovery Environment (Oliver et al., 2013) and Atmosphere of a well-documented cyber-infrastructure CyVerse which was previously known as iPlant Collaborative (Goff et al., 2011). Typically, read alignment rate more than 50% is required to obtain an optimum differential expression analysis but lower mapping rates may be obtained due to the lack of complete annotated genome, poor read quality or the presence of contaminants (Trapnell et al., 2012). The sequence alignment of our data to MedPlant *Narcissus* transcriptome using TopHat showed overall mapping coverage ranging from around 35% to 48% of total input except for one replicates of basal plate (19%) which indicated a moderate mapping coverage. The mapping rate obtained from our results might be due to the lack of complete annotated transcriptome or poor read quality of one replicates of basal plate. Moreover, the reference transcriptome of *Narcissus spp*. used could be from different variety as it was not confirmed as Carlton and reported to be likely Carlton (Kilgore et al., 2014) and thus the presence of SNPs and other polymorphism may affect the mapping. Furthermore, the reference transcriptome assembly constructed based on daffodil leaf, inflorescence and bulb samples (Kilgore et al., 2014). Whereas the *Narcissus* transcriptome data in this study obtained from Carlton basal plate and *in vitro* tissue (Callus), which might also affect the overall mapping for sample divergence. However, read alignment rate ranging from 17% to 94% using TopHat has been reported previously, which was acceptable for differential analysis (Trapnell et al., 2009).

The aligned reads after assembly (Cufflinks) resulted in a total of 2153 Cuff genes which allows the identification of genes in any system and also suggests experiments to establish their regulation pattern (Trapnell et al., 2010). The Cuffdiff outputs provided a number of files (Trapnell et al., 2012) to study 2153 differentially expressed and regulated genes; among which 64.21% transcripts were up-regulated in callus and 35.79% were up-regulated in basal plate, this could be due to the higher mapping rate of callus transcriptome than the basal plate. However, 206 genes or transcripts were detected as significant differentially expressed (q-values ≤ 0.05) in the two conditions under study where higher numbers of significant different genes were up-regulated in basal plate (111) than in callus (95).

Approximately, 79% of the differentially expressed genes were annotated which revealed eleven abundant gene groups of which uncharacterized proteins (10.24%), secondary metabolism related proteins (8.88%), stress related proteins (4.06%), tissue (callus/shoot) differentiation (4.06%), ATP/GTP binding (3.82%), and transcription factors (3.06%) were the mostly enriched groups. Transcriptome analysis of opium poppy cell cultures showed the presence of secondary metabolism related genes such as the most abundance transcripts encoding putative cytochrome P450 with the next abundant enzymes involved in the biosynthesis of *S*-adenosylmethionine such as SAM synthetase. Other abundant transcripts were defense/stress response proteins, peroxidase, pathogenesis-related proteins, and ubiquitin (Desgagné-Penix et al., 2010), which were also found as major gene groups in our samples (Figure 3; Figure 5).

The most abundant transcripts associated with plant growth, development, metabolism, defense, transport, and cellular structure were obtained from our dataset which were commonly reported in other plants (Desgagné-Penix et al., 2010; Délano-Frier et al., 2011; Desgagné-Penix et al., 2012). The most significantly (q-value ≤ 0.05) up-regulated genes detected in basal plate were secondary metabolism related enzymes involved in the later stage of secondary metabolite or alkaloid biosynthesis; such as cytochrome *P450s*, *OMTs*, *NADP/NADPH* dehydrogenases/ reductases and *S*-adenosylmethionine synthetases/ decarboxylases, and lectins (Berkov et al., 2014; Singh and Desgagné-Penix, 2017). Other important transcripts which were found to be up-regulated in basal plate with q-values higher than 0.05 and related to secondary metabolism were 3-ketoacyl-CoA, acyl-CoA, cinnamoyl-CoA, cinnamate 4-hydroxylase, alcohol dehydrogenase, and caffeic acid *O*-methyltransferase (Zhao et al., 2014; Sengupta et al., 2015). A candidate *O*-methyltransferase; (*Np*N4OMT) responsible for the methylation of norbelladine to intermediate precursor 4'-*O*-methylnorbelladine (Kilgore et al., 2014) and cytochrome *P450* (*CYP96T1*) catalyzing the para-para *C-C* phenol coupling (Kilgore et al., 2016a) in the Amaryllidaceae alkaloid biosynthesis have already been proposed in *N. pseudonarcissus*. The plant *O*-methylation reactions are mostly catalyzed by *SAM*-dependent methyltransferases (Liscombe et al., 2012). Other probable enzymes related to alkaloid biosynthesis in Amaryllidaceae were also detected in basal plate such as aldo-keto reductases, and alcohol dehydrogenase, reported for the reduction of ketones, aldehydes, *C-C* double bond, and imines (Sengupta et al., 2015).

Conversely, the significantly up-regulated genes (q-values ≤ 0.05) in callus were related to the enzymes involved in the primary metabolism, such as fructose biphosphate adolase, aminotransferases, dehydrogenases, hydroxyl methyl glutarate, and glutamate synthase. However, other secondary metabolism related enzymes were detected in callus with q-values > 0.05 including cinnamoyl/ coumarate/ enoyl CoA reductases, *N*-methyltransferases, *NADPH*-cytochrome P450s, *S*-adenosylmethionine synthetases, and alcohol dehydrogenases. The methyltransferases detected in callus were either probable *N*-methyltransferases or *(SAM)* dependent *N*-methyltransferases which have been reported in cell cultures of *P. bracteatum* and other benzylisoquinoline alkaloid producing species (Liscombe et al., 2009). The *in vitro* cultures of *C. roseus*, *Camptotheca acuminata* and *R. serpentina* also showed the identification of transcripts related to alkaloid biosynthesis, i.e. *N*-methyltransferases, *O*-methyltransferases, and *NADPH-*cytochrome P450 reductases (Góngora-Castillo et al., 2012). The alkaloids related to putative transcripts *OMT*, *NMT*, *SAM* were found to be expressed in higher level than the P450-dependant enzymes in opium poppy cell cultures (Desgagné-Penix et al., 2010).

Besides the secondary metabolism related genes, most of the transcription factors and *ERF/ AP2* domain containing proteins were detected in basal plate while the stress related and heat-shock proteins were mainly discovered in callus. Transcription factors of *ERF/ AP2* family have been found to trigger shoot regeneration and cell differentiation (Neelakandan and Wang, 2012). Another crucial enzyme for plant growth and shoot regeneration is glutathione-*S*-transferases (Gong et al., 2005), which was identified, in both basal plate and callus. A homeobox protein was recognized in the *Narcissus* basal plate and callus and members of this family reported in *Arabidopsis* and *Brassica* to promote callus induction and somatic embryogenesis (Neelakandan and Wang, 2012). The heat-shock proteins mostly identified in *Narcissus* callus has been reported to be expressed during callus formation in *Arabidopsis* (Ogawa et al., 2007). The enzymes cinnamoyl-CoA, cinnamate 4-hydroxylase, caffeic acid *O*-methyltransferase, and coumarate ligase/reductases detected in *Narcissus* basal plate and callus, play an important role in phenylpropanoid pathway through hydroxylation leading to diverse plant alkaloids(Singh and Desgagné-Penix, 2014) ([Singh and Desgagné-Penix, 2014](#_ENREF_1)).

The gene functions assigned by the Gene Ontology database are mainly based on model organisms (Wolf, 2013) hence; there are no GO annotations available for species closely related to *Narcissus*. However, the UniProt GO annotation tool was applied successfully in previous *Narcissus* project (Pulman, 2015) and was reported in other non-model plants (Guo et al., 2013; Lee et al., 2014). Unlike ours, in Lycoris, an Amaryllidaceae species the GO annotation showed three major GO categories i.e. molecular function, biological process, and cellular component being the most associated transcripts to binding and catalytic activity, metabolic, and cellular process and cell, and cell parts respectively (Wang et al., 2013; Park et al., 2019). The Quick GO-Beta annotation provides Gene Ontology (GO) annotations to proteins in the UniProtKb, which includes a large number of high-quality functional annotations across a broad taxonomic range (Huntley et al., 2015). Therefore, detection of GO terms related to secondary metabolism such as *O*-methyltransferase activity, oxidoreductase activity, *S*-adenosylmethionine dependent methyltransferase activity, NADH dehydrogenase activity were possible which were detected in basal plate but not in callus.

The gene function information obtained from Gene Ontology alone is not sufficient to provide knowledge about biosynthetic pathways. Therefore, it is important to map the genes of interest to candidate metabolic pathways (Wolf, 2013). KEGG is one of the largest pathway databases (Ogata et al., 1999) and also extensively used in non-model plant studies (Xiao et al., 2013). In both basal plate and callus the central metabolic pathways and biosynthesis of secondary metabolites were the most enriched pathways with the highest number of assigned enzymes. Some human disease (e.g. Parkinson's and Huntington's disease) related pathways were also detected in both tissues which is an example where an inappropriate pathway annotation is made since KEGG is a large database that is particularly rich in information on pathways from human and animals, as well as data on plants and microbes. Therefore, it was worthwhile to compare this result with a plant-specific database; the Plant Reactome which is a plant pathway database based on curated *Oryza sativa* reference pathways (Tello-Ruiz et al., 2015). The most notable pathways were phenylpropanoid biosynthesis, amino acid metabolism, flavonoid biosynthesis, and *S*-adenosyl-*L*-methionine cycle. Both KEGG and Plant Reactome pathway analyses showed the detection of secondary metabolism related pathways in both basal plate and callus. The postulated pathway (Figure 8) generated in this study based on KEGG and Plant Reactome results showed that the transcripts (i.e. cytochrome P450s and *OMTs*) involved in the Amaryllidaceae alkaloid biosynthesis (Wang et al., 2013; Kilgore et al., 2014; Kilgore et al., 2016a) were mainly up-regulated in field samples. Whereas, the enzymes involved in initial pathways (fructose biphosphate adolase, aminotransferases, dehydrogenases, hydroxyl methyl glutarate, and glutamate synthase) leading to the biosynthesis of precursors (tyrosine, phenylalanine, and tryptophan) for secondary metabolites were up-regulated in callus which could be in accordance with the higher accumulation of Amaryllidaceae alkaloids in field tissues than *in vitro* tissues in *Narcissus* (Berkov et al., 2009; Tahchy et al., 2011; Ferdausi et al., 2020).

Our previous study showed the relative expression of these putative genes in basal plate and callus. The relative expressions of P450s and NpO4OMT were high in basal plate while PAL and TYDC showed relatively higher expression pattern in callus (Ferdausi et al., 2021b). This finding could be correlated with the detection of transcripts related to P450s and OMTs in basal plate involved in later stage of Amaryllidaceae alkaloids biosynthesis. And the high expression of PAL and TYDC in callus could be correlated with the phenylalanine and tyrosine metabolism related transcripts, involved in the initial biosynthetic pathway (Figure 8). Moreover, carbon metabolism and glycolysis related pathways were detected in callus which could be related to the more active heterotrophic metabolism of the callus tissues. Our previous NMR-based metabolomic study reported the high accumulation of Amaryllidaceae alkaloid precursors such as tyrosine, tyramine, succinylacetone, p-cresol and 3-phenylpropionate in *Narcissus* basal plate. While the metabolites related to carbohydrate metabolism such as sucrose, glucose, galactose, lactose were detected in callus (Ferdausi et al., 2020). Therefore, the previous findings on metabolomics and relative expression of putative genes could be correlated with the findings of the present study.

The link between expression of pathway genes and biosynthesis of their final product is complex. For example, the transcriptome analysis of opium poppy cell culture showed the complete absence of expression of *TYDC*, salutaridine synthase (*SalSyn*), salutaridine reductase (*SalR*), codeine *O*-demethylase (*CODM*), with low levels of (R,S)-reticuline 7-*O*-methyltransferase (*7OMT*) and (R,S)-norreticuline 7-*O*-methyltransferase (*N7OMT*), all of which are known benzylisoquinoline alkaloid biosynthetic enzymes in opium poppy. The absence or low expression of these transcripts was linked to the absence of morphine in opium poppy cell cultures (Desgagné-Penix et al., 2010). However, another study showed the relative expression of peroxidases involved in the biosynthesis of hispidol, flavonoids in *Medicago truncatula* through phenylpropanoid pathway (Farag et al., 2009). This study showed that the peroxidase transcripts were almost equally expressed in both field derived tissues (leaf, stem, flower etc.) and *in vitro* cultures elicited with either methyl jasmonate or yeast extract. In alkaloid biosynthesis in *C. roseus*, the expression of pathway transcripts (e.g. *CYP72A57*, *OMTs*, *NMTs*, *NADPH*-reductase, hydroxylase, peroxidase) involved in the biosynthesis of monoterpene indole alkaloids including vinblastine (Góngora-Castillo et al., 2012) showed equal expression of *CYP72A57* and *OMTs* in field tissues (leaf, stem, flower, and root) and callus cultures. The other transcripts were more highly expressed in field tissues than callus cultures.

Furthermore, the transcripts leading to other alkaloids (isoquinoline, indole, and quinoline) were basically detected in both tissues which could suggest callus as a suitable material for the biosynthesis of other alkaloids/secondary metabolites than Amaryllidaceae alkaloids. This could also suggest a link between Amaryllidaceae alkaloid pathways to the defense or stress response in *Narcissus* as phenylpropanoid, hormone, flavonoid biosynthesis, and mevalonate pathways are linked to the biosynthesis of a diversity of secondary metabolites related to plant defense or stress response (Dixon, 2001).

The 206 (9.5% of total) transcripts were found to be significant differentially expressed (q-value ≤0.05) which might be a reason, that many important transcripts such as *PAL* and *TYDC*, involved in phenylalanine and tyrosine metabolism were not detected in differential expression analysis. Whereas, the candidate genes predicted in this project, involved in Amaryllidaceae alkaloid biosynthesis, such as *PAL* and *TYDC* have been previously identified using experimental methods in *N. pseudonarcissus* (Pulman, 2015; Kilgore et al., 2014, Kilgore et al., 2016a, Ferdausi et al., 2021a) and *L. aurea* (Wang et al., 2013). Therefore, future work should be concerned with the reconstruction of a reference transcriptome using *de novo* assembly techniques such as Trinity (Grabherr et al., 2011), MIRA (Chevreux et al., 2004) or SOAPdenovo (Ward et al., 2012).

**Conclusion**

Secondary metabolism in plant is critically influenced by the plant growth environment as well as tissue type. Therefore, the present study revealed the differential expression pattern of genes or transcripts, biosynthetic pathways leading to the different secondary metabolites, and their molecular regulation in *N. pseudonarcissus* field grown bulb (basal plate) and tissue culture derived callus. The findings also revealed that the major secondary metabolites related transcripts involved in later biosynthetic pathways were mainly up-regulated in basal plate while transcripts related to initial pathways of secondary metabolites biosynthesis and stress response factors were up-regulated in callus. This study also showed the transcriptomic differences between two different tissue types’ organized tissue such as basal plate and unorganized tissue, callus. However, further complete annotation of *Narcissus* reference transcriptome would be better for the confident identification of novel transcripts involved in secondary metabolite production in *Narcissus* field and *in vitro* tissues.

**Data availability statement**

The reference *Narcissus* transcriptome used in this study is available on MedPlant RNA Seq Database (<https://medplantrnaseq.org/>).

**Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Author’s contributions**

Conceptualization, M. J. and A. F.; Methodology, M. J., X. C. and A. F.; Investigation, M. J., X. C., and A. F.; Writing - Original Draft, A. F; Writing – Review & Editing, M. J. and X. C.; Funding Acquisition, A. F; Data acquisition, analysis and interpretation, A. F., Resources, M. J. and X. C.; Supervision, M. J and X. C.

**Funding**

The work was funded by the Commonwealth Scholarship Commission, United Kingdom and University of Liverpool, United Kingdom.

**Acknowledgements**

The authors would like to acknowledge Centre for Genomic Research, University of Liverpool for providing the sequenced data, Prof. Dr. Anthony Hall for his support on RNA-seq analyses, Dr. Ryan Johnson for annotating the transcripts and Dr. Jane Pulman for providing her data on previous *Narcissus* transcriptome.

**Supplementary materials**

The supplementary materials of this study can be found online as Supplementary 1 and Supplementary 2.

**References**

Berkov, S., Georgieva, L., Kondakova, V., Atanassov, A., Viladomat, F., Bastida, J., Codina, C., 2009. Plant sources of galanthamine: phytochemical and biotechnological aspects. Biotech. Biotech. Equipment. 23, 1170-1176. doi/pdf/10.1080/13102818.2009.10817633

Berkov, S., Ivanov, I., Georgiev, V., Codina, C., Pavlov, A., 2014. Galanthamine biosynthesis in plant in vitro systems. Eng. Life Sci. 14, 643-650. doi/pdf/10.1002/elsc.201300159

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E., Wetter, T., Suhai, S., 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14, 1147-1159. ISSN 1088-9051/04

Délano-Frier, J. P., Avilés-Arnaut, H., Casarrubias-Castillo, K., Casique-Arroyo, G., Castrillón-Arbeláez, P. A., Herrera-Estrella, L., Estrada-Hernández, M. G., 2011. Transcriptomic analysis of grain amaranth (*Amaranthus hypochondriacus*) using 454 pyrosequencing: comparison with *A. tuberculatus*, expression profiling in stems and in response to biotic and abiotic stress. BMC Genomics. 12(1), 363. [doi.org/10.1186/1471-2164-12-363](https://doi.org/10.1186/1471-2164-12-363)

Desgagné-Penix, I., Khan, M. F., Schriemer, D. C., Cram, D., Nowak, J., Facchini, P. J., 2010. Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. BMC Plant Biol. 10(1), 252. [doi.org/10.1186/1471-2229-10-252](https://doi.org/10.1186/1471-2229-10-252)

Desgagné-Penix, I., Farrow, S. C., Cram, D., Nowak, J., Facchini, P. J., 2012. Integration of deep transcript and targeted metabolite profiles for eight cultivars of opium poppy. Plant Mol. Biol. 79(3), 295-313. DOI 10.1007/s11103-012-9913-2

Dixon, R.A., 2001. Natural products and plant disease resistance. Nature. 411 (6839), 843-847.
doi.org/10.1038/35081178

Egan, A. N., Schlueter, J., Spooner, D. M., 2012. Applications of next-generation sequencing in plant biology. American J. Bot. 99, 175-185. doi/pdf/10.3732/ajb.1200020

Farag, M. A., Deavours, B. E., De Fátima, Â., Naoumkina, M., Dixon, R. A., Sumner, L. W., 2009. Integrated metabolite and transcript profiling identify a biosynthetic mechanism for hispidol in Medicago truncatula cell cultures. Plant Physiol. 151, 1096-1113. doi.org/10.1104/pp.109.141481

Ferdausi, A., Chang, X., Hall, A., Jones, M., 2020. Galanthamine production in tissue culture and metabolomic study on Amaryllidaceae alkaloids in *Narcissus pseudonarcissus* cv. Carlton. Ind. Crops Prod.144, 112058. [doi.org/10.1016/j.indcrop.2019.112058](https://doi.org/10.1016/j.indcrop.2019.112058)

Ferdausi, A., Chang, X., Jones, M., 2021a. Enhancement of galanthamine production through elicitation and NMR-based metabolite profiling in *Narcissus pseudonarcissus* cv. Carlton in vitro callus cultures. In Vitro Cell. Dev. Biol. Plant. pp.1-12. doi.org/10.1007/s11627-020-10139-z

Ferdausi, A., Chang, X., Hall, A., Jones, M., 2021b. Relative expression of putative genes involved in galanthamine and other Amaryllidaceae alkaloids biosynthesis in *Narcissus* field and in vitro tissues. Gene. p.145424. [doi.org/10.1016/j.gene.2021.145424](https://doi.org/10.1016/j.gene.2021.145424)

Goff, S. A., Vaughn, M., Mckay, S., Lyons, E., Stapleton, A. E., Gessler, D., Matasci, N., Wang, L., Hanlon, M., Lenards, A., 2011. The iPlant collaborative: cyberinfrastructure for plant biology. Front. Plant Sci. 2, 34. [doi.org/10.3389/fpls.2011.00034](https://doi.org/10.3389/fpls.2011.00034)

Góngora-Castillo, E., Childs, K. L., Fedewa, G., Hamilton, J. P., LiscombE, D. K., Magallanes-Lundback, M., Mandadi, K. K., Nims, E., Runguphan, W., Vaillancourt, B., 2012. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PloS one. 7, e52506. . doi:10.1371/journal.pone.0052506

Gong, H., Jiao, Y., Hu, W.W., Pua, E. C., 2005. Expression of glutathione-S-transferase and its role in plant growth and development in vivo and shoot morphogenesis in vitro. Plant Mol. Biol. 57, 53-66. DOI 10.1007/s11103-004-4516-1

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29, 644-652. DOI: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883)

Guo, X., Li, Y., Li, C., Luo, H., Wang, L., Qian, J., Luo, X., Xiang, L., Song, J., Sun, C., 2013. Analysis of the Dendrobium officinale transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. Gene. 527, 131-138. [doi.org/10.1016/j.gene.2013.05.073](https://doi.org/10.1016/j.gene.2013.05.073)

Haas, B. J., Zody, M. C., 2010. Advancing RNA-seq analysis. Nat. Biotech. 28, 421. [doi.org/10.1038/nbt0510-421](https://doi.org/10.1038/nbt0510-421)

Hall, D. E., Zerbe, P., Jancsik, S., Quesada, A. L., Dullat, H., Madilao, L. L., Yuen, M., Bohlmann, J., 2013. Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. Plant Physiol. 161, 600-616. doi.org/10.1104/pp.112.208546

He, M., Qu, C., Gao, O., Hu, X., Hong, X., 2015. Biological and pharmacological activities of amaryllidaceae alkaloids. RSC Adv. 5, 16562-16574. DOI: 10.1039/x0xx00000x

Hotchandani, T., de Villers, J., Desgagné-Penix, I., 2019. Developmental Regulation of the Expression of Amaryllidaceae Alkaloid Biosynthetic Genes in *Narcissus* papyraceus. Genes. 10 (8), 594. doi:10.3390/genes10080594

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., O'Donovan, C., 2015. The GOA database: gene ontology annotation updates for 2015. Nucleic Acids Res. 43, D1057-D1063. [doi.org/10.1093/nar/gku1113](https://doi.org/10.1093/nar/gku1113)

Johnson, M. T., Carpenter, E. J., Tian, Z., Bruskiewich, R., Burris, J. N., Carrigan, C. T., Chase, M. W., Clarke, N. D., Covshoff, S., Edger, P. P., 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PloS one. 7, e50226. [doi.org/10.1371/journal.pone.0050226](https://doi.org/10.1371/journal.pone.0050226)

Karuppusamy, S., 2009. A review on trends in production of secondary metabolites from higher plants by in vitro tissue, organ and cell cultures. J. Med. Plants Res. 3, 1222-1239. [doi.org/10.5897/JMPR.9000026](https://doi.org/10.5897/JMPR.9000026)

Kilgore, M. B., Augustin, M. M., Starks, C. M., O’neil-Johnson, M., May, G. D., Crow, J. A., Kutchan, T. M., 2014. Cloning and characterization of a norbelladine 4′-O-methyltransferase involved in the biosynthesis of the Alzheimer’s drug galanthamine in *Narcissus* sp. aff. *pseudonarcissus.* PloS one. 9, e103223. [doi.org/10.1371/journal.pone.0103223](https://doi.org/10.1371/journal.pone.0103223)

Kilgore, M. B., Augustin, M. M., May, G. D., Crow, J. A., Kutchan, T. M., 2016a. CYP96T1 of *Narcissus* sp. aff. *pseudonarcissus* Catalyzes Formation of the Para-Para'CC Phenol Couple in the Amaryllidaceae Alkaloids. Front. Plant Sci. doi: 10.3389/fpls.2016.00225.

Kilgore, M. B., Holland, C. K., Jez, J. M., and Kutchan, T. M. (2016b). Identification of a Noroxomaritidine Reductase with Amaryllidaceae Alkaloid Biosynthesis Related Activities. J. Biol. Chem. jbc. M116. 717827. [doi.org/10.1074/jbc.M116.717827](https://doi.org/10.1074/jbc.M116.717827)

Kumar, S., Banks, T. W., Cloutier, S., 2012. SNP discovery through next-generation sequencing and its applications. Int. J. Plant Genomics. doi:10.1155/2012/831460

Lee, J., Kang, Y., Shin, S. C., Park, H., Lee, H., 2014. Combined analysis of the chloroplast genome and transcriptome of the Antarctic vascular plant Deschampsia antarctica Desv. PloS one. 9, e92501. [doi.org/10.1371/journal.pone.0092501](https://doi.org/10.1371/journal.pone.0092501)

Liscombe, D. K., Ziegler, J., Schmidt, J., Ammer, C., Facchini, P. J., 2009. Targeted metabolite and transcript profiling for elucidating enzyme function: isolation of novel N‐methyltransferases from three benzylisoquinoline alkaloid‐producing species. The Plant J. 60, 729-743. doi/pdfdirect/10.1111/j.1365-313X.2009.03980.x

Liscombe, D. K., Louie, G. V., Noel, J. P., 2012. Architectures, mechanisms and molecular evolution of natural product methyltransferases. Nat. Prod. Rep. 29, 1238-1250. [doi.org/10.1039/C2NP20029E](https://doi.org/10.1039/C2NP20029E)

Lubbe, A., Gude, H., Verpoorte, R., Choi, Y. H., 2013. Seasonal accumulation of major alkaloids in organs of pharmaceutical crop *Narcissus* Carlton. Phytochem. 88, 43-53. doi.org/10.1016/j.phytochem.2012.12.008

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 5, 621-628.
doi.org/10.1038/nmeth.1226

Mulabagal, V., Tsay, H. S., 2004. Plant cell cultures-an alternative and efficient source for the production of biologically important secondary metabolites. Int. J. Appl. Sci. Eng. 2, 29-48. ISSN 1727-2394

Murashige, T., Skoog, F., 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Planta. 15, 473-497.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 320, 1344-1349. DOI: 10.1126/science.1158441

Nalawade, S. M., Tsay, H. S., 2004. In vitro propagation of some important Chinese medicinal plants and their sustainable usage. In Vitro Cel. Dev. Biol. Plant. 40, 143-154.
doi.org/10.1079/IVP2003504

Neelakandan, A. K., Wang, K., 2012. Recent progress in the understanding of tissue culture-induced genome level changes in plants and potential applications. Plant Cell Rep. 31, 597-620. [doi.org/10.1007/s00299-011-1202-z](https://doi.org/10.1007/s00299-011-1202-z)

Ober, D., 2005. Seeing double: gene duplication and diversification in plant secondary metabolism. Trends Plant Sci. 10, 444-449. [doi.org/10.1016/j.tplants.2005.07.007](https://doi.org/10.1016/j.tplants.2005.07.007)

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M., 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 27, 29-34. [doi.org/10.1093/nar/27.1.29](https://doi.org/10.1093/nar/27.1.29)

Ogawa, D., Yamaguchi, K., Nishiuchi, T., 2007. High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth. J. Exp. Bot. 58, 3373-3383. [doi.org/10.1093/jxb/erm184](https://doi.org/10.1093/jxb/erm184)

Oksman-Caldentey, K. M., Inzé, D., 2004. Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci. 9, 433-440. [doi.org/10.1016/j.tplants.2004.07.006](https://doi.org/10.1016/j.tplants.2004.07.006)

Oliver, S. L., Lenards, A. J., Barthelson, R. A., Merchant, N., Mckay, S. J., 2013. Using the iPlant collaborative discovery environment. Curr. Protoc. Bioinform. 1.22. 1-1.22. 26. doi/abs/10.1002/0471250953.bi0122s42

Osorio, E. J., Berkov, S., Brun, R., Codina, C., Viladomat, F., Cabezas, F., Bastida, J., 2010. In vitro antiprotozoal activity of alkaloids from Phaedranassa dubia (Amaryllidaceae). Phytochem. Let. 3, 161-163. [doi.org/10.1016/j.phytol.2010.06.004](https://doi.org/10.1016/j.phytol.2010.06.004)

Park, C. H., Yeo, H. J., Park, Y. E., Baek, S. A., Kim, J. K., Park, S. U., 2019. Transcriptome analysis and metabolic profiling of *Lycoris radiata*. Biology, 8(3), 63. [doi.org/10.3390/biology8030063](https://doi.org/10.3390/biology8030063)

Pulman, J., 2015. A transcriptomics approach to understanding polymorphic and transcript level differences linked to isoquinoline alkaloid production in triploid varieties of *Narcissus pseudonarcissus*. PhD Thesis. University of Liverpool.

Romero, I.G., Pai, A.A., Tung, J., Gilad, Y., 2014. RNA-seq: impact of RNA degradation on transcript quantification. BMC Biol. 12(1), pp.1-13. <https://doi.org/10.1186/1741-7007-12-42>

Sengupta, D., Naik, D., Reddy, A. R., 2015. Plant aldo-keto reductases (AKRs) as multi-tasking soldiers involved in diverse plant metabolic processes and stress defence: A structure-function update. J. Plant Physiol. 179, 40-55. [doi.org/10.1016/j.jplph.2015.03.004](https://doi.org/10.1016/j.jplph.2015.03.004)

Singh, A., Desgagné-Penix, I., 2014. Biosynthesis of the Amaryllidaceae alkaloids. Plant Sci. Today. 1, 114-120. [doi.org/10.14719/pst.2014.1.3.41](https://doi.org/10.14719/pst.2014.1.3.41)

Singh, A., Desgagné-Penix, I., 2017. Transcriptome and metabolome profiling of *Narcissus* *pseudonarcissus* ‘King Alfred’ reveal components of Amaryllidaceae alkaloid metabolism. Sci. Reps. *7*(1), 1-14. [doi.org/10.1038/s41598-017-17724-0](https://doi.org/10.1038/s41598-017-17724-0)

Strickler, S., Bombarely, A., Mueller, L., 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. American J. Bot. 99(2): 257–266. [doi.org/10.3732/ajb.1100292](https://doi.org/10.3732/ajb.1100292)

Tahchy, A. E., Boisbrun, M., Ptak, A., Dupire, F., Chrétien, F., Henry, M., Chapleur, Y., Laurain-Mattar, D., 2010. New method for the study of Amaryllidaceae alkaloid biosynthesis using biotransformation of deuterium-labeled precursor in tissue cultures. Acta. Biochimica. Polonica. 57, 75-82.[doi.org/10.18388/abp.2010\_2375](https://doi.org/10.18388/abp.2010_2375)

Tahchy, A. E., Bordage, S., Ptak, A., Dupire, F., Barre, E., Guillou, C., Henry, M., Chapleur, Y., Laurain-Mattar, D., 2011. Effects of sucrose and plant growth regulators on acetylcholinesterase inhibitory activity of alkaloids accumulated in shoot cultures of Amaryllidaceae. PCTOC. 106, 381-390. [doi.org/10.1007/s11240-011-9933-7](https://doi.org/10.1007/s11240-011-9933-7)

Takos, A. M., Rook, F., 2013. Towards a molecular understanding of the biosynthesis of amaryllidaceae alkaloids in support of their expanding medical use. Int. J. Mol. Sci. 14, 11713-11741. [doi.org/10.3390/ijms140611713](https://doi.org/10.3390/ijms140611713)

Tello-Ruiz, M. K., Stein, J., Wei, S., Preece, J., Olson, A., Naithani, S., Amarasinghe, V., Dharmawardhana, P., Jiao, Y., Mulvaney, J., 2015. Gramene 2016: comparative plant genomics and pathway resources. Nucleic Acids Res. gkv1179. [doi.org/10.1093/nar/gkv1179](https://doi.org/10.1093/nar/gkv1179)

Trapnell, C., Pachter, L., Salzberg, S. L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 25, 1105-1111. [doi.org/10.1093/bioinformatics/btp120](https://doi.org/10.1093/bioinformatics/btp120)

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotech. 28, 511-515. [doi.org/10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621)

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562-578. [doi.org/10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016)

Verpoorte, R., Contin, A., Memelink, J., 2002. Biotechnology for the production of plant secondary metabolites. Phytochem. Rev. 1, 13-25. [doi.org/10.1023/A:1015871916833](https://doi.org/10.1023/A%3A1015871916833)

Vijay, N., Poelstra, J. W., Künstner, A., Wolf, J. B., 2013. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA‐seq experiments. Mol. Ecol. 22, 620-634. [doi.org/10.1111/mec.12014](https://doi.org/10.1111/mec.12014)

Wang, W., Wang, Y., Zhang, Q., Qi, Y., Guo, D., 2009. Global characterization of Artemisia annua glandular trichome transcriptome using 454 pyrosequencing. BMC Genomics. 10 (1), p.1. [doi.org/10.1186/1471-2164-10-465](https://doi.org/10.1186/1471-2164-10-465)

Wang, R., Xu, S., Jiang, Y., Jiang, J., Li, X., Liang, L., He, J., Peng, F., Xia, B., 2013. De novo sequence assembly and characterization of *Lycoris aurea* transcriptome using GS FLX titanium platform of 454 pyrosequencing. PloS one. 8, e60449. [doi.org/10.1371/journal.pone.0060449](https://doi.org/10.1371/journal.pone.0060449)

Wang, R., Han, X., Xu, S., Xia, B., Jiang, Y., Xue, Y., Wang, R., 2019. Cloning and characterization of a tyrosine decarboxylase involved in the biosynthesis of galanthamine in *Lycoris aurea*. Peer J. 7, e6729. [doi.org/10.7717/peerj.6729](https://doi.org/10.7717/peerj.6729)

Ward, J. A., Ponnala, L., Weber, C. A., 2012. Strategies for transcriptome analysis in nonmodel plants. American J. Bot. 99, 267-276. <https://doi.org/10.3732/ajb.1100334>

Weber, A. P., Weber, K. L., Carr, K., Wilkerson, C., Ohlrogge, J. B., 2007. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol. 144, 32-42. doi.org/10.1104/pp.107.096677

Wolf, J. B., 2013. Principles of transcriptome analysis and gene expression quantification: an RNA‐seq tutorial. Mol. Ecol. Res. 13, 559-572. <https://doi.org/10.1111/1755-0998.12109>

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

Zhao, Y. J., Cheng, Q. Q., Su, P., Chen, X., Wang, X. J., Gao, W., Huang, L. Q., 2014. Research progress relating to the role of cytochrome P450 in the biosynthesis of terpenoids in medicinal plants. Appl. Microbiol. Biotech. 98, 2371-2383. [doi.org/10.1007/s00253-013-5496-3](https://doi.org/10.1007/s00253-013-5496-3)

**Table caption list:**

**Table 1** TopHat output obtained from *Narcissus* RNA-seq data (Discovery Environment, CyVerse) against *Narcissus* transcriptome from MedPlant RNA-Seq Databases

**Table 2** Probable genes related to secondary metabolite production detected in *Narcissus* basal plate and callustranscriptome data (significant genes with q-value ≤ 0.05)

**Table 3** Pathways related to secondary metabolite production and central metabolism, detected in *N. pseudonarcissus* basal plate and callus by mapping to curated *Oryza sativa* pathway database (Plant Reactome).

**Figure caption list:**

**Fig. 1.** An overview of methods used for the *N. pseudonarcissus* cv. Carlton transcriptome (RNA-seq) data analysis

**Fig. 2.** Volcano plot (CummeRbund) indicating the presence of differentially expressed genes between Carlton basal plate and callus. Significantly, expressed genes are indicated by blue dots and red dots are representing the non-significant genes.

**Fig. 3.** Distribution patterns of the eleven abundant gene groups that were differentially expressed between basal plate and callus. The percentages contributed 41.7% to the total annotated genes (83.46%).

**Fig. 4.** Heatmaps generated using CummeRbund for top 50 transcripts sorted by Cuffdiff showing differential expression pattern of transcripts in Carlton basal plate and callus. Blue box = transcripts expressed in callus, red box = transcripts expressed in Carlton basal plate and black box = transcripts expressed in both tissues

**Fig. 5.** The abundance of important gene groups besides secondary metabolism related genes detected as differentially expressed in basal plate and callus with q-values ≤ 1.0.

**Fig. 6.** The UniProt Gene Ontology (GO) analysis of significantly up-regulated transcripts detected in **(a)** basal plate and **(b)** callus (q-value ≤ 0.05). The numbers in bracket represent the total transcripts assigned for the respective GO categories. Several transcripts were assigned to more than one category.

**Fig. 7.** An overview of GO assignment for basal plate and callus GO IDs annotated against significant up-regulated transcripts at q-value ≤ 0.05; using EMBL-EBI Quick GO-Beta annotations. GO terms contributing less than 1% are not included.

**Fig. 8.** Metabolic networks from D-Glucose to secondary metabolites in *N. pseudonarcissus* developed from KEGG and Plant Reactome pathway analyses. The postulated transcripts (italics) corresponding to enzymes expressed in different *Narcissus* tissues are labeled in different colors; callus (blue), basal plate (purple) and both callus and basal plate (red) next to arrows were identified in the *Narcissus* transcriptome data. HMG= Hydroxyl methyl glutarate; ERF= Ethylene response factor; PAL = Phenyl ammonia lyase; TYDC = Tyrosine decarboxylase; AUX=Auxin related enzyme; OMT = *O*-methyltransferase.

**Table 1**

|  |  |  |  |
| --- | --- | --- | --- |
| Samples | Read alignment rate (% of input) | Overall read alignment rate (%) | Concordant pair alignment rate (%) |
| Left read | Right read |
| CBS1 | 35.1 | 33.8 | 34.5 | 30 |
| CBS2 | 19.3 | 19 | 19.1  | 17.1 |
| CBS3 | 48.6 | 47.3 | 47.9 | 40.9 |
| CAL1 | 41.8 | 39.4 | 40.6 | 35.5 |
| CAL2 | 36.5 | 35.5 | 36 | 31.8 |
| CAL3 | 40.3 | 38.1 | 39.2 | 33.4 |

CBS = Carlton basal plate; CAL = Callus; 1, 2, 3 = three replicates of basal plate and callus

**Table 2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Gene ID | UniProt ID | Regulation | log2 (fold change) | q-value | Gene function |
| 1. Phenylpropanoid pathway related  |
|

|  |
| --- |
| XLOC\_001204XLOC\_002108 |

 | O04892A0A124SCB7 | CBS-UPCBS-UP |  -  - | 0.0070.007 | cytochrome P450cytochrome P450 |
| 2. Methyltransferases |
| XLOC\_001033XLOC\_001843XLOC\_000957XLOC\_000735 | [B9SGP1](http://www.uniprot.org/uniprot/B9SGP1)K4CXY9[Q5DNB1](http://www.uniprot.org/uniprot/Q5DNB1)[F6H5H8](http://www.uniprot.org/uniprot/F6H5H8) | CBS-UPCBS-UPCBS-UPCBS-UP | - 3.05 1.78  - | 0.0060.0020.0130.006 | *O*-methyltransferase *O*-methyltransferase  *S*-adenosylmethionine synthase *O*-methyltransferase  |
| XLOC\_002003XLOC\_000918 | A0A0P0YCT2[A0A0V0I0R3](http://www.uniprot.org/uniprot/A0A0V0I0R3) | CAL-UPCAL-UP | 2.01  -  | 0.011 0.0006 | S-methyltransferase hydroquinone methyltransferase |
| 3. NADP/ NADPH related |
| XLOC\_000518XLOC\_001407 | B9T8W8[Q9SLN8](http://www.uniprot.org/uniprot/Q9SLN8) | CBS-UPCBS-UP | 2.11 2.18 | 0.0040.006 | NADH dehydrogenase (NADP(+) reductase |
| XLOC\_001713 | E5G6F3 | CAL-UP | 3.05 | 0.049 | NAD/ NADP binding |
| 4. Oxidoreductase activity related |
|

|  |
| --- |
| XLOC\_001731XLOC\_001235XLOC\_001407 |

 | Q9AV39[T2DPZ9](http://www.uniprot.org/uniprot/T2DPZ9)M0SZ88 | CBS-UPCBS-UPCBS-UP | 2.09 - 2.18 | 0.0450.00060.045 | Os10g0545200 proteincytochrome buncharacterized  |
| XLOC\_001110XLOC\_000371XLOC\_002051XLOC\_002072 | M8AU02[M0TCR0](http://www.uniprot.org/uniprot/M0TCR0)[F2NYJ0](http://www.uniprot.org/uniprot/F2NYJ0)[O24428](http://www.uniprot.org/uniprot/O24428) | CAL-UPCAL-UPCAL-UPCAL-UP | 3.21- - - | 0.0330.00060.00060.0006 | aldehyde dehydrogenaseuncharacterized anaerobic reductasestearoyl-ACP desaturase |
| 5. Hydrolase activity related |
| XLOC\_000089XLOC\_000792XLOC\_001578 | [A0A078EUD2](http://www.uniprot.org/uniprot/A0A078EUD2)[K4NZ15](http://www.uniprot.org/uniprot/K4NZ15)[K4NZ15](http://www.uniprot.org/uniprot/K4NZ15) | CBS-UPCBS-UPCBS-UP | 2.93 - - | 0.0230.00060.0006 | BnaAnng00280D lipaseuncharacterized |
| XLOC\_000006XLOC\_000580XLOC\_001722XLOC\_001883 | Q9LLC2 [M0TI33](http://www.uniprot.org/uniprot/M0TI33)[A5BND5](http://www.uniprot.org/uniprot/A5BND5) F6GXE7 | CAL-UPCAL-UPCAL-UPCAL-UP | 2.88 - - - | 0.011 0.00060.00060.0006 | xyloglucan endo-transglucosylase xyloglucan endo-transglucosylase xyloglucan endo-transglucosylase xyloglucan endo-transglucosylase  |
| Others |
| XLOC\_000536 | [K4P0T2](http://www.uniprot.org/uniprot/K4P0T2) | CBS-UP | - | 0.0007 | lectin  |
| XLOC\_000616 | [W5VXS2](http://www.uniprot.org/uniprot/W5VXS2) | CBS-UP | - | 0.0007 | polyphenol oxidase |
| XLOC\_001502 | [G8XUP0](http://www.uniprot.org/uniprot/G8XUP0) | CBS-UP | - | 0.0007 | Lectin |
| XLOC\_001926 | [C9W8B3](http://www.uniprot.org/uniprot/C9W8B3) | CBS-UP | - | 0.0007 | lectin  |
| XLOC\_002069 | [Q40422](http://www.uniprot.org/uniprot/Q40422) | CBS-UP | - | 0.0007 | mannose specific lectin |
| XLOC\_000485 | [G3GC08](http://www.uniprot.org/uniprot/G3GC08) | CBS-UP | 6.00 | 0.0038 | Lipoxygenase |
| XLOC\_000278 | [A0A0B0MIR6](http://www.uniprot.org/uniprot/A0A0B0MIR6) | CBS-UP | 2.64 | 0.0212 | proactivator polypeptide |
| XLOC\_001282 | K7P8F2 | CBS-UP | 2.58 | 0.0383 | aspartic acid protease |
| XLOC\_001804 | [F4JLV7](http://www.uniprot.org/uniprot/F4JLV7) | CBS-UP | - | 0.0007 | lipid-transfer  |
| XLOC\_001339 | G9M5T0 | CAL-UP | 4.67 | 0.0254 | Lectin |
| XLOC\_000721 | K3ZUW3 | CAL-UP | 1.90 | 0.0113 | Uncharacterized |
| XLOC\_001055 | E0CWD0 | CAL-UP | 2.80 | 0.0007 | alcohol dehydrogenase |
| XLOC\_001373 | [Q9XHL5](http://www.uniprot.org/uniprot/Q9XHL5) | CAL-UP | - | 0.0034 | 3-hydroxy-3-methylglutaryl-coenzyme A reductase |
| XLOC\_000638 | [J5JKX7](http://www.uniprot.org/uniprot/J5JKX7) | CAL-UP | - | 0.0007 | PAP2 superfamily  |
| XLOC\_001712 | [Q5XEP9](http://www.uniprot.org/uniprot/Q5XEP9) | CAL-UP | - | 0.0007 | 3-ketoacyl-CoA synthase  |

CBS-UP = genes up-regulated in basal plate (down regulated in callus); CAL-UP = genes up-regulated in callus (down regulated in basal plate); (-) indicates the transcripts are either expressed in basal plate or callus but not in both, therefore showing no fold change.

**Table 3**

|  |
| --- |
| Carlton basal plate |
| Pathway name | Entities found | Entities total | Entities p-value | Entities FDR | Mapped entities(UniProt ID) | *Narcissus* transcript IDs |
| Plant pathways | 6 | 1499 |  |
| S-adenosyl-L-methionine cycle | 2 | 6 | 0.013 | 0.09 | Q9LGU6; Q0DKY4 | XLOC\_000957, XLOC\_000323, XLOC\_001134, XLOC\_001181, XLOC\_001596, XLOC\_002071  |
| Phenylpropanoid biosynthesis | 3 | 26 | 0.03 | 0.11 | A2Y626; Q6ERR3; B8BB38 | XLOC\_000518, XLOC\_001407, XLOC\_001126, XLOC\_000218, XLOC\_001299, XLOC\_000831, XLOC\_000898 |
| Secondary metabolite biosynthesis | 3 | 203 | 0.94 | 0.94 | A2Y626; Q6ERR3; B8BB38 | XLOC\_000518, XLOC\_001407, XLOC\_001126, XLOC\_000218, XLOC\_001299, XLOC\_000831, XLOC\_000898, XLOC\_000263, XLOC\_000481, XLOC\_000686, XLOC\_000163, XLOC\_000396, XLOC\_001312 |
| Amino acid metabolism | 3 | 211 | 0.95 | 0.95 | Q9LGU6; Q0DKY4; Q0JC10 | XLOC\_000397, XLOC\_000094 |
| Hormone biosynthesis | 2 | 449 | 0.99 | 0.99 | Q9LGU6; Q0DKY4 | XLOC\_000957, XLOC\_000323, XLOC\_001134, XLOC\_001181, XLOC\_001596, XLOC\_002071 |
| Callus |
| Plant pathways | 31 | 1499 |  |
| Phenylpropanoid biosynthesis | 8 | 26 | 0.001 | 0.015 | Q6YYZ2; Q6ERR3; Q0DV32; Q67W82; P17814; B8BB38; Q42982; Q6ETN3 | XLOC\_001712, XLOC\_001713, XLOC\_001353, XLOC\_001227, XLOC\_001353, XLOC\_000670, XLOC\_001171, XLOC\_001310, XLOC\_001760, XLOC\_000155, XLOC\_001462, XLOC\_001484, XLOC\_001597, XLOC\_001851, XLOC\_000446, XLOC\_001557 |
| Flavonoid biosynthesis | 6 | 18 | 0.003 | 0.03 | Q6YYZ2; Q0DV32; Q67W82; P17814; Q42982; Q6ETN3 | XLOC\_001227, XLOC\_001353 |
| Glutamate biosynthesis  | 2 | 2 | 0.011 | 0.05 | Q0JKD0; Q0DG35 | XLOC\_000259, XLOC\_000583' XLOC\_000892, XLOC\_00145, XLOC\_00373 |
| Mevalonate pathway | 2 | 15 | 0.33 | 0.47 | Q9XHL5; Q0J0M8 | XLOC\_001373, XLOC\_001760 |
| S-adenosyl-L-methionine cycle | 1 | 6 | 0.37 | 0.47 | Q2QLY5 | XLOC\_001333, XLOC\_001791, XLOC\_000186, XLOC\_001350, XLOC\_001633 |
| Amino acid metabolism | 4 | 211 | 0.99 | 0.99 | Q2QLY5; Q0JKD0; Q0DG35; P37833 | XLOC\_001611, XLOC\_000822, XLOC\_001825, XLOC\_002003, XLOC\_000364 |

p-value = probability that the overlap between the query and the pathway has occurred by chance; FDR = probability corrected for multiple comparisons.